Vision Research 50 (2010) 1519–1531

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Inconsistency of perceived 3D shape

M. Di Luca [a,b], F. Domini [b,d,*], C. Caudek [c]

[a] Max Planck Institute for Biological Cybernetics, Tuebingen, Germany
[b] Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI, USA
[c] Department of Psychology, Università degli Studi di Firenze, Firenze, Italy
[d] Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto (Tn), Italy

ARTICLE INFO

Article history:
Received 5 February 2010
Received in revised form 5 May 2010

Keywords:
3D shape
Perception
Representation
Depth
Slant
Curvature
Cue integration

ABSTRACT

Internal consistency of local depth, slant, and curvature judgments was studied by asking participants to match two 3D surfaces rendered by different mixtures of 3D cues (velocity, texture, and shading). We found that perceptual judgments were not consistent with each other, with cue-specific distortions. Adding multiple cues did not eliminate the inconsistencies of the judgments. These results can be predicted by the Intrinsic Constraint (IC) model according to which the perceptual metric local estimates are a monotonically increasing function of the Signal-to-Noise Ratio of the optimal combination of direct information of 3D shape (Domini, Caudek, & Tassinari, 2006).

1. Introduction

The fundamental problem of three-dimensional (3D) shape perception is to identify the perceptual processes yielding (a) a conscious experience of solid shape, and (b) the visuomotor mechanisms allowing a successful interaction with the environment. The first of these two aspects is the focus of this work. Two main schools of thought have emerged in this respect. According to one, the goal of visual processing is to create "visual representations" that are isomorphic with the environmental objects. Euclidean geometry can be used to describe the geometric properties of both the environmental objects and the observer's "internal representations". Such class of models will therefore be called "Euclidean".

"Euclidean" models treat 3D shape perception as the solution of the inverse problem of generating 3D descriptions from images (e.g., Horn, 1986; Kersten, Mamassian, & Yuille, 2004). Akin to Helmoltzian perceptual inference, 3D shape perception is understood as the re-construction of the information that has been lost in the transformation from the distal to proximal stimulus. The "shape-from-X" problem (where X can be stereo, texture, shading, etc.) is an ill-defined problem, because a veridical 3D metric re-construction cannot be obtained from the retinal images alone, but it requires additional knowledge about the ("missing") scene parameters and a-priori knowledge about the structure and motion of the environmental objects (e.g., Bülthoff & Yuille, 1990). One example is provided by the rigidity assumption for the solution of the Structure-from-Motion problem (Ullman, 1979).

Another school of thought postulates that only some geometrical properties of the environmental objects are represented in an isomorphic fashion in the brain. Specifically, it is hypothesized that 3D shape perception is based on the information that is directly specified by the retinal images (not inferred through the estimation of "missing" parameters). Lappin and Craft (2000) presented such view by saying that "[s]patial vision requires reliable and approximately one-to-one spatial correspondences between environmental objects, retinal images, and their representations in the visual nervous system and perception. The spatial information for vision is defined by such mutual correspondences" (p. 6). In the same spirit, other researchers have speculated that spatial vision relies on "image primitives" that provide an invariant and reliable one-to-one correspondence with specific properties of the environmental objects. The relative binocular disparities, for example, directly specify the viewer-centered depth-order relationships, the second-order derivatives of the retinal disparities differentiate planar and curved surfaces, and so on. In general, retinal signals like disparity fields, velocity fields, texture gradients, and luminance gradients directly specify local "affine" structure (Koenderink, 1990; Todd & Bressan, 1990). We will therefore term the computational models relying on such image primitives as "Local Affine".

* Corresponding author at: Brown University, Providence, RI, USA.
E-mail address: Fulvio_Domini@brown.edu (F. Domini).
URL: http://www.cog.brown.edu/research/3Dspl/index.html (F. Domini).

0042-6989/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2010.05.006

Is 3D shape perception better described by "Euclidean" or "Local Affine" theories? Neither seems to suffice. "Euclidean" models are at odds with the biases of perceived 3D shape. In fact, observers' judgments of Euclidean 3D properties are often inaccurate (Bradshaw, Parton, & Glennerster, 2000; Caudek & Domini, 1998; Caudek & Proffitt, 1993; Caudek & Rubin, 2001; Di Luca, Domini, & Caudek, 2004; Domini & Caudek, 1999; Domini & Caudek, 2003a, 2003b; Domini, Caudek, & Richman, 1998; Fantoni, 2008; Fantoni, Caudek, & Domini, 2010; Glennerster, Rogers, & Bradshaw, 1996; Hecht, van Doorn, & Koenderink, 1999; Tittle, Todd, Perotti, & Norman, 1995; but see also Brenner & van Damme, 1999; Durgin, Proffitt, Olson, & Reinke, 1995; Johnston & Passmore, 1994a, 1994b), with a large variance within and across observers (Todd, 2004; Todd & Norman, 2003). "Euclidean" models attribute perceptual distortions to incorrect estimations of the "missing" parameters. For example, an incorrect estimate of the viewing distance is hold responsable for the distortions in perceived shape-from-stereo (e.g., Johnston, 1991). But this account can be regarded at best as an ex post facto explanation: in fact, it is not possible to quantify a priori the mis-estimation of a "missing" parameter. As a consequence, no predictions can be made about perceptual biases in perceived 3D shape.

On the other hand, the scope of "Local Affine" theories is limited. Such approach is consistent with the empirical findings showing that local affine judgements are reliable and accurate at hyperacuity levels (van Boxtel, Wexler, & Droulez, 2003; Cornilleau-Pérès et al., 2002; Dijkstra, Cornilleau-Pérès, Gielen, & Droulez, 1995; Lappin & Craft, 2000), but it cannot be applied to perception of metric 3D shape.

The present investigation intends to contribute to the development of a general theory capable of explaining both the biases of metric judgments and the precision of affine discriminations. In this paper, we will first present our hypothesis and then we will test the proposed approach in two experiments.

1.1. Direct information about 3D shape

1.1.1. Metric maps for depth, slant, and curvature

The stimuli used in the present investigation were (single- or multiple-cue) renderings of an elliptic paraboloid of revolution aligned with the line of sight (z). Within an orthogonal coordinate system (x,y,z), such a surface is usually described in terms of the *depth-map*

$$z(x,y) = \frac{C}{2}(x^2 + y^2). \quad (1)$$

The constant C determines the surface elongation along the z dimension. In Fig. 1, the depth separation between the tip of the paraboloid and another point of its surface is represented by the length of the red segment.

The depth-map provides one of many equivalent descriptions of this 3D structure. Being smooth and locally differentiable, the surface can also be described in terms of a *slant-map* $s(x,y)$, where $s(x,y) = \sqrt{\left(\frac{\delta z}{\delta x}\right)^2 + \left(\frac{\delta z}{\delta y}\right)^2}$. The slant at a surface location is the angle between the fronto-parallel plane and the plane tangent to the surface (the green square in Fig. 1). For an elliptic paraboloid of revolution, $s(x,y) = C\sqrt{x^2 + y^2}$.

Another equivalent metric description is provided by the *curvature-map* $c(x,y)$, where $c(x,y) = \sqrt{k_{min}^2 + k_{max}^2}$, and k_{min} and k_{min} are the principal curvatures at the point (x,y). For an elliptic paraboloid of revolution, the curvature is constant: $c(x,y) = C$. In Fig. 1, the principal curvatures at the tip of the paraboloid are represented by the two blue arcs. The principle curvature values are the inverse of the radii of these two arcs.

1.1.2. Local affine maps for depth, slant, and curvature

Fig. 1 represents four paraboloids characterized by different values of the constant C. For the Euclidean geometry, the depth-map, slant-map, and curvature-map are equivalent: given one of these descriptions, another can be found through derivation or integration. As seen in Fig. 1, the value of C uniquely determines the local metric depth, slant, and curvature.

For the Affine geometry, conversely, the metric of the depth-, slant-, and curvature-maps becomes meaningless: all four surfaces shown in Fig. 1 are identical, because a local affine representation preserves the depth-map, the slant-map, and the curvature-map only up to an unknown scaling constant k. In the following, we will refer to $k_z z(x,y)$, $k_s s(x,y)$, and $k_c c(x,y)$ as to the *local affine* depth-, slant-, and curvature-maps.

1.1.3. Estimation of local affine maps

In our previous work, we extended the Local Affine approach (a) by showing that the local affine depth-, slant-, and curvature-maps can be estimated in an optimal manner by combining different retinal signals (Domini et al., 2006), and (b) by presenting a probabilistic model that provides a metric scaling to these estimates (Domini and Caudek, 2003; Domini et al., 1997; Tassinari et al., 2008).

