



UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE

Repository istituzionale dell'Università degli Studi di Firenze

Spatial integration in structure from motion

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

Spatial integration in structure from motion / DI LUCA M.; DOMINI F.; C. CAUDEK. - In: VISION RESEARCH. - ISSN 0042-6989. - STAMPA. - 44:(2004), pp. 3001-3013. [10.1016/j.visres.2004.07.004]

Availability:

This version is available at: 2158/203590 since:

Published version:

DOI: 10.1016/j.visres.2004.07.004

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

Publisher copyright claim:

(Article begins on next page)

Spatial integration in structure from motion

Massimiliano Di Luca^a, Fulvio Domini^{a,*}, Corrado Caudek^b

^a *Department of Cognitive and Linguistic Sciences, Brown University, P.O. Box 1978, Providence, RI 02912, USA*

^b *Psychology Department, University of Trieste, Italy*

Received 17 February 2003; received in revised form 22 September 2003

Abstract

In three experiments we investigated whether the perception of 3D structure from the optic-flow involves a process of spatial integration. The observer's task was to judge the 3D orientation of local velocity field patches. In two conditions, the patches were presented either in isolation, or as part of a global optic-flow. In Experiment 1, the global optic-flow was a linear velocity field. In Experiment 2, the patches were embedded in a randomly perturbed linear velocity field. In Experiment 3, the local patches belonged to a smoothly curved surface.

The results of these three experiments lead to two main conclusions: (1) a process linking spatially separated patches into global entities does affect the perception of local surface orientation induced by the optic-flow, and (2) linearity or smoothness of the global velocity field are not necessary conditions for spatial integration.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Vision; Structure from motion; Spatial integration

1. Introduction

The study of the visual processes involved in the reconstruction of 3D structure from dynamic information has, in the last few years, established three main facts. First, the visual system uses only first-order temporal properties (i.e., two views) in order to derive 3D shape from moving images (e.g., Todd & Bressan, 1990; Todd & Norman, 1991). The main theoretical implication of this finding is that perceived structures, in general, do not have the same Euclidean properties as the projected structures (for a review, see Norman & Todd, 1992), since three or more views are needed for a veridical reconstruction of 3D shape (Hoffman, 1982; Ullman, 1979). Second, a number of recent empir-

ical studies have revealed that geometric properties of perceived structures are derived through heuristic processes that provide a non-veridical solution to the SfM problem (Caudek & Domini, 1998; Domini & Caudek, 1999). In particular, perceived local orientation and motion of projected surfaces depend on properties of the optic-flow that are not related in a one-to-one mapping with the distal properties that they represent (Domini & Caudek, 1999; Todd & Perotti, 1999). Third, perceived local properties of smooth surfaces are not consistent with a coherent Euclidean or affine global representation (Domini & Braunstein, 1998; Domini, Caudek, & Richman, 1998).

These findings seem to be in apparent contradiction with our perceptual experience of coherent 3D shapes. If local properties of smooth surfaces are derived in a non-veridical manner and they are internally inconsistent, how do we perceive smooth global surfaces? In this paper we suggest that perceived surface orientation cannot be understood solely in terms of a local analysis of

* Corresponding author. Tel.: +1 401 863 1356; fax: +1 401 863 2255.

E-mail address: fulvio_domini@brown.edu (F. Domini).

the stimulus information. In particular, we will present empirical results showing that the perceived orientation of local patches specified by dynamic random-dot displays depends on the properties of the surrounding stimulus regions. Moreover, we will show that the effects of neighboring regions on the perceived orientation of a local patch are not specific to the case of a smooth optic-flow. To motivate our experiments, in the next section we will briefly describe a computational formulation of the analysis of local optic-flow. This formulation provides a good account of empirical results on local surface slant perception from dynamic information.

2. Local slant perception

The relative motion between an observer and a three-dimensional surface can be described as illustrated on Fig. 1. A coordinate system (x, y, z) can be located at the viewing point with the z -axis corresponding to the viewing direction. If we assume that the main components of ego-motion are a horizontal translation (T_x , corresponding to lateral head-motion) and vertical rotation (ω_y , corresponding to head rotation), then the image velocities (\dot{u}) projected on the image plane (u, v) located at a distance f from the origin of the coordinate axes can be described by the following equation (Longuet-Higgins & Prazdny, 1980):

$$\dot{u} = \frac{-T_x f}{z} - \omega_y f - \omega_y \frac{u^2}{f} \quad (1)$$

If only a local portion of the visual field is considered, then the horizontal (α_u) and vertical (α_v) visual angles can be approximated by $\alpha_u \approx \frac{u}{f}$ and $\alpha_v \approx \frac{v}{f}$. In this case the above equation can be rewritten in terms of visual angles by dividing the left and right sides by f . As a consequence, the horizontal velocity ($\dot{\alpha}_u = \frac{\dot{u}}{f}$) can be approximated by

$$\dot{\alpha}_u \approx \frac{-T_x}{z} - \omega_y \quad (2)$$

since $\alpha_u^2 = \frac{u^2}{f^2}$ is negligible. If the viewed surface is smooth, it can be locally approximated by a planar patch having the following equation:

$$z = g_x x + g_y y + d \quad (3)$$

where g_x and g_y are the horizontal and vertical depth gradients and d is the distance of the planar surface from the origin of the coordinate axes (see Fig. 1, panel). If (3) is substituted in (2), we obtain (see Appendix A):

$$\dot{\alpha}_u \approx -\left(\frac{T_x}{d} + \omega_y\right) + \frac{T_x}{d}(g_x \alpha_u + g_y \alpha_v) \quad (4)$$

Eq. (4) shows that, in general, *motion parallax* and *structure from motion* (SfM) are both contributing to the pattern of retinal velocities that results in a *linear velocity field*. Following the traditional definitions of these two dynamic sources of information, motion parallax is produced by pure observer translation ($\omega_y = 0$) and SfM by pure surface relative rotation. The second case arises when the observer fixates the point of the surface defined by the intersection of the planar patch and the z -axis. In order to keep fixation on this point, the vertical rotation must compensate the horizontal translation, i.e., $\omega_y = -\frac{T_x}{d}$.

Usually the information provided by a linear velocity field is described in terms of three parameters: the mean translation component (V_u) and the horizontal (φ_u) and vertical (φ_v) velocity gradients (Domini & Caudek, 1999; Litter & Braunstein, 1998; Todd & Perotti, 1999). In fact, Eq. (4) can be written as:

$$\dot{\alpha}_u \approx V_u + \varphi_u \alpha_u + \varphi_v \alpha_v \quad (5)$$

where $V_u = -(\frac{T_x}{d} + \omega_y)$, $\varphi_u = \frac{T_x}{d} g_x$ and $\varphi_v = \frac{T_x}{d} g_y$.

An important property of the linear velocity field is the deformation (def), i.e., the intensity of the velocity field gradient along the direction in which the velocity variation is highest (Koenderink, 1986). It can be shown that $\text{def} = \sqrt{\varphi_u^2 + \varphi_v^2}$.

It is important to note that the assumption of pure observer translation (motion parallax) or pure surface rotation (SfM) leads to two different interpretations of the velocity field. In the motion parallax case, the velocity field completely specifies the 3D surface interpretation whereas, in the SfM case, the velocity field is inherently ambiguous. In order to best clarify this point, let us describe the surface orientation in terms of slant ($\sigma = \sqrt{g_x^2 + g_y^2}$) and tilt ($\tau = \frac{g_y}{g_x}$). For both the motion parallax and SfM interpretations, tilt is specified (in a specific instant of time) by the instantaneous velocity field:

$$\tau = \frac{\varphi_v}{\varphi_u} = \frac{\frac{T_x}{d} g_y}{\frac{T_x}{d} g_x} = \frac{g_y}{g_x} \quad (6)$$

In the motion parallax case, since $V_u = -\frac{T_x}{d}$ and $\text{def} = \sqrt{\varphi_u^2 + \varphi_v^2} = \frac{T_x}{d} \sigma$, slant σ is specified by the ratio $|\frac{\text{def}}{V_u}|$. In the SfM case, however, the slant of the surface is not univocally specified by the velocity field. In fact, since $\omega_y = -\frac{T_x}{d}$, $V_u = 0$ and $\text{def} = |\omega_y \sigma|$. In this case, the information provided by def is ambiguous, since there are infinite combinations of slant (σ) and angular rotation (ω_y) that produce the same deformation (van Veen & Werkhoven, 1996).

