Available online at www.sciencedirect.com
SCIENCE DIRECT®
PERGAMON
Vision Research 43 (2003) 1753–1764
Vision Research
www.elsevier.com/locate/visres

Recovering slant and angular velocity from a linear velocity field: modeling and psychophysics

Fulvio Domini [a,b,*], Corrado Caudek [a,b]

[a] Department of Cognitive and Linguistic Sciences, Brown University, P.O. Box 1978, Providence, RI 02912, USA
[b] Department of Psychology, University of Trieste, via S. Anastasio, 12, I-34134, Trieste, Italy

Received 29 January 2002; received in revised form 28 February 2003

Abstract

The data from two experiments, both using stimuli simulating orthographically rotating surfaces, are presented, with the primary variable of interest being whether the magnitude of the simulated gradient was from expanding vs. contracting motion. One experiment asked observers to report the apparent slant of the rotating surface, using a gauge figure. The other experiment asked observers to report the angular velocity, using a comparison rotating sphere. The results from both experiments clearly show that observers are less sensitive to expanding than to contracting optic-flow fields. These results are well predicted by a probabilistic model which derives the orientation and angular velocity of the projected surface from the properties of the optic flow computed within an extended time window.

Keywords: Vision; Structure from motion; Optic flow

1. Introduction

The problem of how the human visual system extracts three-dimensional (3D) information from motion (the so-called structure-from-motion (SFM) problem) has been debated for a long time (e.g., Braunstein, Hoffman, & Pollick, 1990; Braunstein, Hoffman, Shapiro, Andersen, & Bennett, 1987; Caudek & Proffitt, 1993; Landy, Dosher, Sperling, & Perkins, 1991; Lappin, Doner, & Kottas, 1980; Treue, Husain, & Andersen, 1991), and several models have been proposed trying to account both for veridical performance and biases in human performance (e.g., Domini & Caudek, 1999; Hildreth, Ando, Andersen, & Treue, 1995; Hildreth, Grzywacz, Adelson, & Inada, 1990; Todd & Bressan, 1990; Todd & Perotti, 1999). Most models of human SFM proposed so far assume that the goal of the interpretation process is the recovery of the Euclidean or affine structure of the projected objects, and of their motion in 3D space. In a series of papers (for a review, see Domini & Caudek, in press; Proffitt & Caudek, 2002), however, we have shown that (1) the perceived 3D structure and motion derived from a moving projection cannot be described by Euclidean or affine geometry, and (2) the perceptual solution to the SFM problem, in general, is not veridical.

To account for these results, we proposed a theory based on a few simple hypotheses: (1) the visual system measures a local property of the optic flow called deformation (the deformation is a scalar property defined by the instantaneous velocities of three feature points—see Koenderink, 1986), (2) the visual system groups together feature points that generate similar deformations (Caudek & Rubin, 2001; Domini, Caudek, & Proffitt, 1997), and (3) the visual system derives the local orientation and the 3D motion of the planar patches identified by the grouping process by means of a heuristic procedure (Domini & Caudek, 1999).

The purpose of the present work is to further develop our previous model so as to account for the smaller sensitivity to expanding than to contracting optic-flow fields revealed by human observers, as demonstrated by our recent research (e.g., Domini, Caudek, & Skirko, 2003; Domini, Vuong, & Caudek, 2002). Before discussing the details of our model of human SFM, however,

* Corresponding author at: Department of Cognitive and Linguistic Sciences, Brown University, P.O. Box 1978, Providence, RI 02912, USA. Tel.: +1-401-863-1356; fax: +1-401-863-2255.
E-mail address: fulvio_domini@brown.edu (F. Domini).

0042-6989/03/$ - see front matter
doi:10.1016/S0042-6989(03)00181-0

we must clarify how its purpose differs fundamentally from the purpose of the SFM algorithms developed in the field of Computer Vision.