Domini et al. (2006) hypothesized that, for any retinal signal s measured at (x,y), it is always possible to find a mathematical description $s_p(x,y)$ that is proportional to the local metric 3D property p (e.g., depth, slant, or curvature):

$$s_p(x,y) = k_{sp} p(x,y) + \varepsilon_{sp}, \quad (2)$$

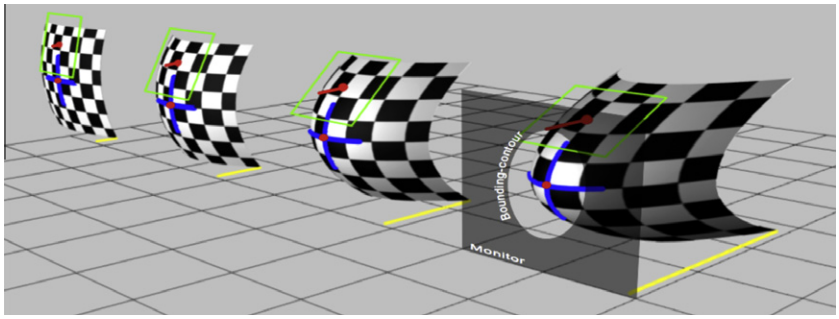

Fig. 1. Schematic representation of the elliptic paraboloid of revolution simulated in this study. The checkerboard pattern is shown for illustrative purposes only (a picture of the actual stimuli is shown in Fig. 7). The yellow segment at the bottom of each paraboloid indicates the magnitude of the constant C that determines the elongation of the surface. In the two experiments, participants were asked to judge (a) the curvature at the tip of the surface (the blue cross represent the two principal curvatures), (b) the slant of a local surface patch (the green square represents the plane tangent to the surface), and (c) the depth separation between the tip and another point on the surface (red segment). The fours paraboloids in the Figure differ for their Euclidean, but not for their Affine, structure.

where k_{sp} is a constant that depends on the visual scene and ε_{sp} is the measurement noise of $s_p(x,y)$. The measurement noise is assumed to be Gaussian with zero mean and standard deviation σ_{sp}.

Definition 1. The appropriate mathematical description of the signal s provides an estimate $s_p(x,y)$ of the local affine property-map p.

Consider the recovery of surface curvature from the optic flow. If we are looking for a description that is invariant within an affine space, then $v_c(x,y)$ directly specifies the local affine curvature-map with the additive uncertainty $\varepsilon_{vc} : \mathscr{E}[v_c(x,y)] = k_{vc}c(x,y)$. For the points (x_0,y_0) and (x_1,y_1) belonging to a local neighborhood, $v_c(x,y)$ allows judgements of the type: $c(x_0,y_0) > c(x_1,y_1)$, that is, the curvature at (x_0,y_0) is larger than the curvature at (x_1,y_1). In fact, $\mathscr{E}[v_c(x_0,y_0)] > \mathscr{E}[v_c(x_1,y_1)] \Longleftrightarrow c(x_0,y_0) > c(x_1,y_1)$.

1.1.4. Estimation of local affine maps from velocity signals

In this section, we will show how the local affine depth-, slant-, and curvature-maps can be estimated from the velocity field. The relative velocity field produced by the projection of a surface's rotation about the vertical axis can be approximated by the following equation:

$$v_z(x,y) \approx \frac{\omega}{z_f} z(x,y) + \varepsilon_{vz}, \tag{3}$$

where ω is the angular velocity, z_f is the absolute distance to a reference point on the surface, and ε_{vz} is additive Gaussian noise. Eq. (3) shows that the velocity field is proportional to the *depth-map*.

The gradient of the velocity field, defined as the *def* component of the optic flow, is proportional to the *slant-map*:

$$v_s(x,y) = def(x,y) \approx \frac{\omega}{z_f} s(x,y) + \varepsilon_{vs}, \tag{4}$$

where ε_{vs} is Gaussian noise affecting the *def* measurement (Koenderink, 1986).

The intensity of the second-order spatial derivative of the velocity field is proportional to the *curvature-map*:

$$v_c(x,y) \approx \frac{\omega}{z_f} c(x,y) + \varepsilon_{vc}, \tag{5}$$

where ε_{vc} is additive Gaussian noise.

We hypothesize that the additive Gaussian noise of Eqs. (3)–(5) is constant only within a small neighborhood. In other words, σ_{vz} varies across different local neighborhoods; the same can be said for σ_{vs} and σ_{vc}.

It is beyond the scope of the present work to describe how the local affine maps can be estimated from other image signals (e.g., shading or texture). What is important for the present discussion is the hypothesis that, for each image signal, it is possible to find a mathematical description that is proportional to the local depth-, slant-, or curvature-map.

1.1.5. The intrinsic constraints model

So far, we discussed the problem of estimating the local affine depth-, slant-, and curvature-maps from a single image signal. Now, let us focus on the estimation of local surface curvature and let us consider the case of multiple image signals (Cutting & Vishton, 1995). In the previous sections, we have shown that a local estimate that is *proportional* to true curvature can be derived from each signal. These estimates are represented as random variables, say $v_c(x,y)$ for curvature from velocity and $t_c(x,y)$ for curvature from texture. Domini et al. (2006) proposed that the visual system combines these estimates in an optimal weighted sum

$$r_c(x,y) = w_v v_c(x,y) + w_t t_c(x,y), \tag{6}$$

where the weights are

$$w_s \propto \frac{\mathscr{E}(s_c)}{\sigma_{sc}^2} \tag{7}$$

for $s = v,t$. Such combination is optimal in the sense that it maximizes the Signal-to-Noise Ratio (SNR) of the combined estimate $r_c(x,y)$.

The SNR of the optimal combination of Eq. (6) is termed $\rho_c(x,y)$:

$$\rho_c(x,y) = \frac{\mathscr{E}[r_c(x,y)]}{\sqrt{\mathscr{V}[r_c(x,y)]}}. \tag{8}$$

It can be shown the SNR of the combined estimate obtained from the two signals is given by:

$$\rho_c(x,y) = \sqrt{\mathrm{SNR}_{vc}^2(x,y) + \mathrm{SNR}_{tc}^2(x,y)}. \tag{9}$$

For multiple signals s_i,

$$\rho_c(x,y) = \sqrt{\Sigma_i \mathrm{SNR}_{s_i c}^2(x,y)}. \tag{10}$$

If only one-cue is present, for example velocity, then $\rho_c = \mathrm{SNR}_{vc}$.

Eq. (10) describes what has been termed the *Intrinsic Constraint* (IC) model. For a discussion of the similarities and the differences between IC and the modified weak fusion (MWF) model (Landy, Maloney, Johnston, & Young, 1995), see Domini and Caudek (2009), MacKenzie, Murray, and Wilcox (2008).[1]

1.2. Metric judgements

We have shown how the information provided by multiple image signals (velocity, texture, shading, etc.) can be combined to obtain the most precise estimate of a local affine property-map (e.g., curvature). But we have not said anything about how *metric* judgments can be obtained. Domini et al. (2006) hypothesized that the (metric) curvature c' perceived at the point (x,y) is a monotonic function of $\rho_c(x,y)$:

$$c'(x,y) = f[\rho_c(x,y)]. \tag{11}$$

Domini and Caudek (in press) speculated that perceived 3D shape is found through heuristic procedures that depend both on the function f of Eq. (11) and on the context in which each local surface patch is embedded. When the viewing conditions are kept constant, therefore, the perceptual estimates of curvature (or of any other metric property p) should only depend on $\rho_c(x,y)$.

1.3. Rationale of the present investigation

In two experiments, we tested a basic assumption of the models described in the Introduction. According to "Euclidean" theories, the perceptual estimates of depth, slant, and curvature must be mutually consistent. According to the approach proposed by Domini et al. (2006), instead, these perceptual estimates are inconsistent with each other.

The logic underlying the IC model is summarized in Fig. 2 for the particular case of a velocity-only stimulus. Image operators extract the zero-, first-, and second-order spatial properties of the velocity field (see Eqs. (3)–(5)). v_z, v_s, and v_c are the estimates of the local affine depth (z), slant (s), and curvature (c) maps, respectively. We hypothesize that the SNR of these estimates determines, through a heuristic function, the magnitude of perceived depth, slant, and curvature.

[1] According to MWF, the most widely used model of cue combination, different image signals are independently processed so as to obtain *unbiased* estimates of the *metric* properties of the environmental objects. These unbiased estimates are then combined with weights proportional to the reliability of each estimate. MWF is optimal in the sense that it maximizes the reliability of the *metric* solution. Note that MWF does not attempt to maximize the reliability of the *local affine* solution.