¹ Since the stimuli used in the experiments here reported are always smaller than 8° of visual angle, this approximation is appropriate.

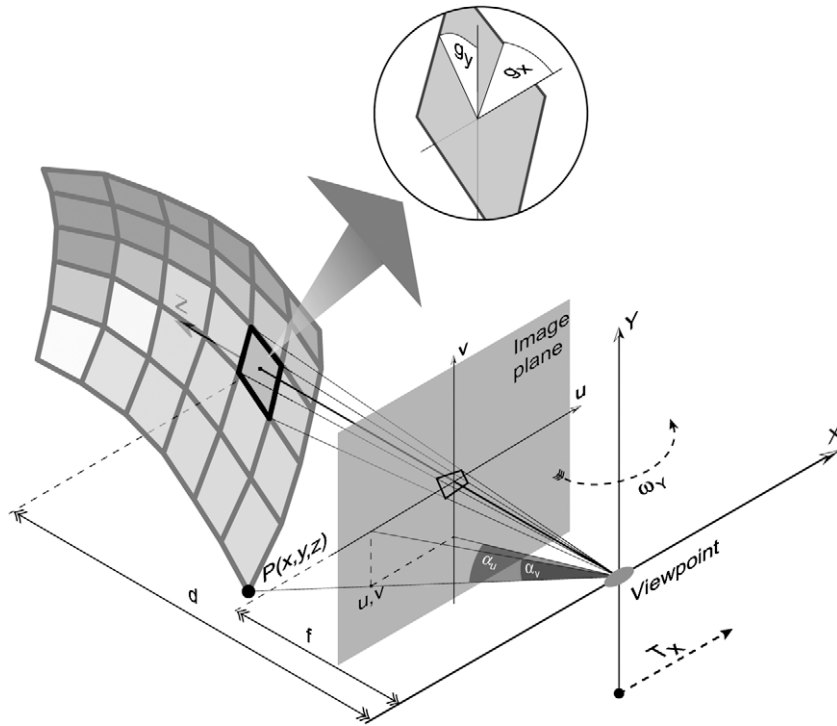


Fig. 1. Schematic representation of the relative motion between an observer and a three-dimensional surface. The coordinate system (x,y,z) has origin at the viewpoint and the z -axis is the line of sight. The image plane is located at a distance f from the origin of the coordinate system. A 3D point $P(x,y,z)$ projects on the image plane a point (u,v) . α_u and α_v indicate the horizontal and vertical visual angles subtended by P . The inset represents a local planar patch at a distance d from the viewing point which orientation can be described in terms of the horizontal and vertical depth gradients g_x and g_y .

Whereas several empirical investigations focused on motion parallax (Braunstein & Tittle, 1988; Domini et al., 1998; Gibson, Gibson, Smith, & Flock, 1959; Liter & Braunstein, 1998), only recently pure SfM has been studied (Domini et al., 1998). In a recent work, Domini and Caudek (1999) proposed that the ambiguity of the velocity field could be overcome by selecting, among the infinite pairs of slant and angular velocities compatible with a given def, the most likely one. In particular, they have shown that, if the a priori distributions of slant and angular velocity are uniform and limited, then the posteriori probability distribution, $p(\omega_y, \sigma/\text{def})$, has a maximum (ω_y^*, σ^*) at $\omega^* = k_\omega \sqrt{\text{def}}$ and $\sigma^* = \frac{1}{k_\omega} \sqrt{\text{def}}$ (see Fig. 2, V_{u0}). Several empirical investigations support the hypothesis that the perceptual interpretation of pure SfM does indeed conform to the analysis proposed by Domini and Caudek (e.g., Todd & Perotti, 1999).

Pure motion parallax and pure SfM are the two extremes of a continuum in which these two “sources of information” are combined. In fact, def and V_u do not specify whether the relative motion between the observer and the planar surface is due to a pure translation, or whether some component of vertical rotation is involved as well. If we define $\omega = -\omega_y$ as the amount of rotation that compensates the observer translation, and if we assume that this quantity ranges from 0 (pure translation)

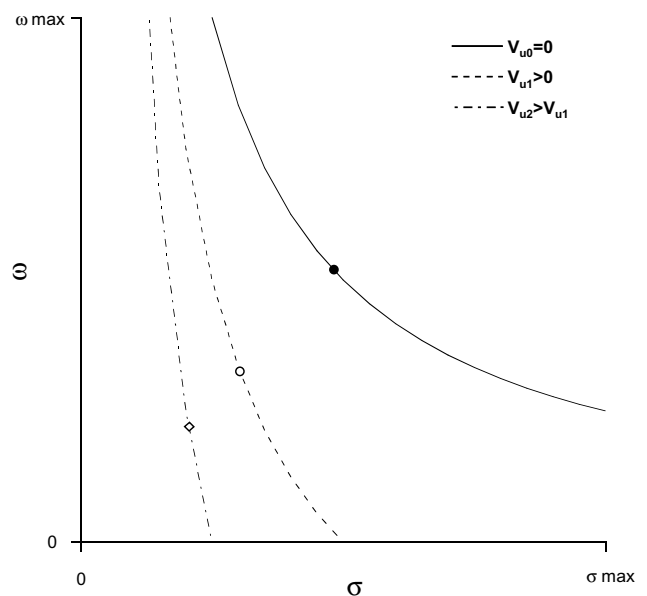


Fig. 2. The three curves represent the loci of all σ, ω pairs that are compatible with a given value of def and three different translational components of the velocity field (V_u). The solid curve represents the case studied by Domini and Caudek (1999) in which $V_u = 0$.

to $\frac{T_x}{d}$ (pure rotation), then the equations that describe the generic observer motion become:

$$V_u = -\left(\frac{T_x}{d} - \omega\right) \quad (7)$$

$$\text{def} = \frac{T_x}{d} \sigma$$

If T_x and ω , cannot be determined through vestibular information provided by head and body motion,² then the above equations are ambiguous, as in the case of a pure SfM, since the ratio $|\frac{\text{def}}{V_u}|$ does not specify a unique value of slant (σ). The mean translational component V_u , however, provides additional information that is not already contained in def. In fact, if the ratio $\frac{T_x}{d}$ is derived from the first equation of (7) and it is substituted in the second one, then def can be expressed as:

$$\text{def} = \sigma(|V_u| + \omega) \quad (8)$$

As for the pure SfM case ($V_u = 0$), Eq. (8) shows that infinite pairs (σ, ω) produce the same value of def. The curve that specifies these infinite solutions, however, shifts when the intensity of the mean translational component is increased and def is kept constant, as shown in Fig. 2 (V_{u1} and V_{u2}). Since the increase of V_u shifts downward the curve that represents the family of possible solutions of slant and angular velocity, the heuristic proposed by Domini and Caudek (1999) derives smaller σ, ω values as V_u increases and def is kept constant (see Fig. 2). According to their proposal, therefore, perceived slant and angular velocity should be inversely related to the intensity of the translational component ($|V_u|$).

It should be noted that this analysis applies only to the first-order structure of the image motion, but not to its second-order structure and that (a) the second-order differential structure associated to surface shape can be recovered independently of the first-order structure and entails a one-to-one mapping between the image properties and the projected 3D shape (Lappin & Craft, 2000); (b) previous studies as well as the present one show that observers are not veridical in recovering first-order structure properties associated with surface slant, relative depth and angular velocity (e.g., Todd & Perotti, 1999); (c) judgments of surface shape (second-order structure) have been found to be more accurate than judgments of first-order structure (e.g., Perotti, Todd, Lappin, & Phillips, 1998).