2. Computational models of structure-from-motion

The SFM problem is one of the most studied problems in Computer Vision and, from that point of view, it can be described in the following terms. Given a sequence of photographic images of a fixed 3D scene, taken by a moving camera, the task is that of estimating (i) a 3D geometric model of the scene (structure), and (ii) the camera's motion parameters (which are often expressed in terms of the instantaneous translation and rotation of the camera). Some of the SFM modeling work deals with a moving camera and deriving the observer motion parameters. Much of the SFM work, however, assumes a stationary viewer and moving object, and recovers the 3D structure (and sometimes aspects of the 3D motion) of the object. Some of the computer algorithms proposed in the literature represent a scene as a set of discrete 3D feature points. In this simplified framework, the goal of the artificial SFM system is to derive the 3D coordinates of these feature points from their projected motion. Theoretical studies performed in the last two decades have almost completely solved the geometry of this problem. In fact, several geometric methods have been proposed to derive the veridical 3D configuration of a set of feature points from few orthographic or perspective views of a 3D object. These methods allow the recovery of the 3D projected structure by embedding in the interpretation process only a few assumptions about the relative motion between the camera and the feature points. The most common of these assumptions are, for example, rigidity (Hoffman & Bennett, 1985; Hoffman & Flinchbaugh, 1982; Ullman, 1979, 1984, 1986), fixed-axis motion (Bennett & Hoffman, 1985; Hoffman & Bennett, 1986; Ullman, 1979) and constant 3D angular velocity (Hoffman & Bennett, 1986; Ullman, 1979).

An important factor in the development of viable computational models (both for machine and human vision) is the presence of noise in the measurements of image motion in real-world situations. Different strategies have been adopted to design robust SFM systems, but two main categories of algorithms can be distinguished, based on how image motion is measured: (1) The algorithms based on feature-correspondence (e.g., Longuet Higgins, 1981; Tsai & Huang, 1984), and (2) the algorithms based on the optic-flow (e.g., Barron, Fleet, & Beauchemin, 1994; Subbarao, 1988; Subbarao, 1989; Subbarao & Waxman, 1986; Waxman, Kamgar-Parsi, & Subbarao, 1987). Most of the feature-based algorithms treats the SFM problem like the stereo problem: The 3D structure of the projected objects is derived from two perspective (e.g., Faugeras, 1993; Luong & Faugeras, 1996) or orthographic (e.g., Tomasi & Kanade, 1993) views of a set of feature points. The feature-correspondence methods are quite robust, but they are limited by the fact that the correspondence problem (matching the same feature-points in the two different views) is hard to solve (Oliensis, 1996). The algorithms belonging to the second category avoid the correspondence problem by measuring either the instantaneous velocity of projected feature points or spatial and temporal intensity gradients. Since the measured optic-flow is very small (being based on the maximum displacements of few pixels between successive frames—e.g., Barron et al., 1994), however, the algorithm output is more sensitive to input noise (for a discussion, see Chiuso, Brockett, & Soatto, 2000).

In order to reduce the influence of measurement noise on the SFM derivation, a series of methods have been proposed to improve the performance of both classes of algorithms. Most of these methods make use of temporal integration (Chiuso et al., 2000). Ullman (1984) has been one of the first researchers to suggest the idea that SFM may be better achieved by analyzing extended sequences of the projected objects. In most cases, his incremental-rigidity scheme converges to a veridical solution by using each new sequence-frame to improve a constantly updated internal 3D model, and Ando (1991) has recently proposed a more robust variation of Ullman's model by introducing a surface interpolation procedure in the 3D reconstruction process.

The algorithms that are more relevant for the following discussion are those that have analyzed the problem of planar surface motion (Subbarao & Waxman, 1986) and its representation through affine parameters (Koenderink & Van Doorn, 1991; Negahdaripour & Lee, 1992). The work of Koenderink and Van Doorn (1991), in particular, has prompted a large number of psychophysical investigations aimed at establishing whether the perceptual space obtained from the optical flow has an affine structure (e.g., Norman & Todd, 1993; Todd & Bressan, 1990; Todd, Oomes, Koenderink, & Kappers, 2001).

Attempts have also been made to develop computational models of human performance. One of these models has been proposed by Hildreth et al. (1995), and is inspired by Ullman's (1984) incremental-rigidity scheme (see also Treue, Andersen, Ando, & Hildreth, 1995). The model of Hildreth and collaborators is motivated by the psychophysical findings showing that the perceptual analysis of the optic flow (and the successive or concurrent reconstruction of 3D shape) occurs within an extensive temporal window (e.g., Atchley, Andersen, & Wuestefeld, 1998; Burr & Santoro, 2001; Caudek, Domini, & Di Luca, 2002; Eby, 1992; Hildreth et al., 1990; Todd, Akerstrom, Reichel, & Hayes, 1988; Treue et al., 1991; van Damme & van de Grind, 1996). Con-

sistent with these findings, Hildreth et al. assumed that 3D shape reconstruction is mediated by a process of temporal integration. After the initial measurement of the 2D image velocities, their model computes the 3D velocities so as to maximize the rigidity of the 3D configuration. From these 3D velocities, the model recovers the depths of the projected features. To cope with the error of the image-motion measurements, in a successive temporal-integration stage, the model averages the depths estimates over an extended time-period by means of an approach based on Kalman filtering (Anderson & Moore, 1979; Gelb, 1974). Finally, in a surface-reconstruction stage, the model fits a smooth 3D surface to the sparse depth values. The model of Hildreth et al. does not always converge on the veridical solution and, thus, it is able to account for several properties of human SFM. There are, however, some psychophysical findings that have been recently reported, which cannot be accounted for neither by Hildreth's model, nor by any other model based upon standard Computer Vision algorithms. These findings will be summarized in the next section.