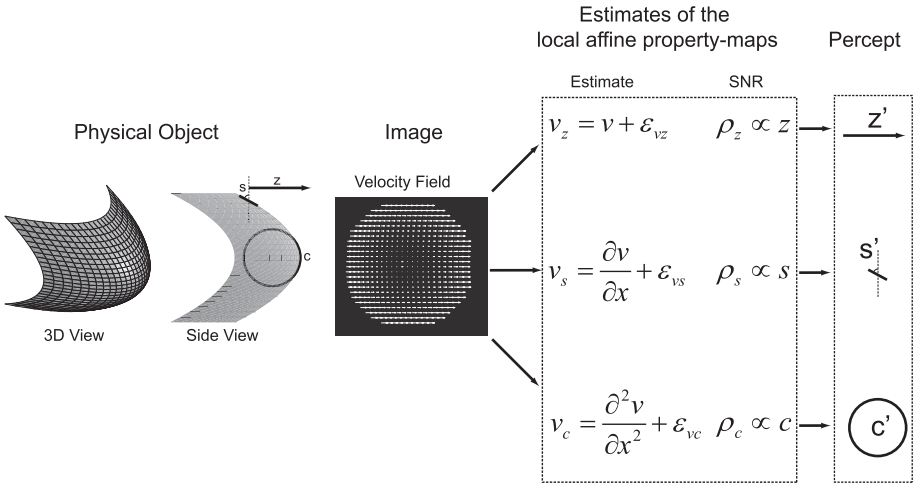

Fig. 2. The 3D rotation of an elliptical surface about the vertical axis generates a retinal velocity field. Image operators extract estimates of the zero-order (v_z), first-order (v_s), and second-order (v_c) spatial properties of the velocity field, which are proportional to depth (z), slant (s), and curvature (c), respectively. The SNRs of these estimates (i.e., ρ_z, ρ_s, and ρ_c) are the IC predictors of the perceived local *metric* properties z', s', and c'.

The purpose of the present investigation was to test whether judgments of depth, slant, and curvature are consistent with each other. Suppose that the perceived (local) curvature of a velocity-only stimulus matches that of a shading-only stimulus. This does not mean that the two stimuli will be perceived as having also the same (local) depth or slant. According to IC, two stimuli are matched in perceived curvature because the local estimates proportional to curvature, which are recovered from the two stimuli, have the same SNR value. In two stimuli, however, the encodings of local depth, slant, and curvature may be characterized by different SNRs. If the perceptual estimates depend on such SNRs, therefore, judgments of depth, slant, and curvature may not be consistent with each other. This hypothesis was tested in Experiment 1 for the case of single-cue stimuli.

Now, consider the case of multiple-cue stimuli (see Fig. 3). According to IC, the visual system recovers from each image signal an estimate that is proportional to the local property p (v_z, v_s, and v_c for velocity, s_z, s_s, and s_c for shading, and so on). The estimates of the same property p (e.g., curvature) are then combined according to Eq. (6). We hypothesize that the perceptual *metric* judgment of the property p is a monotonically increasing function of the SNR of the combined estimate $r_p(x,y)$.

Two predictions follow from these premises. (1) The judgments of the *metric* 3D properties are expected to increase as more image signals are added to a stimulus display. (2) An increase in the number of image signals does not reduce the inconsistency of the estimated depth, slant, and curvature magnitudes. These predictions were tested in Experiment 2.

2. Experiment 1

The experiment comprised two parts: the selection of the appropriate stimulus parameters and the test of the basic assumption of the IC and "Euclidean" models described above (see Table 1).

Selection of stimulus parameters. The stimuli were single-cue renderings of elliptic paraboloid of revolution. Participants were asked to judge which of two sequentially-presented stimuli evoked the largest amount of perceived curvature at the center of the bounding-contour of the stimulus displays (see Fig. 1). One of the stimuli provided texture-only information and remained constant across all trials (*standard*). The other stimulus was rendered by velocity-only or shading-only information and was varied in each trial (*comparison*).

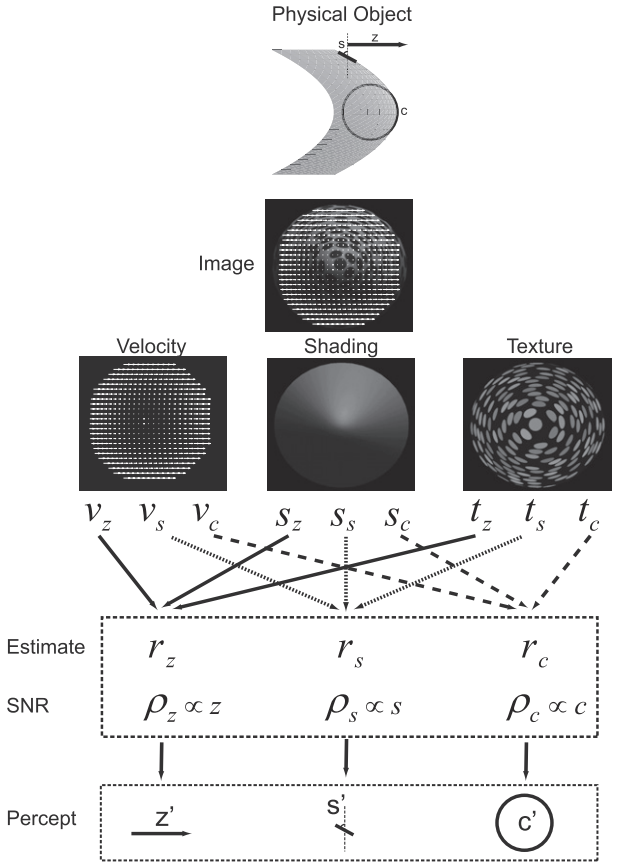

Fig. 3. According to IC, from each image signal [velocity (v), texture (t), or shading (s)] the visual system recovers an estimate that is proportional to local depth (z), slant (s), or curvature (c). The estimates v_z, v_s, and v_c are recovered from velocity; s_z, s_s, and s_c are recovered from shading; t_z, t_s, and t_c are recovered from texture. The visual system then combines all the estimates that are proportional to each specific property p (e.g., curvature). The IC combination rule maximizes the SNR of the recovered local affine property-map. According to IC, the magnitudes of z', s' and c' are determined by the SNRs of the recovered depth-, slant-, and curvature-maps.

According to IC, a curvature match is found when the relevant estimates (ρ_p) recovered from the two stimuli have the same SNR. To obtain a curvature match, we manipulated the SNR of the *comparison* stimulus by varying the relevant scene parameters. For the velocity signal, the SNR is:

Table 1
Design of Experiment 1.

	Standard	Comparison		Independent variable	Match
Part 1	Texture-only	Velocity-only	Shading-only	Scene parameters	Curvature
Part 2	Texture-only	Velocity-only	Shading-only	Elongation	Depth Slant Curvature

$$\mathrm{SNR}_{v_c} = \frac{v_c(x,y)}{\sigma_{vc}} = \frac{\omega}{z_f \sigma_{vc}} c(x,y), \tag{12}$$

where the subscript $_v$ stands for "velocity" and $_c$ stands for "curvature" (see Eq. (5)). The SNR depends on the scene parameters ω and z_f, the distal property c, and the noise σ_{vc}. By varying ω, we changed the SNR by leaving the simulated 3D structure unchanged. In an equivalent manner, the SNR of the shading-only *comparison* stimulus was varied by changing the illuminant direction (see also Caudek, Domini, & Di Luca, 2002; Christou, Koenderink, & van Doorn, 1996; Curran & Johnston, 1994, 1996; Erens, Kappers, & Koenderink, 1993; Johnston & Passmore, 1994a, 1994b; Koenderink, van Doorn, & Kappers, 1992, 1996; Pentland, 1982; Todd, Koenderink, van Doorn, & Kappers, 1996; Todd & Mingolla, 1983). The goal was to generate the renderings of two pairs of elliptic paraboloids of revolution, both having the same simulated elongation, that evoked the same amount of perceived curvature at the target location: a texture-only *standard* and shading-only *comparison*, and a texture-only *standard* and velocity-only *comparison*.

Test of IC and "Euclidean" models. In each trial, participants reported which of two successively-presented stimuli (*standard* or *comparison*) had the largest amount of depth (z), slant (s), or curvature (c) – see Fig. 1. Depending on their response, the elongation of the (shading-only or velocity-only) *comparison* stimulus was varied within a staircase procedure; the elongation of the texture-only *standard* was kept constant. The "scene parameters" (i.e., the angle of rotation for the velocity-only stimulus and the illuminant direction for the shading-only stimulus) were set to the values found in the first part of the experiment.

According to "Euclidean" theories, two geometrically-identical stimuli that are perceived as having the same curvature should also be perceived as having the same slant and depth. For the velocity-only and texture-only pair, we should have that

$$z_v = z_t, \tag{13}$$
$$s_v = s_t, \tag{14}$$
$$c_v = c_t. \tag{15}$$

According to IC, a local depth, slant, or curvature match occurs only if

$$\mathrm{SNR}_{vz} = \frac{\omega_0}{z_f \sigma_{vz}} z_v = \mathrm{SNR}_{tz}, \tag{16}$$
$$\mathrm{SNR}_{vs} = \frac{\omega_0}{z_f \sigma_{vs}} s_v = \mathrm{SNR}_{ts}, \tag{17}$$
$$\mathrm{SNR}_{vc} = \frac{\omega_0}{z_f \sigma_{vc}} c_v = \mathrm{SNR}_{tc}. \tag{18}$$

As the consequence of the manipulation of the first part of the experiment, we know that observers perceive the same amount of curvature at the tip of the two surfaces: $c_v = c_t$. Regardless of this, IC predicts that the two stimuli will not have the same perceived depth or slant. The encodings of curvature, slant, and depth, in fact, may be affected by different levels of noise. The same reasoning can also be applied to the shading-only and texture-only pair.