3. Local vs. global

The computational formulation described in the previous section provides a good account of the empirical data about the perception of *local* orientation in dy-

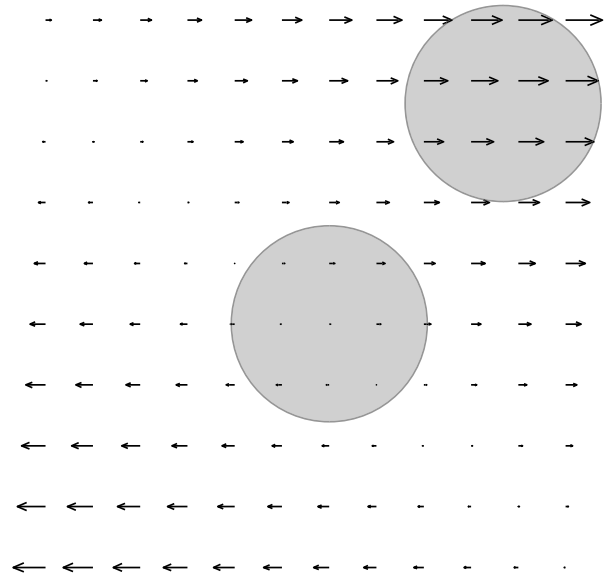


Fig. 3. Schematic representation of the instantaneous velocity field produced by the orthographic projection of a planar surface slanted in depth and rotating about the vertical axis. The two apertures evidence two areas respectively characterized by a nil and a positive translational component.

namic displays. Several studies, in fact, have shown that surface slant is an increasing function of deformation and a decreasing function of the translational component V_u (Domini & Caudek, 1999; Litter & Braunstein, 1998; Todd & Perotti, 1999). In the following discussion, however, we will show that this formulation is not sufficient to explain global surface perception, since the heuristic procedure outlined above does not derive local orientation in a veridical fashion and, therefore, necessarily leads to global inconsistencies.

We will now discuss the particular case in which the optic-flow is produced by the projection of a planar surface rotating about the vertical axis.³ If the visual angle subtended by the stimulus is not larger than 8° (as for the displays of the present investigation), the projected velocity field is approximately linear and can be described by Eq. (5) (see Fig. 3). Domini and Caudek (1999) investigated the perception of such a velocity field with a translational component $V_u = 0$ (pure SfM) and found that perceived slant is an increasing function of the deformation.

Let us now consider two local regions of such an optic-flow: a central and a peripheral region (see Fig. 3). The central region is characterized by a null translational component ($V_u = 0$); the peripheral region, however, has the same value of def as the central region,

² Even though some recent studies have shown that extra-retinal information is used by the perceptual system to help the interpretation of the optic-flow (Wexler, Lamouret, & Droulez, 2001), there is no evidence so far that these parameters are used to derive a veridical solution.

³ In the present context, the concept of “local” is defined with reference to a planar surface since we restrict our analysis to the first-order properties of the image. It is important to notice that also second-order shape is locally defined.

but a translational component $V_u \neq 0$. According to the model described above, when viewed in isolation, these two regions should give rise to different perceived slants. In particular, perceived slant should be bigger for the central than for the peripheral region. This, however, is at odds with the fact that a linear velocity field is usually perceived as a planar surface, implying that, for a large velocity field, perceived slant must depend on some grouping process linking spatially separated patches into coherent global entities.

The grouping processes that integrate local patches into a global surface require operations over an extended area, but still, their goal can be regarded as achieving a representation of surface orientation for each optic-flow patch. Braddick and Qian (2001) proposed that the representation of surfaces, rather than being indexed by local spatial locations, must be indexed by objects—perhaps the kind of representation that has been referred to as an “object file” (Triesman, 1988). It is not clear what role eye movements play in the emergence of global surfaces from local patches. The fact remains, however, that global surfaces have different properties than local patches. The present research starts to investigate the effects of grouping on perceived surface orientation, when a local optic-flow patch is presented in isolation or is embedded in a larger optic-flow field.

The purpose of Experiment 1 was to establish whether the computational formulation provided above is sufficient to explain the perception of local surface orientation in dynamic random-dot displays simulating a rotating planar surface. The results show that this formulation is indeed compatible with the observers' judgments when local regions are presented in isolation. When the same regions are embedded in a global optic-flow, however, the observers' settings deviate from the predictions of the heuristic proposed by Domini and Caudek (1999). The difference between these two conditions, therefore, provide evidence that the processing of local surface orientation is affected by the presence of a surrounding flow field. In Experiment 2, we asked whether grouping affects the perception of local surface orientation only in the case of smoothly connected optic-flow patches, or also for an un-structured surrounding optic-flow. We found that the smoothness of the velocity field is not a necessary condition for the grouping effects observed in Experiment 1. In Experiment 3, we extended the results of the first two experiments to the case of curved surfaces.

4. General methods

4.1. Observers

All the observers were undergraduate students from Brown University, Providence, and they were paid for

their participation. They were naive to the purpose of the research and were not familiar with experiments involving structure from motion displays. All had normal or corrected-to-normal vision.

4.2. Design

All the independent variables were studied within observers. Each subject viewed a fixed number of trials of each condition in one block, with the order of trials completely randomized. The dependent variable was the perceived orientation of the surface and was coded in terms of slant and tilt.

4.3. Apparatus

The stimulus displays were presented on a Sony Trinitron 19" color monitor controlled via a HP Visualize computer. The resolution of the monitor was 1280×1024 and the refresh rate was 60 Hz. The graphic buffer used was 32 bits deep. The monitor was viewed monocularly through a circular window from a distance of 200 cm. The window limited the visible portion of the monitor to a circular region of 28 cm in diameter (8° of viewing angle).

4.4. Stimuli

The displays were composed of high-luminance anti-aliased dots on a low-luminance background and simulated either planar (Experiments 1–3) or curved (Experiment 4) surfaces oscillating back and forth around the vertical axis. 2000 dots were randomly positioned in a circular region of 28 cm in diameter. In order to measure the perceived orientation of the surface we superimposed an adjustable gauge figure on the random-dot stimulus (as in Domini & Caudek, 1999, Experiment 3; originally from Koenderink, Van Doorn, & Kappers, 1992). This gauge figure was depicted as the orthographic projection of a wire-frame hemisphere composed of 12 meridians and four parallels. The hemisphere had a diameter of 30 arcmin of visual angle (see Fig. 4).

4.5. Procedure

The observers were instructed to judge the local orientation of the surface by adjusting the orientation of the gauge figure through mouse movement. They were told that the base of the hemisphere should be perceived as parallel to the random-dot surface. When they were satisfied with the orientation of the gauge figure, they pressed the mouse button twice to initiate the next trial. Viewing was monocular and a chin rest restricted head motion. Eye movements were not restricted. The experimental room was dark during the whole duration of the

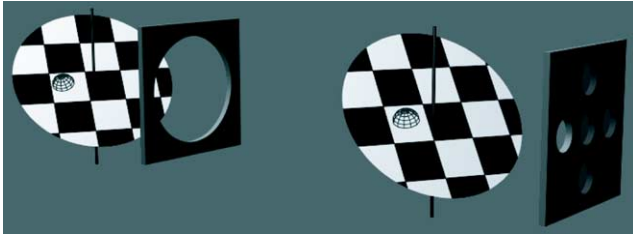


Fig. 4. Schematic drawing of the experimental setting used in Experiment 1. The vertical line represents the axis of rotation. The wire-frame hemisphere represents the gauge figure used by the observers to estimate local slant. *Left panel:* the whole surface is visible through the mask. *Right panel:* the surface and the probe are visible only through one of the apertures.

experiment. The responses were not timed and no restriction was placed on the viewing time. A training session always preceded the actual experiment, but none of the conditions shown in the experiment were used. No feedback was provided.