3. Psychophysical findings incompatible with previous computational models

Three sets of psychophysical findings are relevant here. These findings show that (1) the metric properties of the perceived 3D objects are unrelated to the 3D Euclidean parameters of the projected objects; (2) the global perceived 3D structure cannot be represented by either Euclidean or affine geometry; (3) the assumptions used by most computational models to restrict the space of possible solutions of the SFM problem are not psychologically plausible.

An example of the first set of findings is provided by the research of Domini and Caudek on the perceptual interpretation of the optic flow produced by the orthographic projection of a rotating random-dot planar surface (see Domini & Caudek, in press; Proffitt & Caudek, 2002). In these studies, we have shown that perceived 3D shape depends only on the first-order temporal information provided by the optic flow, even when second-order information is available (see also Liter, Braunstein, & Hoffman, 1993; Todd & Bressan, 1990). We have shown, moreover, that perceived 3D orientation and angular velocity are not related to the simulated parameters, even when the stimuli are sufficient for any algorithm to derive a veridical solution (see also Todd, Tittle, & Norman, 1995). Finally, we have demonstrated that there is a unique nonlinear relationship between the local properties of the velocity field (first-order optic flow) and the perceived 3D properties. The perceptual solution, moreover, can be well accounted for by a maximum-likelihood interpretation of the stimulus displays (Domini & Caudek, 1999; see also Freeman, 1994).

An example of the second set of findings is provided by a study by Domini, Caudek, and Richman (1998). They asked observers to provide depth-order judgments relative to two probe dots positioned on a random-dot surface oscillating under orthographic projection about a fixed axis, and found that (1) the ordinal structure of the projected objects was not preserved in the relative-depth judgments of the observers; (2) lines that were parallel in the projected objects were not parallel in the perceived shapes; (3) the local signing of the perceived depth-order relations was incompatible with an internally consistent 3D structure. In a related study, Domini and Braunstein (1998) asked observers to judge the depth separation of two probe dots located on a random-dot planar surface. Also in that case, they found that the perceived 3D shape was internally inconsistent: Different paths of integration gave different results and the algebraic sum of the depth judgments along a closed path was not zero.

An example of the third set of findings is provided by a study by Domini et al. (1997) about the discrimination between rigid and nonrigid 3D motion. Consistent with what was found by other researchers (e.g., Griffiths & Zaidi, 1998; Hogervorst, Kappers, & Koenderink, 1997; Perotti, Todd, & Norman, 1996; Sparrow & Stine, 1998), Domini et al. found that the projection of rigid motion does not necessarily support the perception of a 3D rigidly moving object. This finding questions the hypothesis that a rigidity constraint is used to disambiguate the SFM problem, as assumed by several algorithms (e.g., Hoffman & Bennett, 1986; Ullman, 1979, 1984). Likewise, it has been shown that, in certain circumstances, human observers tend to perceive a fixed-axis rotation as a rotation about an axis with a changing orientation, or they tend to perceive a rotation about an axis with a changing orientation as a fixed-axis rotation (Caudek & Domini, 1998). This questions the fixed-axis assumption (e.g., Bennett & Hoffman, 1985; Hoffman & Bennett, 1986). Furthermore, Domini, Caudek, Turner, and Favretto (1998) showed that, in the appropriate stimulus conditions, human observers report perceiving constant 3D angular velocity as nonconstant, and nonconstant 3D velocity as constant. This questions the constant-3D-angular-velocity assumption (e.g., Hoffman & Bennett, 1985).