A test of IC requires the knowledge of the SNRs of Eqs. (16)–(18). Domini and Caudek (2009, 2010) demonstrated that these quantities can be empirically found by the ratios between the Points of Subjective Equality (PSEs) and the Just Noticeable Differences (JNDs) computed from the psychometric functions estimated through the procedure described in the following section.

2.1. Method

2.1.1. Apparatus and stimuli

The stimuli were generated by means of custom software making use of the OpenGL libraries. Stimuli were rendered at 60 fps and the screen update was synchronized with the monitor refresh. Stimuli were presented on a ViewSonic P70f color monitor controlled via a Dell Dimension 8100 with a Nvidia FX9600 graphic card. The resolution of the monitor was 1280×1024 and the refresh rate was 60 Hz.

Participants viewed the monitor monocularly by wearing an eye-patch. With a chin-rest, the viewing distance was kept at 250 cm to minimize the influence of accommodation cues (e.g. Mather, 1997; Watt, Akeley, Ernst, & Banks, 2005). The monitor was contained within a viewing-box; a squared viewing-window measuring 20×20 cm (4°35′ of viewing angle) limited the visible portion of the monitor. When stimuli were not presented, the monitor was occluded so as to reinforce the impression that the stimuli were tangible objects inside a box. To prevent dark adaptation, a dim light source (5 W) was positioned away from the line of sight.

The stimuli simulated the parallel projection of an elliptic paraboloid of revolution:

$$z = S \times 0.66(x^2 + y^2), \tag{19}$$

where S is the stretch parameter, z is the amount of depth simulated behind the monitor's surface, x and y are the screen positions. The visible bounding-contour of the surface was circular, with a radius $r_b = 2.5$ cm (viewing angle 1°8′).

The *standard* and *comparison* quadratic surfaces were rendered by only one visual cue: velocity, texture, or shading.

Velocity: The low-albedo smooth surface was scattered with about 80 high-albedo dots (diameter =0.1 cm). The dots were randomly distributed on the image plane, with a minimum distance of 0.125 cm among their centers. The simulated surface performed one sinusoidal oscillation at 1 Hz around a vertical axis at half of the surface's depth.
Texture: A volumetric texture technique was employed to generate a sculptured representation of the surface. The surface intersected a number of spheres having a diameter of 0.5 cm. These spheres were positioned with their centers on the surface so that they did not intersect each other. Points of the surface within the spheres were simulated with a different albedo than points outside. An anti-aliasing technique was used to create a smooth border between the two regions (see Fig. 7).
Shading: A standard Phong model with Lambertian reflectance function determined the amount of screen luminance at every point (Foley & Van Dam, 1983). Since the object was a simple convex shape, no cast-shadow algorithm was needed. The light source consisted of nine point light-sources at optical infinity arranged in a 3×3 array. The difference in the angle of incident light for the row and columns of the array was 5°. The light source was positioned 15° on the right from the vertical (see Fig. 7).

2.1.2. Participants

A total of ten undergraduate and graduate students from Brown University participated in the experiment. They all had normal or

corrected-to-normal vision. All participants were naïve to the purpose of the study, but most of them were experienced psychophysical participants. All participants were run individually and completed the experiment in three 1 h sessions consisting of two blocks of trials.

2.1.3. Procedure

Selection of stimulus parameters. The *comparison* was a shading-only or a velocity-only stimulus; the *standard* was a texture-only stimulus. For both stimuli, the stretch parameter *S* took on the value of 1.0.

Each trial comprised the successive presentation of a *standard* and a *comparison* stimulus, with the order of presentation randomized. Participants were asked to determine which of the two stimuli evoked the greatest amount of perceived curvature at the tip of the quadratic profile. The 3D rotation speed (for the velocity-only *comparison*) or the illuminant direction (for the shading-only *comparison*) were varied within a staircase procedure until a curvature match was obtained.

Four interleaved staircases (3up-1down, 2up-1down, 1up-2down, 1up-3down) were employed. The staircases started at 60°, 50°, 20°, 10° for the direction of illumination and at 30°, 25°, 10°, and 5° for the amount of rotation. The staircase-step was 5° before the first reversal and 1° afterwards. Each staircase stopped after four reversals. If all the staircases were not completed within the 25 min session, the participant's data were discarded.

The data were fitted with cumulative Gaussians free to vary in position (PSE) and slope (JND) using the software package `psignifit` (Wichmann & Hill, 2001). From the fitted psychometric functions, we determined the point of subjective equality (PSE) and the difference threshold [just-noticeable difference (JND)].[2] The values of these PSEs were then used for generating the stimuli of the main part of the experiment. Different values for the angle of rotation and for the illuminant direction were used for different participants, according to their PSEs.

Test of IC and "Euclidean" models. Participants reported which of two successively-presented stimuli had the greatest amount of curvature, slant, or depth at the target location, depending on the icon presented at the beginning of the trial (see Fig. 1). Each trial began with the presentation of an icon indicating the requested judgment (depth, orientation, or curvature). After 300 ms, the icon disappeared and after 500 ms the *standard* and *comparison* stimuli were sequentially presented in random order at different monitor positions. Each stimulus was shown for 1000 ms. Participants used the mouse buttons to provide their response. The next trial started 500 ms after each key-press. A short practice session was provided at the beginning of each block of trials.

In all trials, the stretch parameter of the texture-only *standard* stimulus was kept constant: $S = 1.0$. The stretch parameter of the velocity-only or shading-only *comparison* stimulus was manipulated within a staircase procedure. The PSEs for the stretch parameter *S* of the *comparison* stimuli were determined by means of four interleaved staircases. The staircases started at 2.0, 1.5, 1.0, and 0.5. The staircase-step was 0.125. Participants performed about 600 trials during the experiment.

[2] The PSE corresponds to the angle of rotation or the illuminant direction at which the psychometric function reaches 0.5. That is, the PSE is the point at which the texture-only and the velocity-only surfaces, or the texture-only and the shading-only surfaces, were equally often judged as having the same curvature at at the tip of their elliptic profile (see Fig. 1). The JND is defined as the difference between the PSE and the 0.84 point on the psychometric function (= the angle of rotation or the illuminant direction for which the *comparison* stimulus is judged to be more curved 84% of the time). The 0.84 point is usually used to define the JND because then the JND corresponds to the standard deviation σ of the cumulative Gaussian fitted to the data (e.g., Helbig & Ernst, 2007).

2.1.4. Design

A 2 × 3 within-subject design was used (see Table 1), with two *cue*-conditions (motion or shading *comparison*-stimuli) and three *task*-conditions (curvature, slant, depth judgments).

2.2. Results and discussion

Selection of stimulus parameters. Across participants, the average PSE for the velocity-only *comparison* stimuli was equal to 14.5 ± 4.3° (s.e.); the average PSE for shading-only *comparison* stimuli was equal to 24.9 ± 4.1° (s.e.). Average JNDs for the angle of rotation and for the angle of illumination were 7.98 ± 1.81° (s.e.) and 10.24 ± 2.63° (s.e.), respectively. These results indicate that the manipulation of the angle of rotation and of the illuminant direction affected the perceived curvature of the *comparison* stimulus (see Appendix).

Test of IC and "Euclidean" models. Fig. 4 shows the results for the six experimental conditions. On the *y*-axis of each panel is shown the average depth, slant, and curvature of the *comparison* stimulus at the PSE. The dashed lines represent the depth, slant, and curvature values simulated for the *standard* stimulus. Values above the dashed lines indicate that, at the PSE, the *comparison* stimulus had larger amounts of simulated curvature, slant, and depth than the *standard*. The average JNDs corresponding to the six PSEs reported in the Figure are shown in Table 2. Note that, as the JNDs increased, a larger value of the stretch parameter was needed for the *comparison* stimulus in order to obtain a perceptual match.

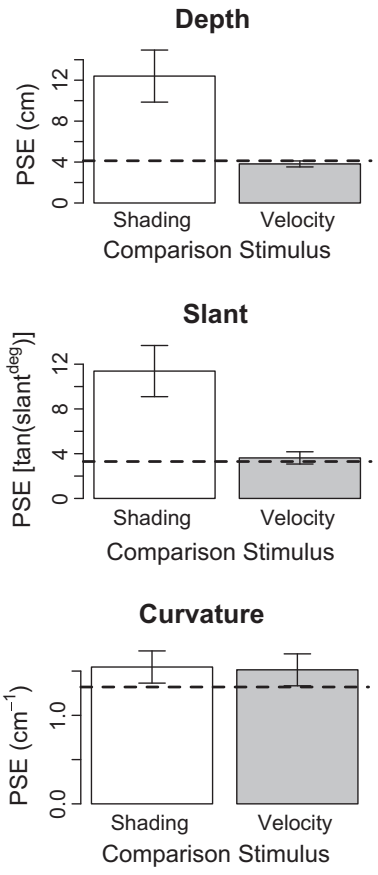

Fig. 4. Experiment 1. Average PSE values of the shading-only and velocity-only *comparison* stimuli computed from the psychometric functions for the depth-, slant-, and curvature-matching tasks. The dashed lines represent the curvature, slant, and depth values simulated for the *standard* stimulus. Error bars represent one standard error of the mean.