5. Experiment 1

The purpose of Experiment 1 was to determine whether perceptual grouping affects the perception of local surface orientation induced by dynamic random-dot displays. We reasoned that, if perceived local slant is influenced by the surrounding velocity field—not just by the deformation and the intensity of the translational component of the local optic-flow—then some form of grouping must take place. In the two main experimental conditions, observers were shown (1) local regions of a larger velocity field in isolation (the surrounding optic-flow was occluded), and (2) the whole velocity field (see Fig. 4).

5.1. Method

5.1.1. Observers

Eleven Brown University undergraduates participated to the experiment.

5.1.2. Design

Two independent variables were studied in this experiment: (1) the translational component V_u of the local regions (0.00°/s and 3.61°/s), and (2) the viewing condition (local patch seen in isolation or as part of a global velocity field).

5.1.3. Stimuli

In order to manipulate the average velocity of the local patches, the observers judged different regions of the global velocity field (see Fig. 4, left panel). The perceived orientation of five local regions: a central region and four peripheral regions was also measured (Fig. 4, right

panel). The central region was characterized by a null translational component. All the peripheral regions had a translational component of 3.61°/s. In different blocks these regions could be either seen in isolation (by using a 7 cm/2° circular aperture), or as part of the whole velocity field (see Fig. 4). The local regions shown through the circular aperture were 1° of visual angle apart. The global velocity field was seen through a circular window of 8° in diameter. The motion of the dots was consistent with a linear constant velocity field having a deformation of 0.76 s⁻¹ and a tilt (the arc-tangent of the ratio between the vertical and horizontal velocity gradients) of either +45° or -45°. Dot density was kept constant during each stimulus sequence. One oscillation cycle about the vertical axis took 1 s (60 frames). Each observer viewed four presentations of the 20 conditions.

5.2. Results and discussion

Fig. 5 plots mean perceived slant as function of the translational component of the local velocity field in the two experimental conditions. For the partial-viewing condition, the data are compatible with the qualitative prediction of the model of Domini and Caudek (1999) and with previous reports on slant perception (e.g., Todd & Perotti, 1999). When each optic-flow patch is presented in isolation, perceived slant is a decreasing function of the translational component of the optic-flow. This trend is significantly reduced, however, when

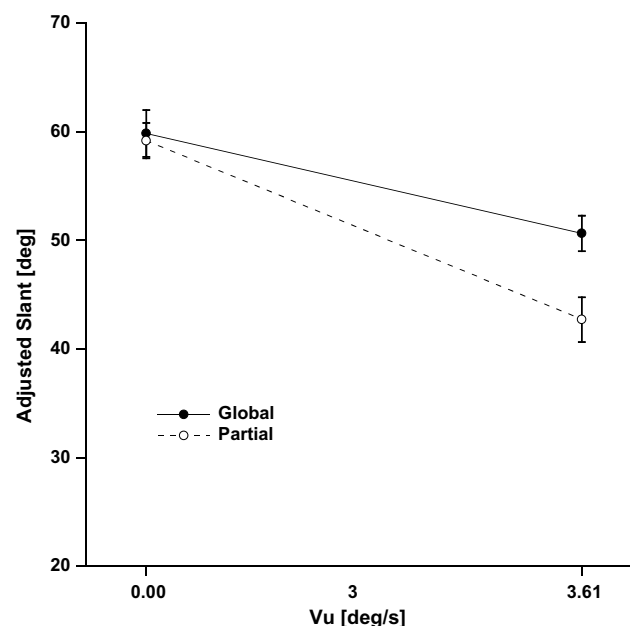


Fig. 5. Average judged slant in Experiment 1 as a function of V_u in the global- and partial-viewing conditions. Vertical bars represent one standard error. In the present and in the following figures, the standard errors are defined on the between-observers variability.

the whole velocity field is visible. The different results obtained in the local- and global-viewing conditions, thus, indicate that the process linking spatially separated patches into global entities does affect the perception of local surface orientation.

A 2 (viewing condition: global vs. partial) \times 2 (translational component: 0°/s and 3.61°/s) repeated-measures analysis of variance (ANOVA) on perceived slant revealed a main effect of viewing condition ($F(1,10) = 6.06$, $p < 0.05$) and a main effect of the translational component ($F(1,10) = 24.52$, $p < 0.01$). The interaction between the two independent variables was significant ($F(1,10) = 12.95$, $p < 0.01$).

The variability of observers' judgments is reported in Table 1 for each observer and each condition. Weber fractions were computed by dividing the standard deviations by the mean judged orientation for each stimulus and each observer, ignoring the non-linearity of this angular variable (see Fig. 6). A repeated measures ANOVA on these Weber fractions substantially replicated the results of the previous analysis. A significant effect was found for the translational component ($F(1,10) = 37.367$, $p < 0.001$); the main effect of viewing condition was not significant ($F(1,10) = 2.799$, n.s.); the interaction between the two independent variables was marginally significant ($F(1,10) = 4.474$, $p = 0.06$).

Since in the present experiment eye movements were not restricted, it would be possible to argue that observers might have always performed a local computation around their fixation point. Under these conditions, the increase in perceived slant in the global condition (relative to that for peripheral stimuli presented in a local region) could be due to observers fixating points closer to the center of the global stimulus where the net translational motion is decreased. We addressed this issue by running an additional experiment, here not re-

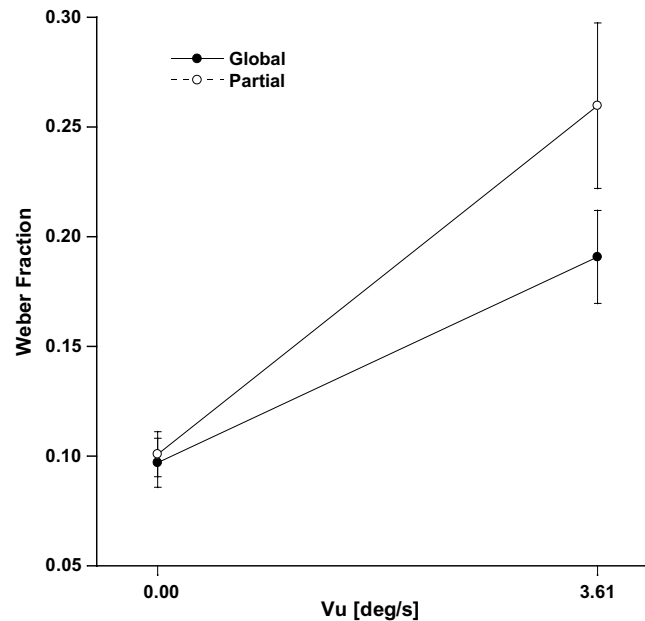


Fig. 6. Weber fractions as a function of V_u in the *global*- and *partial*-viewing conditions of Experiment 1. Vertical bars represent one standard error.

ported. In this control experiment, observers were instructed to maintain fixation to the center of the display, while adjusting the gauge figure in the periphery. This task was very difficult to perform, as indicated by the enormous variability of the observers' settings. Since in Experiment 1, conversely, the variances of the observers' settings did not significantly differ between the global and partial viewing conditions (the variance was actually slightly smaller in the global condition), a possible explanation of the results of Experiment 1 in terms of eye movements can be ruled out.

Table 1

Standard deviations defined on the within-observer variability for all observers in each condition of Experiment 1

| Subject | Global | | Partial | |
|-------------|--------|-------|---------|-------|
| | 0.00 | 3.61 | 0.00 | 3.61 |
| 1 | 7.64 | 9.27 | 7.32 | 14.47 |
| 2 | 4.15 | 9.80 | 2.38 | 5.85 |
| 3 | 3.75 | 12.10 | 7.17 | 14.06 |
| 4 | 8.11 | 14.55 | 7.58 | 10.28 |
| 5 | 3.63 | 6.57 | 4.14 | 7.91 |
| 6 | 2.72 | 6.17 | 6.66 | 11.68 |
| 7 | 3.20 | 8.89 | 6.21 | 11.82 |
| 8 | 7.74 | 11.23 | 8.87 | 18.84 |
| 9 | 9.17 | 5.31 | 2.98 | 4.95 |
| 10 | 6.58 | 9.01 | 6.36 | 8.85 |
| 11 | 7.05 | 11.11 | 5.59 | 6.63 |
| RMS average | 5.79 | 9.46 | 5.93 | 10.49 |

Translational component V_u : 0.00°/s vs. 3.61°/s, and global vs. partial viewing conditions.

6. Experiment 2

The results of the previous experiment reveal an effect of the global flow on the perception of local slant in dynamic random-dot displays. We should note, however, that in the previous experiment: (1) the local patches were part of a smoothly connected surface, and (2) all local patches shared the same horizontal and vertical gradients. The purpose of Experiment 2 was to establish whether the smoothness and linearity of the optic-flow are necessary conditions for attributing different perceptual interpretations to local patches presented in isolation or embedded in a larger flow field.

The stimulus displays used in Experiment 2 were different from those used in the previous experiment in two respects. First, the contours of the local patches were visible. It may be argued, in fact, that the effect of the translational component may vanish if additional

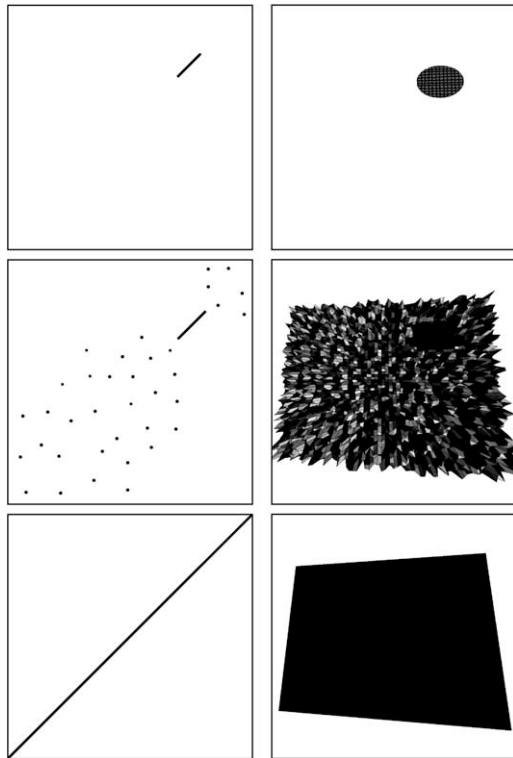


Fig. 7. Schematic representation of the stimuli used in the three viewing condition of Experiment 2. *Top*: patch in isolation with deforming contours; *middle*: local patch embedded in a 3D volume; *bottom*: patch embedded in a linear velocity field. *Left figures*: the stimulus displays are schematically represented as if viewed from the side. *Right figures*: a schematic 3D representation of the stimulus displays. The top row represents the *local-viewing* condition; the middle row represents the *noisy global-viewing* condition; the bottom row represents the *linear global-viewing* condition.

dynamic information (such as contour deformation) is available (Fig. 7, top row). Second, the global optic-flow simulated the projection of a rotating planar surface whose point positions were randomly perturbed along the z -axis (Fig. 7, central row). This display was perceived as a 3D volume of randomly distributed dots containing a small planar region. If the grouping effects observed in the previous experiment occur only when local patches are part of a global *linear* velocity field, then the same slant should be perceived when a local patch is viewed in isolation or as a part of such a 3D volume.

6.1. Method

6.1.1. Observers

Six observers participated to the experiment.

6.1.2. Design

Two independent variables were examined in this experiment: (1) the translational component V_u of the local patches (0.00°/s, 0.96°/s, 1.93°/s, 2.89°/s, 3.86°/s)

and (2) the viewing condition (local patch embedded in a linear velocity field, local patch embedded in a 3D volume, local patch seen in isolation).

6.1.3. Stimuli

The displays corresponding to the three viewing conditions are schematically represented in Fig. 7. The global linear velocity field was identical to the velocity field generated in the previous experiment ($\text{def} = 0.76 \text{ s}^{-1}$). The random volume condition was generated by perturbing the velocities of the linear velocity field with uniform random distribution with mean 0.00°/s and spread $\pm 1.24^\circ/\text{s}$. A portion of the velocity field was left unperturbed. This portion corresponded to a circular area on the 3D surface that projected an ellipsoidal contour on the image plane. The observers judged the perceived slant of this region. The local patch seen in isolation was identical to the local patch embedded in the random volume. The translational component of the local velocity field was manipulated as in the first experiment by showing different regions of the global velocity field. The local regions were selected along a direction orthogonal to the tilt of the optic-flow from the stimulus center to the stimulus periphery such that their average velocities were 0.00°/s, 0.96°/s, 1.93°/s, 2.89°/s, 3.86°/s respectively. Each subject viewed eight presentations of the 15 conditions.

6.2. Results and discussion

Fig. 8 plots mean perceived slant as function of the translational component of the local velocity field in the three experimental conditions. The data show that (1) for local optic-flow patches seen in isolation, perceived slant is a decreasing function of the translational component V_u (even if deforming-contours information is available), (2) perceived slant is not affected by the translational component V_u when the patches are embedded in a linear velocity field (*linear* global-viewing condition), and (3) there is no difference between the *linear* and the *noisy* global-viewing conditions. Even though in the *noisy* global-viewing condition the surrounding velocity field is perceived as a volume of randomly distributed dots, the perception of local patches is still affected by the “perturbed” global field. These data indicate, therefore, that the smoothness of the velocity field is not a necessary condition for assigning a different perceptual interpretation to a local patch viewed in isolation or as part of a larger flow field.

A 3 (local-viewing, noisy global-viewing and linear global-viewing conditions) \times 5 (translational component: 0.00°/s, 0.96°/s, 1.93°/s, 2.89°/s, 3.86°/s) repeated-measures analysis of variance on perceived slant (ANOVA) revealed a main effect of viewing condition ($F(2,10) = 11.58$, $p < 0.01$) and a main effect of the translational

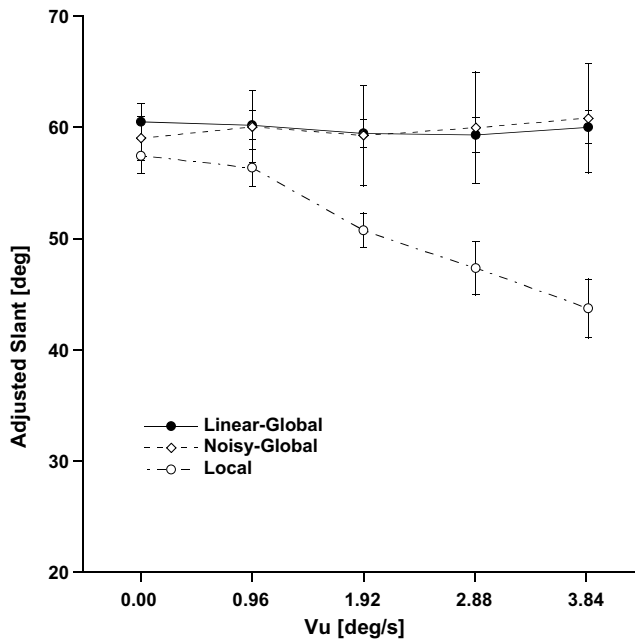


Fig. 8. Average judged slant as a function of the translational component V_u in the *local-viewing*, *noisy global-viewing* and *linear global-viewing* conditions of Experiment 2. Vertical bars represent one standard error.

component ($F(4, 20) = 3.98$, $p < 0.05$). The interaction between the two independent variables was also significant ($F(8, 40) = 8.08$, $p < 0.01$).