Table 2
JNDs for Part 2 of Experiment 1.

	Comparison stimulus	
	Velocity-only	Shading-only
Curvature	0.244	0.144
Slant	0.272	0.595
Depth	0.175	0.580

Inferential statistics on the observers' responses are based on a Linear Mixed-Effects (LME) model specifying participants as a random factor to control for their associated intraclass correlation.[3] A 2 (cue of the *comparison* stimulus: velocity or shading) × 3 (judgments of depth, slant, or curvature) LME model revealed a significant 2-way interaction, $\chi^2_2 = 13.594$, $p = .001$. If judgments of perceived curvature, slant, and depth were consistent with each other (this is the basic assumption of the "Euclidean" theories), once *standard* and *comparison* stimuli had been matched in terms of perceived curvature, they should also be matched in terms of slant and depth. The three data points in Fig. 5 (left panel) should therefore coincide with the point (1, 1). For curvature-matched stimuli, instead, we found that the average PSE was larger for the shading-only than for the velocity-only *comparison* stimuli for the slant judgments, $t_{55} = 4.896$, $p = .001$. Moreover, the average PSE was larger for the shading-only than for velocity-only *comparison* stimuli for the depth judgments, $t_{55} = 4.333$, $p = .001$. As a control, we also examined the PSEs for the curvature judgments. As expected, the average PSEs for these judgments did not differ across velocity-only and shading-only *comparison* stimuli, $t_{55} = 0.050$, $p = .967$.

A more direct test of IC can be performed according to Eqs. (16)–(18) by estimating the SNRs with the PSE/JND ratios (see Domini & Caudek, 2009, 2010). Here, we computed the reciprocal of the SNRs because they are less sensitive to the variability of the JND estimates. Note that the ratio JND/PSE may be more easily comprehended as a coefficient of variation, defined on the standard deviation and mean. The coefficient of variation is proportional to the Weber fraction (e.g., Bizo, Chu, Sanabria, & Killeen, 2006). Fig. 5 (right panel) shows the average 1/SNR of the shading-only *comparison* stimuli as a function of the velocity-only *comparison* stimuli. For a perfect correspondence between observed data and predictions, the three points on the graph should lie on the 45° dashed line. When considering the 95% confidence intervals, we can conclude that the observed data and the predictions of IC are very closely matched, indeed.

3. Experiment 2

In Experiment 2, we tested whether the results of Experiment 1 could be replicated with a larger number of depth cues. The experimental design was similar to Experiment 1, with two main differences: (a) the *comparison* stimulus had three cues, whereas the *standard* had either one or two cues, and (b) the *standard* stimulus took on two different elongations (see Table 3).

According to the "Euclidean" theories, improvement in depth constancy is expected if multiple, mutually consistent depth-cues are provided (e.g., Hoffman, Girshick, Akeley, & Banks, 2008; Landy et al., 1995; Richards, 1985; Watt et al., 2005).

[3] We used the `lmer` program (`lme4` package) in the `R` system for statistical computing (R Development Core Team, 2010). As indicated by Baayen (2008), *p* values and confidence intervals are generated from the posterior distribution of parameter estimates with Markov Chain Monte Carlo methods, using the `mcmcsamp` program in the `lme4` package with default specifications (e.g., $n = 1000$ samples; locally uniform priors for fixed effects; locally non-informative priors for random effects).

According to IC, instead, a larger number of depth-cues does not guarantee better (more veridical) performance. The inconsistency of metric judgements can persist (or even worsen) for multiple-cue stimuli, because the local estimates of the affine property maps do not carry metric information. With more image signals, instead, the estimates of the metric properties should take on larger values, because ρ_p (i.e., the SNR of the combined estimate of the local affine structure) increases.

With the visual cues used in Experiment 2, the IC model can be tested as follows. Let us consider the curvature-matching task – the same considerations also apply to the slant- and depth-matching tasks. According to IC, a perceptual match occurs when the same value ρ_p is generated by the three-cue *comparison* and by the one-cue or two-cue *standard* stimuli. For the single-cue *standard* stimuli, we can write:

$$SNR_{vc} = \frac{k_{vc}}{\sigma_{vc}} c_0, \tag{20}$$

$$SNR_{tc} = \frac{k_{tc}}{\sigma_{tc}} c_0, \tag{21}$$

$$SNR_{sc} = \frac{k_{sc}}{\sigma_{sc}} c_0, \tag{22}$$

where c_0 denotes the amount of curvature simulated in the experiment.

The SNR of a combined two-cue stimulus is given by Eq. (9). By substituting the above SNRs in Eq. (9), we obtain:

$$\rho_{vt} = \sqrt{\left(\frac{k_{vc}}{\sigma_{vc}}\right)^2 + \left(\frac{k_{tc}}{\sigma_{tc}}\right)^2} c_0, \tag{23}$$

$$\rho_{vs} = \sqrt{\left(\frac{k_{vc}}{\sigma_{vc}}\right)^2 + \left(\frac{k_{sc}}{\sigma_{sc}}\right)^2} c_0, \tag{24}$$

$$\rho_{ts} = \sqrt{\left(\frac{k_{tc}}{\sigma_{tc}}\right)^2 + \left(\frac{k_{sc}}{\sigma_{sc}}\right)^2} c_0. \tag{25}$$

For the three-cue *comparison* stimulus, the SNR becomes:

$$\rho_{vts} = \sqrt{\left(\frac{k_{vc}}{\sigma_{vc}}\right)^2 + \left(\frac{k_{tc}}{\sigma_{tc}}\right)^2 + \left(\frac{k_{sc}}{\sigma_{sc}}\right)^2} c, \tag{26}$$

where c is the amount of simulated curvature that is varied through the staircase procedure.

Let c_a, c_b, and c_{ab} (with $a,b \in \{v,t,s\}$ and $a \neq b$) be the curvature at the PSE of the *comparison* stimulus matched with a single-cue or a two-cue *standard*, respectively. According to IC, there is a specific relation between c_{ab}, on the one side, and c_a and c_b, on the other.

For example, at the PSE, the perceptual matches of the velocity-only and texture-only *standards* with the three-cue *comparison* will be:

$$\sqrt{\left(\frac{k_{vc}}{\sigma_{vc}}\right)^2 + \left(\frac{k_{tc}}{\sigma_{tc}}\right)^2 + \left(\frac{k_{sc}}{\sigma_{sc}}\right)^2} c_v = \frac{k_{vc}}{\sigma_{vc}} c_0, \tag{27}$$

$$\sqrt{\left(\frac{k_{vc}}{\sigma_{vc}}\right)^2 + \left(\frac{k_{tc}}{\sigma_{tc}}\right)^2 + \left(\frac{k_{sc}}{\sigma_{sc}}\right)^2} c_t = \frac{k_{tc}}{\sigma_{tc}} c_0. \tag{28}$$

At the PSE, the perceptual match of the velocity-texture *standard* with the three-cue *comparison* will be:

$$\sqrt{\left(\frac{k_{vc}}{\sigma_{vc}}\right)^2 + \left(\frac{k_{tc}}{\sigma_{tc}}\right)^2 + \left(\frac{k_{sc}}{\sigma_{sc}}\right)^2} c_{vt} = \sqrt{\left(\frac{k_{vc}}{\sigma_{vc}}\right)^2 + \left(\frac{k_{tc}}{\sigma_{tc}}\right)^2} c_0. \tag{29}$$

From Eqs. (27)–(29), it follows that

$$c_{vt} = \sqrt{c_v^2 + c_t^2}. \tag{30}$$

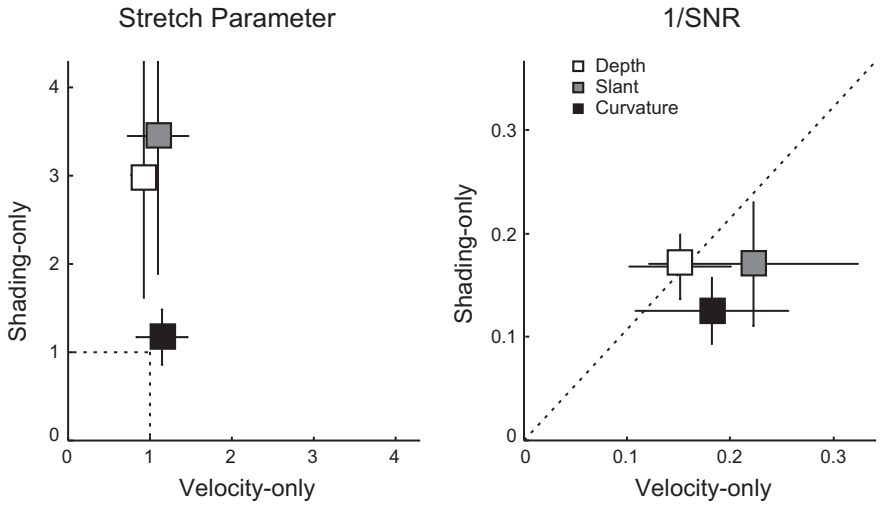

Fig. 5. Experiment 1. Left: Average PSE of the shading-only *comparison* stimulus as a function of the average PSE of the velocity-only *comparison* stimulus. Right: Average estimated 1/SNR of the shading-only *comparison* stimulus as a function of the average estimated 1/SNR of the velocity-only *comparison* stimulus. Judgments of local depth, slant, and curvature are coded by color (white, gray, and black, respectively). Error bars represent 95% confidence intervals.