The variability of observers' judgments is reported in Table 2 for each observer and each condition. As for Experiment 1, the Weber fractions were computed for each stimulus and each observer (see Fig. 9). Also in this case, a repeated-measures ANOVA on these Weber fractions replicated the results of the previous analysis. A significant effect was found for the viewing condition ($F(2, 10) = 4.771$, $p < 0.05$); the main effect of the translational component was not significant ($F(1, 5) = 2.791$, n.s.); the interaction between the two independent variables was significant ($F(2, 64) = 8.149$, $p < 0.001$).

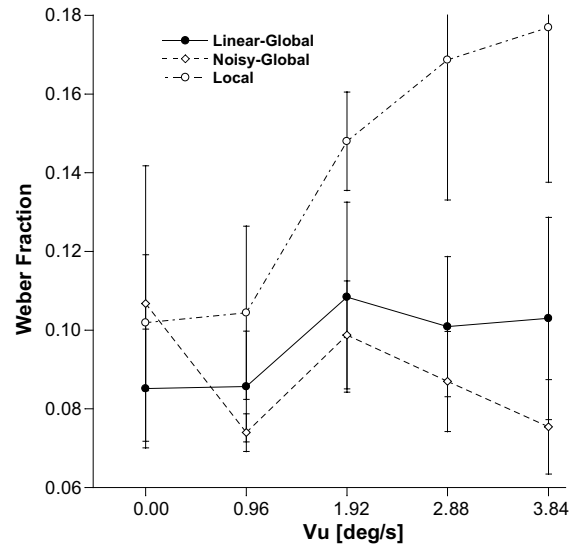


Fig. 9. Weber fractions as a function of V_u in the *local-viewing*, *noisy global-viewing* and *linear global-viewing* conditions of Experiment 2. Vertical bars represent one standard error.

7. Experiment 3

The previous experiments reveal an effect of the surrounding optic-flow on the perceptual interpretation of local surface orientation in dynamic random-dot displays. Whereas perceived slant for an optic-flow patch viewed in isolation was affected by the translational velocity component of the optic-flow, this effect was reduced or vanished when the same patch was viewed as a part of a larger linear-velocity field, or a randomly perturbed linear-velocity field.

A parsimonious way of explaining these results is to hypothesize that *the visual system estimates the average slant of the global optic-flow and assigns this value to each local patch of the velocity field*. In both the previous experiments, in fact, the overall translational component of the global optic-flow was equal to zero and, thus, the average slant of the whole velocity field could be derived by using the heuristic model proposed by Domini and

Table 2
Standard deviations defined on the within-observer variability for all observers in each condition of Experiment 2

| Subject | Linear local | | | | | Noisy global | | | | | Linear global | | | | |
|-------------|--------------|-------|------|-------|------|--------------|------|------|------|------|---------------|------|-------|-------|-------|
| | 0.00 | 0.96 | 1.92 | 2.88 | 3.84 | 0.00 | 0.96 | 1.92 | 2.88 | 3.84 | 0.00 | 0.96 | 1.92 | 2.88 | 3.84 |
| 1 | 7.75 | 10.06 | 8.97 | 12.57 | 10.6 | 11.84 | 4.69 | 8.12 | 5.99 | 6.21 | 6.92 | 5.95 | 12.93 | 6.00 | 12.71 |
| 2 | 2.85 | 4.74 | 6.67 | 8.35 | 8.70 | 2.67 | 3.52 | 3.30 | 2.99 | 2.41 | 5.17 | 3.60 | 3.17 | 3.51 | 3.67 |
| 3 | 5.73 | 4.24 | 7.58 | 8.44 | 8.97 | 2.63 | 5.59 | 4.39 | 4.17 | 2.51 | 3.88 | 3.68 | 4.59 | 4.71 | 3.12 |
| 4 | 5.11 | 4.23 | 6.07 | 4.72 | 3.36 | 5.85 | 4.37 | 4.99 | 5.49 | 4.82 | 2.32 | 3.11 | 6.30 | 4.64 | 4.22 |
| 5 | 8.07 | 6.12 | 7.81 | 4.98 | 5.65 | 9.89 | 3.92 | 6.85 | 7.50 | 5.65 | 6.89 | 6.59 | 4.64 | 5.99 | 3.99 |
| 6 | 4.77 | 4.60 | 7.41 | 6.45 | 5.24 | 2.17 | 4.53 | 7.08 | 4.59 | 5.46 | 4.78 | 7.57 | 6.77 | 10.96 | 9.49 |
| RMS average | 5.71 | 5.67 | 7.42 | 7.59 | 7.09 | 5.84 | 4.44 | 5.79 | 5.12 | 4.51 | 4.99 | 5.08 | 6.40 | 5.97 | 6.20 |

Translational component V_u : 0.0°/s, 0.96°/s, 1.92°/s, 2.88°/s, 3.84°/s; viewing conditions: linear local, noisy global, linear global.

Caudek (1999). This hypothesis would explain why the perceived slant of all local patches in the “global flow” condition is equal to the perceived slant of the central patch ($V_u = 0$) in the “partial viewing” condition (see Fig. 8).

The previous explanation, however, does not apply to the case of a smoothly curved surface since, in that case, the local regions of the surface are always perceived as having different orientations. One could speculate, therefore, that a planar surface (or a randomly-perturbed planar surface) is a special case that is not representative of more generic stimulus conditions. In Experiment 3, therefore, we investigated the effects of grouping in the case of a non-linear velocity field. In this experiment, we generated SfM displays that simulated the orthographic projection of a random-dot hemisphere oscillating about the y -axis. In two conditions, the axis of rotation was either in front or behind the base of the hemisphere (see Fig. 10); two non-linear velocity fields were therefore created, having similar deformations, but different local velocities.

7.1. Method

7.1.1. Observers

Six naive observers participated in this experiment.

7.1.2. Design

Three independent variables were examined: (1) The position of the axis of rotation (in front vs. behind), (2) the viewing condition (local patch embedded in a global velocity field vs. local patch seen in isolation), and (3) the translational component V_u of the local patches (0.00°/s, 1.15°/s, 2.00°/s corresponding to three different radial positions).

7.1.3. Stimuli

The motion of the dots simulated the orthographic projection of a hemisphere oscillating about a vertical

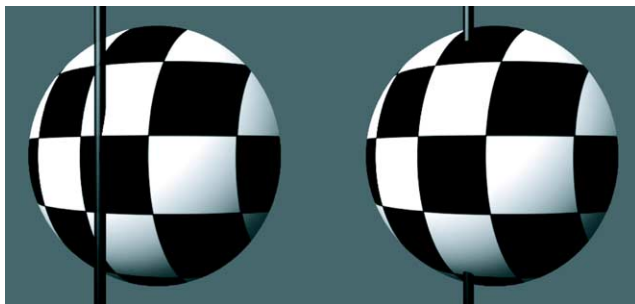


Fig. 10. Schematic representation of the stimuli used in Experiment 3. *Left panel*: axis in front. *Right panel*: axis behind. In the actual experiment, the stimuli were random-dot displays simulated as oscillating about the vertical axis of rotation.

axis. The axis of rotation could be either *in front* (0.98 times the ray of the sphere from its center) or *behind* (0.64 times the ray of the sphere from its center). The manipulation of the axis of rotation served the purpose of changing the translational component of the local patches. Fig. 11 shows how different radial positions of the patches correspond to different translational components. If the axis of rotation is *in front* (open squares on Fig. 11), the translational component increases from the center to the periphery. If the axis of rotation is *behind* (open circles on Fig. 11), the translational component decreases from the center to the periphery. It is important to note that the translational components of the patches in the central radial locations do not depend on the position of the axis of rotation. The average value of def for the local patches was 0.18, 0.47, 0.62 s^{-1} for the three different radial positions, respectively. Each observer viewed eight presentations of the 36 stimulus conditions.

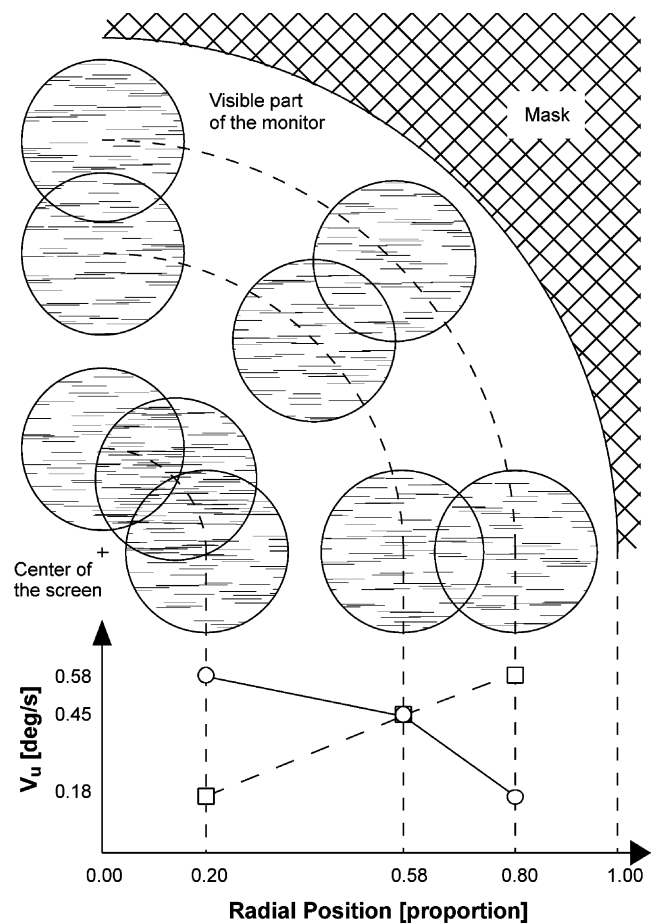


Fig. 11. *Top*: image positions of the local optic-flow patches in Experiment 3. *Bottom*: the relationship between the radial position of the patches and the translational component of the velocity field. *Dashed lines and squares*: axis-behind; *continuous line and circles*: axis-in-front.

7.2. Results and discussion

Fig. 12 plots mean perceived slant as function of the radial position in the four experimental conditions. The data show that, also for a non-linear velocity field, the observers' settings in the global-viewing condition differ from those in the partial-viewing condition. For the patches viewed in isolation (open squares and circles), the qualitative trend of the data is consistent with the heuristic proposed by Domini and Caudek (1999). According to this heuristic, perceived slant depends on the deformation and on the translational component of the local optic-flow. Hence, the function relating perceived local slant to the radial position of the local patch should have opposite slope-signs in the two axes-conditions. The translational component of the local optic-flow, in fact, is an increasing function of radial position if the axis of rotation is in front, and a decreasing function of radial position if the axis of rotation is behind. In both conditions, however, the deformation of the local patches is identical and, thus, we should expect that perceived slant increases with radial position (circles) when the axis is *behind* (since def increases and the translational component decreases) and increase less or decrease (squares) when the axis is *in front* (since both def and the translational component increase). The data for the partial-viewing condition clearly show this interaction (see Fig. 12). In the global-viewing condition, on the other hand, perceived slant always increases with radial position, independently of whether the axis of rotation is in front or behind (since def increases with radial

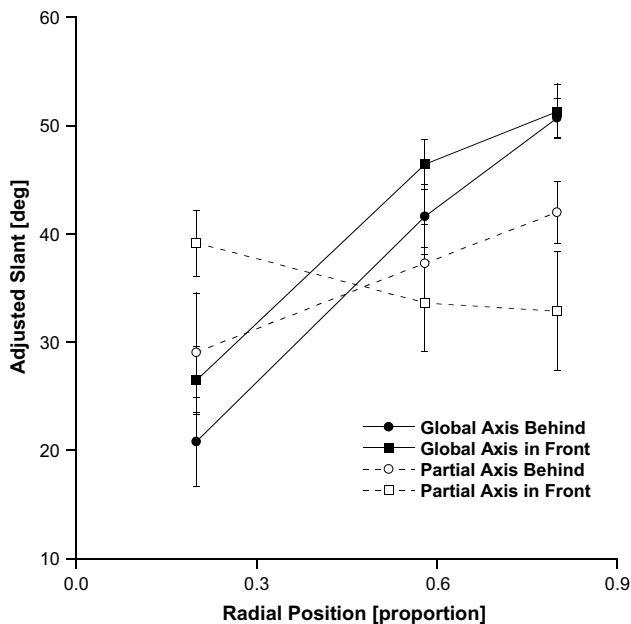


Fig. 12. Average judged slant as a function of the radial position in the four experimental conditions of Experiment 3. Vertical bars represent one standard error.

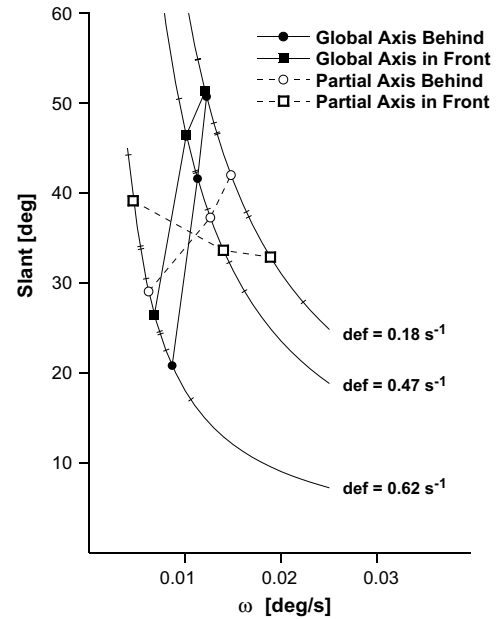


Fig. 13. Average judged slant as a function of angular rotation plotted on the constraint lines defined by simulated def in the global and partial viewing conditions of Experiment 3. Bars represent one standard error. The three hyperbolas represent the loci of the σ , ω pairs compatible with the def magnitudes used in the experiment. Each data point is identified by the σ coordinate on the def constraint line. Domini and Caudek (1999) demonstrated that, for def magnitudes comparable to those used in the present experiment, the observers' settings for perceived orientation (σ) and angular rotation (ω) magnitudes are very consistent.

position). The different qualitative trends of the data for the global and partial conditions cannot be explained by the local properties of the optic-flow (identical in both conditions) and, therefore, must be attributed to the processes linking spatially separated patches into global entities.⁴

A 2 (axis of rotation) \times 2 (viewing condition: global and partial) \times 3 (radial position: inner, central and peripheral) repeated-measures ANOVA revealed a main effect of radial position ($F(2, 10) = 43.49$, $p < 0.01$). The following two-way interactions were significant: radial position and viewing condition ($F(2, 10) = 43.13$, $p < 0.01$), radial position and axis of rotation ($F(2, 10) = 9.72$, $p < 0.01$), viewing condition and axis of rotation ($F(1, 5) = 8.50$, $p < 0.05$). The three-way interaction between radial position, viewing condition and axis of rotation was also significant ($F(2, 10) = 4.60$, $p < 0.05$).

Fig. 13 plots the average slant judgments on the constraint lines representing the three def magnitudes simulated in the present experiment. The figure shows that

⁴ In an analysis here not reported, we found that the perceived 3D shapes inferred from the observers' settings in the *partial*- and *global*-viewing conditions were consistent with different magnitudes of affine stretching of the simulated 3D shape (see Perotti et al., 1998).

the slant judgments in the partial viewing condition are associated to a larger range of angular-rotation magnitudes than in the global viewing condition. This conclusion (even if it is inferred from perceived slant only) provides further supports to the hypothesis that the difference between the partial and global viewing conditions is due to a process of spatial integration. As it should be expected for a rigid surface rotation, in fact, similar angular-rotation magnitudes tend to be associated to different optic-flow patches in the global-viewing condition (but not necessarily in the partial-viewing condition).