Table 3
Design of Experiment 2.

	Standard	Comparison		Independent variable	Judgment
Part 1	Texture-only	Velocity-only	Shading-only	Scene parameters	Curvature
Part 2	One-cue Two-cue	Three-cues Three-cues		Elongation	Curvature Slant Depth

The same relationship holds for any pair of cues and any local property:

$$p_{c1c2} = \sqrt{p_{c1}^2 + p_{c2}^2}, \tag{31}$$

where p denotes the magnitude of the metric 3D property (e.g., z, s, c), and the subscripts $_{c1}$, $_{c2}$ denote the two cues.

In conclusion, if a perceptual match depends on the SNR, as hypothesized by IC, then Eq. (31) describes the relation between the PSE of the *comparison* stimulus matched with a two-cue *standard* and the PSEs of the *comparison* stimuli matched with the one-cue *standard* stimuli.[4]

3.1. Method

3.1.1. Participants

Four undergraduate students, one graduate student, and the first author participated in the experiment. All participants completed the experiment in six 1 h sessions.

3.1.2. Procedure and stimuli

Selection of stimulus parameters. The procedure was similar to Experiment 1. Participants matched the perceived curvature of a texture-only *standard* stimulus with the perceived curvature of a motion-only or shading-only *comparison*. The simulated elliptic paraboloid of revolution was $z = S \times 0.495(x^2 + y^2)$.

Test of IC and "Euclidean" hypotheses. The *comparison* stimulus was rendered by three cues (motion, texture, and shading); the *standard* stimulus was rendered either by one (shading, motion, or motion) or by two cues (shading-motion, texture-motion, or texture-shading). In this way, *standard* and *comparison* stimuli were always defined by different sets of cues.[5]

The stimuli were similar to those of Experiment 1, except for the presence of small differences in simulated albedo, which allowed reliable motion signals. To this purpose, small dots were added to the stimuli providing shading information. Surfaces rendered by texture information did not contain these additional dots. For consistency, the dots were present also with stationary shading-only stimuli.

The stretch parameter S of the *comparison* stimulus was varied across trials within a staircase procedure. For each experimental condition, a 1-up-1-down staircase procedure was used to find the stretch parameter required for perceiving the *comparison* stimulus with the same curvature, slant, or depth as the *standard* (see Fig. 1). In different sessions, the staircase started with the stretch parameter of the *comparison* stimulus set to a value either 0.5 times higher or 1.8. The number of trials per participant ranged between 2200 and 3400. The *standard* stimuli provided one or two cues, with a stretch parameter that was either 1.00 or 1.66.

3.1.3. Design

There were 36 experimental conditions within-observer (see Table 3): six *cue* conditions (texture-only, motion-only, shading-only, texture-motion, texture-shading, motion-shading *standard* stimuli) crossed with two *stretch-parameter* conditions ($S = 1.00$ or 1.66), and three *task* conditions (judgment of depth, orientation, or curvature). For each condition, each observer run one 1-up-1-down staircase.

[4] In the above discussion, we obviously assume that the stretch parameter S of the *standard* stimulus remains constant.

[5] The two elliptic paraboloids of revolution were rendered by different cues or by different combinations of cues in order to prevent participants to perform the task by simply matching the 2D features of the stimulus displays.

3.2. Results and discussion

Selection of stimulus parameters. As in Experiment 1, the manipulation of the viewing parameters (i.e., the angle of rotation and the illuminant direction) affected the amount of perceived curvature. The PSEs computed for each participant were used to choose the angle of rotation and the illuminant direction for the second part of the experiment.

Test of IC and "Euclidean" hypotheses. The left panel of Fig. 6 shows the average PSEs computed for the 36 experimental conditions of Experiment 2. The average amounts of depth, slant, or curvature of the *comparison* stimuli at the PSE are shown on the *y*-axis; the *x*-axis indicates the visual cues comprising the *standard* stimuli.

A 6 (cue of *standard* stimuli) × 3 (surface property) × 2 (stretch parameter) LME analysis revealed that a model with a 3-way interaction did not differ significantly from a model with the three main effects and a Cue × Property interaction, $\chi^2_{17} = 25.776,\ p = .079$. The 2-way interaction between *cue* and *property* was significant, $\chi^2_{10} = 30.179,\ p = .001$. This result replicates what found in Experiment 1: two geometrically-identical stimuli that are perceptually matched in terms of curvature are not matched in terms of perceived slant or depth.

The test of IC performed according to Eq. (31) is shown on the right panel of Fig. 6. The *x* and *y* axes report the IC's predictions and the observed PSEs, respectively. A perfect match between the model's predictions and the data is given by the 45° line. The LME regression of the observed PSEs on the model's predictions produces an intercept of 0.164 not significantly different from zero, $p = .308$, 95% *CI*: [-0.181, 0.504], and a slope of 1.086, not significantly different from 1.0, 95% *CI*: [0.985, 1.193]. An alternative ("Euclidean") LME model was run, with the stretch parameter of the *standard* stimulus as the only fixed effect. The first model, with the IC's predictions and the individual differences, explained 84% of the variance of the PSEs. The second model, with the stretch parameter and the individual differences, explained only the 17% of the variance of the PSEs. We can thus conclude that the predictions of IC are in very close agreement with the empirical data. It is important to stress that the IC predictions were formulated with no free parameters.

4. General discussion

In two experiments, we investigated the consistency of perceptual judgments of local depth, slant, and curvature. IC hypothesizes that (a) the noise levels affecting the encoding of a surface property (e.g., curvature) vary with retinal image location, (b) different levels of noise affect the encoding of a surface property in different image signals, or combinations of image signals, and (c) the SNR of the local encoding of a surface property determines the perceptual *metric* estimate of that property (see Fig. 3). From these hypotheses it follows that judgments of depth, slant, and curvature should not be consistent with each other.

An example of this is provided in Fig. 7, where two elliptic paraboloids of revolution having the same elongation are rendered by texture information (left panel) and shading information (right panel). Typically, the curvature at the center of the display is perceived to be larger for the shading stimulus. For a local curvature-match, therefore, we should reduce the elongation of the shading stimulus. In Fig. 7, however, observers perceive a larger relative depth between the tip of the paraboloid and the bounding-contour for the texture than for the shading rendering. For a depth-match, therefore, we should reduce the elongation of the texture stimulus. Finally, near the boundary of the display, local slant appears to be shallower for the shading stimulus. For a local slant-match, therefore, we should increase the elongation of the shading stimulus. This example shows that the perceptual judgments of depth, slant, and curvature are mutually inconsistent. Therefore, they cannot be described by an affine mapping between the environmental objects and the perceived 3D shape.

Participants varied the elongation of a elliptic paraboloid of revolution in a two-interval forced-choice (2IFC) task until it was perceptually matched to another elliptic paraboloid in terms of depth, slant, or curvature. The two surfaces were rendered by different cues. From the results of this task, we can conclude that perceptual judgements are inconsistent if (a) different elongations of the two surfaces are found at the PSE, and (b) these elongations depend on the surface's property being judged (depth, slant, curvature).

For selecting the appropriate stimulus parameters, in Experiment 1 the angular velocity was varied within a staircase procedure until the perceived *curvature* of a velocity-only paraboloid matched that of a texture-only paraboloid. The same was done for the illuminant direction of a shading-only paraboloid. All three stimuli (velocity-only, shading-only, texture-only) had the same simulated elongation. The angular velocity and the illuminant direction at the PSE were then used to generate the stimuli of the main part of the Experiment.

In the main part of Experiment 1, participants were asked to compare the shading-only and texture-only paraboloids in terms of perceived *depth*, *slant*, or *curvature*. The same was done for the velocity-only and texture-only paraboloids. The elongation of the velocity-only or shading-only paraboloid was varied within a staircase procedure until a perceptual match was found. For depth and slant judgments, we found that the shading-only paraboloid had a larger elongation at the PSE than the texture-only paraboloid; for the curvature judgments, instead, the two paraboloids had the same elongation at the PSE. In other words, it was not possible to find a depth elongation for which the shading-only paraboloid was perceived as having the same depth, slant, and curvature as the texture-only paraboloid. Such results violate the fundamental assumption of the "Euclidean" theories, because it implies that the estimated depth-, slant, and curvature-maps were not consistent with each other.