8. General discussion

In three experiments we showed that the perception of surface orientation in structure from motion cannot be explained by a purely local analysis of the optic-flow. In Experiment 1, observers judged the perceived slant of local optic-flow patches. In two conditions, these patches were seen either in isolation, or as part of the global optic-flow. Despite the fact that local information was the same, perceived slant was judged differently in the two conditions. When the optic-flow patches were viewed in isolation through a circular window, their perceived slant was a decreasing function of the translational component of the local optic-flow; when the whole global flow was visible, perceived slant was still affected by the translational component of the local optic-flow, but by a smaller degree. The difference in the partial vs. global viewing conditions thus provides evidence that local optic-flow processing is influenced by the surrounding flow field.

Experiment 2 revealed that the grouping effects found in Experiment 1 do not necessitate the linearity or the smoothness of the global velocity field. Similar magnitudes of slant, in fact, were reported for optic-flow patches embedded in a smooth velocity field or in a randomly perturbed linear velocity field. This result is surprising since the perturbed velocity field appeared like a cloud of random-dots—not a planar surface—and, in these circumstances, one may expect that spatial integration does not occur.⁵

⁵ In Experiments 1 and 2, the deformation component of the velocity field was not manipulated since we already demonstrated that, within stimulus conditions similar to those of the present experimental setting, perceived slant magnitudes are unrelated to the slant magnitudes that, in principle, can be derived from the second-order properties of the velocity field (Domini, Caudek, & Proffitt, 1997). While in our previous research we demonstrated that judgments of surface orientation depend primarily on the first-order properties of the optic-flow, the aim of the present investigation was to investigate the influence of the surrounding field on the judgments of local surface orientation.

By using the projection of an oscillating random-dot sphere, in Experiment 3 we investigate a more general stimulus condition than in the first two experiments. Also in these circumstances, however, observers reported different magnitudes of perceived slant when the same local regions were viewed in isolation, or as parts of the global optic-flow.

There are two main theoretical implications of these results. (1) The perceptual processes deriving 3D properties from dynamic information cannot be accounted for by a purely local analysis of the optic-flow (e.g., Domini & Caudek, 1999; Todd & Perotti, 1999). Local-slant judgments, in fact, are affected by the surrounding optic-flow. (2) The perception of a smooth surface is not a necessary condition for spatial integration, as indicated by the results of Experiment 2 where the surrounding optic-flow was produced by the rigid motion of a cloud of dots.

It remains a goal for future research to understand the perceptual mechanisms that govern spatial integration. This is an especially difficult task, since human structure-from-motion only makes use of ambiguous information of the first temporal order and, therefore, (in general) does not derive a veridical 3D structure from a moving image. It could be speculated, however, that the local (first-order) analyses of contiguous optic-flow patches may mutually constrain each other. A local rigidity constraint, for example, could bias towards similar values the perceived rotation of neighboring patches. The global shape may then result from these local interactions, without the guarantee of being veridical. An alternative interpretation is that perceived surface structure may depend on the second-order image properties (Lappin & Craft, 2000). The hypothesis that perceived local orientation is affected by the global context, in fact, is not inconsistent with the view that global shape is accurately perceived up to an affine transformation of the image.

The above considerations can be related to our previous research on temporal integration in SfM (Domini, Vuong, & Caudek, 2002). In that investigation, observers were shown two optic-flow sequences presented side by side. Each sequence was made up of two successive segments, the *history* and the *comparison*. The velocity gradients ($\phi_{x1} > \phi_{x2}$) used for the *comparison* segments were such that, when shown alone, observers reliably associated (in at least the 80% of the cases) the larger perceived slant to the velocity field having the largest gradient (ϕ_{x1}). When ϕ_{x1} was preceded in the *history* segment by a very small velocity gradient, and ϕ_{x2} was preceded by a very large velocity gradient, however, the opposite result was found: Observers consistently judged the velocity field with the smallest gradient (ϕ_{x2}) in the *comparison* phase as having the largest perceived slant. Domini et al. explained these findings by means of a temporal-integration model assuming that

(1) a 3D representation is derived heuristically from the first-order velocity field, and (2) perceived local surface orientation is updated over time by averaging the slant magnitudes specified by the current optic-flow, on the one hand, with the slant and angular rotation magnitudes perceived in previous moments of time, on the other. With reference to this previous analysis, a similar mechanism might be postulated for spatial integration, the only difference being that the dimension along which integration occurs is space rather than time.

Appendix A

To derive Eq. (4) from Eqs. (2) and (3) it should be noted that the relationship between the screen coordinates (u, v) in terms of visual angles ($\alpha_u \approx \frac{u}{f}$, $\alpha_v \approx \frac{v}{f}$) and the 3D coordinates of a point $P(x, y, z)$ is given by

$$\begin{aligned}\alpha_u &\approx \frac{u}{f} = \frac{x}{z} \\ \alpha_v &\approx \frac{v}{f} = \frac{y}{z}\end{aligned}\quad (\text{A.1})$$

If we substitute these in the equation of the planar surface (Eq. (3)), we obtain:

$$z = g_x \alpha_u z + g_y \alpha_v z + d \quad (\text{A.2})$$

From Eq. (A.2) we can derive z and substitute it in the equation of the velocity field (Eq. (2)). This substitution leads to Eq. (4).

References

- Braddick, O., & Qian, N. (2001). The Organization of Global Motion and Transparency. In J. M. Zanker & Z. Jochen (Eds.), *Motion Vision—Computational, Neural, and Ecological Constraints* (pp. 86–112). Springer-Verlag.
- Braunstein, M. L., & Tittle, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance*, 14(4), 582–590.
- Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 609–621.
- Domini, F., & Braunstein, M. L. (1998). Recovery of 3-D structure from motion is neither euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 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., & 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 and Psychophysics*, 60(7), 1164–1174.
- Domini, F., Vuong, Q., & Caudek, C. (2002). Temporal integration in structure from motion. *Journal of Experimental Psychology: Human Perception and Performance*, 28(4), 816–838.
- Gibson, E. J., Gibson, J. J., Smith, O. W., & Flock, H. (1959). Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, 58, 40–51.
- Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America*, 72(7), 888–892.
- Koenderink, J. J. (1986). Optic flow. *Vision Research*, 26(1), 161–179.
- Koenderink, J. J., Van Doorn, A. J., & Kappers, A. M. (1992). Surface perception in pictures. *Perception and Psychophysics*, 52(5), 487–496.
- Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: From retinal images to perceived shapes. *Psychological Review*, 107(1), 6–38.
- Liter, J. C., & Braunstein, M. L. (1998). The relationship of vertical and horizontal velocity gradients in the perception of shape, rotation, and rigidity. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1257–1272.
- Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London: Series B*, 208, 385–397.
- Norman, J. F., & Todd, J. T. (1992). The visual perception of 3-dimensional form. In G. A. Carpenter & S. Grossberg (Eds.), *Neural networks for vision and image processing* (pp. 93–110). Cambridge, MA, US: The MIT Press.
- Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception and Psychophysics*, 60(3), 377–388.
- Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception and Psychophysics*, 48(5), 419–430.
- Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception and Psychophysics*, 50(6), 509–523.
- Todd, J. T., & Perotti, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception and Psychophysics*, 61(8), 1577–1589.
- Triesman, A. (1988). Features and objects. *Quarterly Journal of Experimental Psychology*, 40, 201–237.
- Ullman, S. (1979). *The interpretation of visual motion*. Oxford, England: Massachusetts Institute of Technology Pr.
- van Veen, H. A. H. C., & Werkkhoven, P. (1996). Metamerisms in structure-from-motion perception. *Vision Research*, 36(14), 2197–2210.
- Wexler, M., Lamouret, I., & Droulez, J. (2001). The stationarity hypothesis: an allocentric criterion in visual perception. *Vision Research*, 41(23), 3023–3037.