In Experiment 2, the simulated elongation of a paraboloid rendered by three cues (velocity, shading, and texture) was varied until it was perceptually matched in depth, slant, or curvature to a fixed paraboloid rendered by either one (velocity, shading, or texture) or two cues (texture-motion, texture-shading, or motion-shading). Therefore, the stimulus displays were richer than in Experiment 1. None-withstanding this difference, different elongations of the three-cue paraboloid were found at the PSE for judgments of depth, slant, or curvature. Moreover, larger elongations were found at the PSE for two-cue (rather than one-cue) *standard* stimuli.

Note that "Euclidean" models cannot provide *a priori* quantitative accounts of these systematic biases, nor they can explain why the estimates of different metric properties (such as depth, slant, curvature) are inconsistent with each other. A quantitative account of the present results, instead, can be provided by the process exemplified in Fig. 3. Figs. 5 and 6 show that the *quantitative* predictions of IC are in good agreement with the data of both Experiments 1 and 2. Similar results were found by Shah, Domini, and Caudek (submitted for publication) for depth judgments of natural (as opposed to virtual) stimuli defined by velocity and/or disparity information.

IC is based on the analysis of the local affine structure. We do not propose, however, that perceived global 3D shape should be described in terms of an affine transformation of the environmental object (see Fig. 7). At the contrary, in our previous work we have advocated the opposite view. For example, we found that

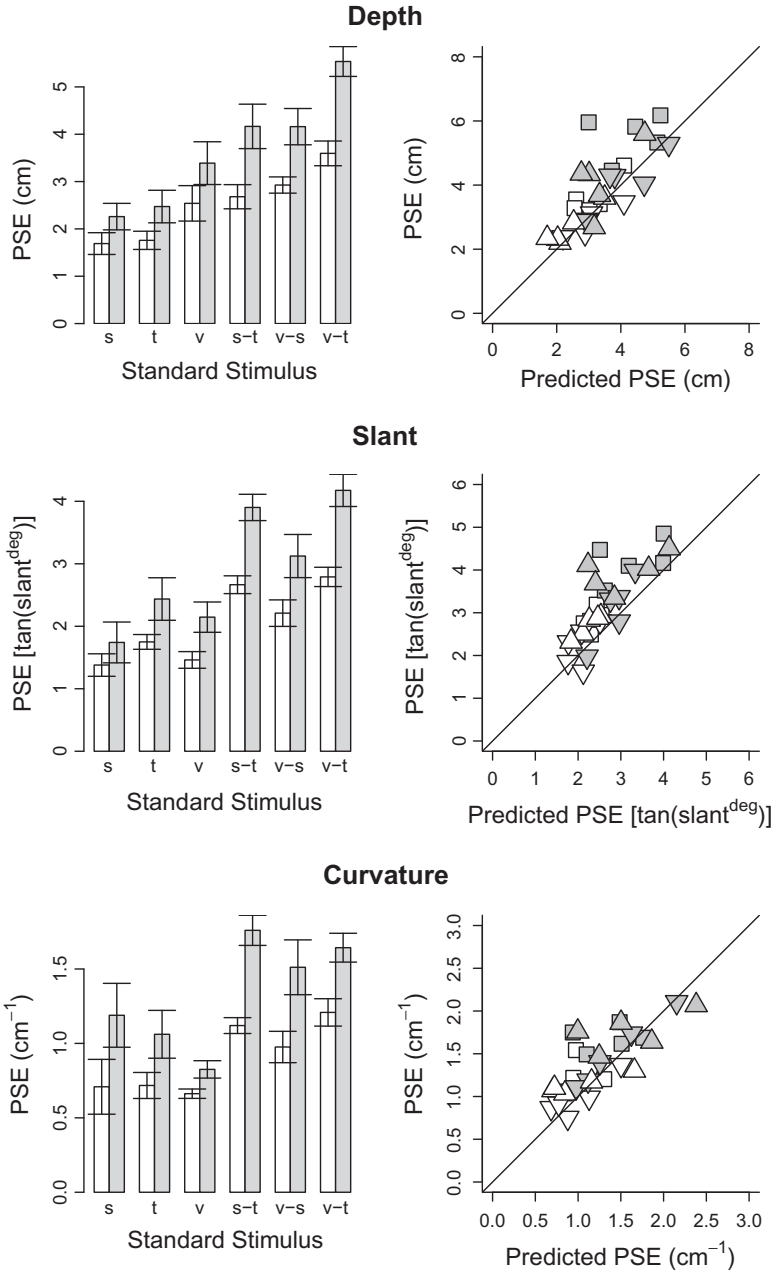

Fig. 6. Experiment 2. The three rows show the results of the depth-, slant-, and curvature-matching tasks (top, center, and bottom, respectively). The stretch parameter S of the *standard* stimulus is coded by color: 1.0: white; 1.66: gray. Left: The y-axis of each panel shows the average PSE value of the (variable) three-cue *comparison* stimulus. The x-axis indicates the image signals of the (fixed) *standard* stimulus to which the *comparison* was perceptually matched. The shading, texture, and velocity signals are denoted by "s" , "t", and "v", respectively. A single letter denotes a one-cue *standard*; two hyphenated letters denote a two-cue *standard*. Error bars represent one standard error of the mean. Right: The y-axis of each panel shows the observed PSE of the *comparison* stimulus matched to a two-cue *standard* (e.g., "s–t"); the x-axis shows the predictions of the IC model computed by means of the PSEs of the *comparison* stimuli matched to the corresponding one-cue *standard* stimuli (e.g., "s" and "t") – see Eq. (31). Each point corresponds to one observer. The symbols in the plots are coded as follows: ■ = s–t; ▼ = v–s; ▲ = v–t.

perception of global Structure-from-Motion is neither Euclidean nor Affine (Domini & Braunstein, 1998; Domini et al., 1998), in spite of the fact that first-order motion information is ambiguous only up to a one-parameter family of solutions. For other cues, like disparities or texture, the perceptual distortions are well documented to be non-affine (Todd, Thaler, Dijkstra, Koenderink, & Kappers, 2007). We can safely assume, however, that within a local neighborhood the perceptual map is affine. This requires the existence of mathematical transformations of the local image signals that are linearly related to the properties of the environmental object (depth, slant, or curvature).

In several investigations, IC has provided successful accounts of perceptual performance in local affine and metric tasks, for both single-cue and combined-cue stimuli (Di Luca, Domini, & Caudek, 2007; Domini & Caudek, 2009; Domini & Caudek, 2010; Domini et al., 2006; Tassinari, Domini, & Caudek, 2008). To account for the psychophysical data, IC makes use of heuristic procedures that generate biased metric estimates of 3D properties. A natural

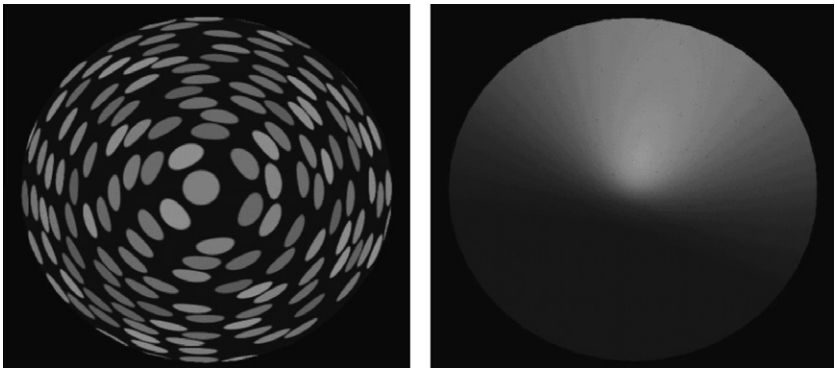

Fig. 7. Texture-only and shading-only rendering of the elliptical paraboloid of revolution defined by Eq. (1) with $C = 0.33$. This is an example of the stimuli that were actually used in Experiment 1.

question then arises: How is it possible to reconcile the predictions of IC (and the empirical data coming with it) with the obvious fact that we successfully interact with our visual environment (apparently without any systematic bias)?

A possible answer to such question is that visually-guided behavior may not require an internal representation of external space but, rather, it may be based on more concrete representations directly linked with the sensory channels. The input patterns may be mapped in a task-dependent way into output patterns without requiring the full specification of the metric depth-map of the spatial layout (Bradshaw et al., 2000; Gårding, Porill, Mayhew, & Frisby, 1995; Glennerster et al., 1996; Morasso & Sanguineti, 1997; Todd, 2004).

Consistently with such point of view, we hypothesize that visually-guided behavior is mainly driven by *direct information* about the external 3D space (see also Thaler & Goodale, 2010). Such "weaker" non-metric representation has the advantage of being more constrained from a mathematical point of view than the computation of metric structure. Therefore, it is more robust towards sensor noise (Robert, Zeller, Faugeras, & Hebert, 1997). Moreover, non-metric representations require a lower computational cost (Beardsley, Reid, Zisserman, & Murray, 1995; Faugeras, 1995; Koenderink & van Doorn, 1991; Todd, 2004). All these advantages make IC better suited for a biological system than an "Euclidean" analysis.

5. Conclusions

We found that local judgements of depth, slant, and curvature of smoothly-curved surfaces rendered by single-cue or multiple-cue stimuli are not congruent with each other. Such findings are inconsistent with "Euclidean" models of 3D shape recovery, but they can be explained by the IC model (Domini et al., 2006).

Acknowledgments

We thank two anonymous reviewers for their constructive comments on an earlier version of this article. We are grateful to Daniel Rothman and William Warren for their help and support. Massimiliano Di Luca is funded by EU project IST-2006-027141 "Immersence". This project was founded by NSF Award No. BCS 0643234.

Appendix A

We add here some important considerations concerning the results of the first part of Experiment 1. For motion-only stimuli, we found that perceived curvature increased as angular rotation increased (Domini et al., 1997; Liter, Braunstein, & Hoffman, 1993; Todd & Norman, 1991; Todd & Perotti, 1999). For shading-only stimuli, perceived curvature increased as the illuminant direction was shifted away from the viewing direction (Caudek et al., 2002; Curran & Johnston, 1994, 1996; Johnston & Passmore, 1994a, 1994b). These results are compatible with IC, because the variation of the scene parameters affects the measurement noise of the local image primitives. They are also consistent with the "Euclidean" models, because different scene parameters may induce different biases in the estimation of the "missing parameters" (for example, the parameter ω in Eq. (3)).

The present data could be interpreted by saying that the average JNDs of these judgments were very large relative to the average PSEs – about 55% and 41% for the rotation and illumination angles, respectively, based on the ratios of the average JNDs and PSEs. These seem to be large Weber fractions, which may indicate that (a) the rotation angle and illumination directions were poor determinants of the perceived depth elongations, and/or (b) simple shapes specified by texture cannot be reliably matched with those specified by motion or shading.

We should consider, however, that the precision of participants' judgments can be expressed either in terms of the image properties or in terms of the parameters of the 3D scene. If judgments are coded in terms of the parameters of the 3D scene (i.e., illuminant direction, 3D rotation speed), the present data indicate that the precision of participants' judgments was indeed poor. This does not mean, however, that the precision of participants' judgments is low when it is evaluated with respect to the direct information about 3D shape provided by the retinal images.

For the shading-only stimuli, we can compare our data to those of Pentland (1982). Pentland reports an indirect estimate of the discrimination threshold for the illuminant direction of about 12°. Our JND of 10.23° is similar to the value reported by Pentland, thus indicating that our participants did not under-performed in this discrimination task.

A similar consideration can be made about the JND of 7.98° for the 3D rotation speed. Previous investigations indicate that the perceptual estimate of 3D rotation, further from being veridical, is a positive function of the deformation component of the optic flow (e.g., Domini et al., 1997). As a consequence, it is misleading to think that a JND of 7.98° is large. First, we should describe the participants' task as the discrimination of velocity gradients, rather than 3D angular speeds. Second, we should keep in mind that the same first-order velocity field can be produced by a small rotation and a large surface slant, or by a large rotation and a small surface slant (e.g., Koenderink, 1986). For a constant discrimination performance, therefore, the JND can take on different magnitudes

expressed in terms of 3D angular rotation, depending on the slant of the distal surface.

References

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge University Press.

Beardsley, P. A., Reid, I. D., Zisserman, A., & Murray, D. W. (1995). Active visual navigation using non-metric structure. In *Proceedings of the 5th international conference on computer vision* (pp. 58–65). Boston: IEEE Computer Society Press.

Bizo, L. A., Chu, J. Y. M., Sanabria, F., & Killeen, P. R. (2006). The failure of Weber's law in time perception and production. *Behavioural Processes, 71*, 201–210.

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

Brenner, E., & van Damme, W. (1999). Perceived distance, shape and size. *Vision Research, 39*, 975–986.

Bülthoff, H. H., & Yuille, A. (1990). Shape-from-X: Psychophysics and computation. *Proceedings of the SPIE, 1383*, 165–172.

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.

Christou, C., Koenderink, J. J., & van Doorn, A. J. (1996). Surface gradients, contours and the perception of surface attitude in images of complex scenes. *Perception, 25*, 701–713.

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

Curran, W., & Johnston, A. (1994). The effect of light source position on perceived curvature. *Investigative Ophthalmology and Visual Science, 35/4*, 1741.

Curran, W., & Johnston, A. (1996). The effect of illuminant position on perceived curvature. *Vision Research, 36*, 1399–1410.

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* (2nd ed.. In W. Epstein & S. Rogers (Eds.). *Handbook of perception and cognition* (Vol. XIX, pp. 69–117). San Diego, CA, US: Academic Press, Inc.

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

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

Di Luca, M., Domini, F., & Caudek, C. (2007). The relation between disparity and velocity signals of rigidly moving objects constrains depth order perception. *Vision Research, 47*, 1335–1349.

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

Domini, F., & Braunstein, M. L. (1998). Recovery of 3d structure from motion is neither euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1273–1295.

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). Recovering 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. (2009). The intrinsic constraint model and Fechnerian sensory scaling. *Journal of Vision, 9*(2), 1–15.

Domini, F., & Caudek, C. (2010). Matching perceived depth from disparity and from velocity: Modeling and psychophysics. *Acta Psychologica, 133*, 81–89.

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

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.

Erens, R. G., Kappers, A. M., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics, 54*, 145–156.

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

Fantoni, C., Caudek, C., & Domini, F. (2010). Systematic distortions of perceived planar surface motion in active vision. *Journal of Vision*, *10*(5):12, 1–20.

Faugeras, O. (1995). Stratification of three-dimensional vision: Projective, affine, and metric representations. *Journal of the Optical Society of America A, 12*, 465–484.

Foley, J. D., & Van Dam, A. (1983). *Fundamentals of interactive computer graphics*. Reading, MA: Addison Wesley.

Gårding, J., Porill, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research, 35*, 703–722.

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

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.

Helbig, H. B., & Ernst, M. O. (2007). Knowledge about a common source can promote visual-haptic integration. *Perception, 36*, 1523–1533.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision, 8*(3), 1–30.

Horn, B. K. P. (1986). *Robot vision*. New York: McGraw Hill.

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

Johnston, A., & Passmore, P. J. (1994a). Shape from shading. i: Surface curvature and orientation. *Perception, 23*, 169–189.

Johnston, A., & Passmore, P. J. (1994b). Shape from shading. ii: Geodesic bisection and alignment. *Perception, 23*, 191–200.

Kersten, D., Mamassian, P., & Yuille, A. L. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271–304.

Koenderink, J. J. (1986). Optic flow. *Vision Research, 26*, 161–179.

Koenderink, J. J. (1990). *Solid shape*. Cambridge, Massachusetts, London, England: MIT Press.

Koenderink, J. J., & van Doorn, A. J. (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. L. (1992). Surface perception in pictures. *Perception & Psychophysics, 52*, 487–496.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1996). Pictorial surface attitude and local depth comparison. *Perception & Psychophysics, 58*, 163–173.

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.

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception, 22*, 441–1465.

MacKenzie, K. J., Murray, R. F., & Wilcox, L. M. (2008). The intrinsic constraint approach to cue combination: An empirical and theoretical evaluation. *Journal of Vision, 8*(8), 1–10.

Mather, G. (1997). The use of image blur as a depth cue. *Perception, 26*, 1147–1158.

Morasso, P., & Sanguineti, V. (Eds.). (1997). *Self-organization, computational maps and motor control*. Elsevier.

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

R Development Core Team (2010). *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0. Available from: http://www.R-project.org.

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

Robert, L., Zeller, C., Faugeras, O., & Hebert, M. (1997). Applications of non-metric vision to some visually-guided robotics tasks. In Y. Aloimonos (Ed.). *Visual navigation: From biological systems to unmanned ground vehicles, advances in computer vision* (Vol. II). New Jersey, USA: Lawrence Erlbaum Associates Mahwah. 89–134.

Shah, R., Domini, F., & Caudek, C. (submitted for publication). Do we perceive a flattened world on the monitor screen?

Tassinari, H., Domini, F., & Caudek, C. (2008). The intrinsic constraint model for stereo-motion integration. *Perception, 37*, 79–95.

Thaler, L., & Goodale, M. A. (2010). Beyond distance and direction: The brain represents target locations non-metrically. *Journal of Vision, 10*(3), 1–27.

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

Todd, J. T. (2004). The visual perception of 3d shape. *Trends in Cognitive Sciences, 8*, 115–121.

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., Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1996). Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 695–706.

Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 583–595.

Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics, 50*, 509–523.

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

Todd, J. T., Thaler, L., Dijkstra, T. M. H., Koenderink, J. J., & Kappers, A. M. L. (2007). The effects of viewing angle, camera angle and sign of surface curvature on the perception of 3D shape from texture. *Journal of Vision, 7*(12), 1–16.

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

van Boxtel, J. J. A., Wexler, M., & Droulez, J. (2003). Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision, 3*(5), 318–332.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5*, 834–862.

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