The findings described above suggests that human SFM should not be regarded as the veridical geometric solution to the inverse-projection problem, but rather as the most likely interpretation of the ambiguous stimulus-information provided by the first-order optic flow (Domini & Caudek, 1999). The perceptual solution to the SFM problem, moreover, cannot be represented in either Euclidean (Todd & Bressan, 1990) or affine space (Domini & Braunstein, 1998; Domini, Caudek, & Richman, 1998). Together, the findings described in this section question the idea that the Computer Vision

algorithms may be regarded as adequate models of human SFM. Even though an extension of Ullman's incremental rigidity scheme predicts some aspects of human performance (see Hildreth et al., 1995), it is difficult to imagine how any standard Computer Vision algorithm may be capable of accommodating all the properties of human SFM and, in particular, those concerning the violations of Euclidean and affine geometry. Computer Vision algorithms, in fact, are not intended to explain the limitations of human performance, but rather to provide a robust and unbiased 3D interpretation to a set of 2D projections. A model of human SFM, conversely, must account the specific properties (both accurate responses and biases) of human performance.

The model that we have previously put forward (Domini & Caudek, 1999) predicts the psychophysical findings described above, but it is limited in two respects. First, it implicitly assumes that the visual system is able to measure the properties of the instantaneous optic flow. In other words, it disregards the fact that the optic flow is measured by the visual system within an extended time period (e.g., Burr & Santoro, 2001; Caudek et al., 2002; Treue et al., 1995). Second, it does not account for the different perceptual interpretations that are derived from purely contracting or expanding flow fields (Domini et al., 2003; Domini et al., 2002). The purpose of the present paper is to modify our model so as to overcome the two limitations outlined above. In doing so, we will show that the different interpretations of contracting and expanding flow-fields follow naturally from the hypothesized strategy for measuring the relevant properties of the optic flow. Before discussing the revision of our model, however, we must briefly summarize the relevant properties of the instantaneous velocity field.

4. A maximum-likelihood approach for interpreting the instantaneous velocity field

Every smooth surface can be locally approximated by a planar patch. In a viewer-centered coordinate system, the 3D orientation of a planar patch can be described either in terms of the two depth-gradients g_x and g_y, or in terms of *slant* ($\sigma = (g_x^2 + g_y^2)^{1/2}$) and *tilt* ($\tau = a \times \tan(g_y/g_x)$). Slant is the tangent of the angle between the surface and the image plane and tilt is the direction in which the surface slopes (see Fig. 1).

If the planar patch undergoes a rotation of magnitude $\Delta\alpha$ about the y-axis, [1] the y coordinate of each point P on the surface remains constant, whereas the x coordinate of each point changes by an amount Δx according to the following equation (see Appendix A):

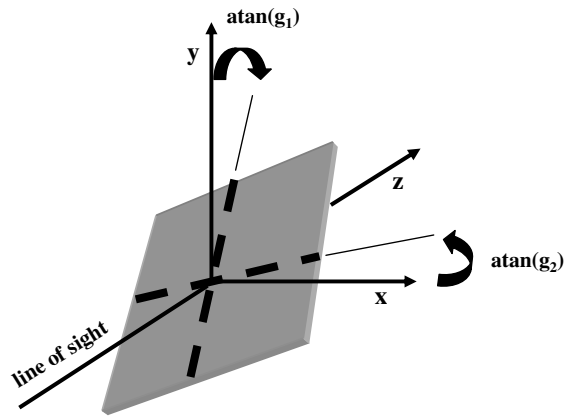

Fig. 1. Schematic representation of a planar surface with its center coinciding with the center of the (x, y, z) coordinate system. g_1 and g_2 are the horizontal and vertical components of the depth gradient, respectively.

$$\Delta x = (-g_x \sin(\Delta\alpha) + \cos(\Delta\alpha) - 1)x - g_y \sin(\Delta\alpha)y \quad (1)$$

Since we assume an orthogonal projection on the image plane [2] (defined by the axes x and y), Δx represents the 2D displacement of each projected point P. This vector field (which we will call displacement field) is linear in the image coordinates x and y, and its horizontal (Φ_x) and vertical (Φ_y) gradients are equal to:

$$\begin{aligned} \Phi_x &= -g_x \sin(\Delta\alpha) + \cos(\Delta\alpha) - 1 \\ \Phi_y &= -g_y \sin(\Delta\alpha) \end{aligned} \quad (2)$$

Eq. (1) should be compared with the equation that specifies the instantaneous velocity field (Domini & Caudek, 1999):

$$v_x = -g_x \omega x - g_y \omega y \quad (3)$$

where ω is the magnitude of the instantaneous 3D angular velocity, and v_x is the instantaneous 2D velocity of the projected point P. Eq. (3) is also linear in the image coordinates x and y, and $\varphi_x = -g_x\omega$, $\varphi_y = -g_y\omega$ are the vertical and horizontal velocity gradients, respectively. Since these gradients depend on the choice of the coordinate system, a more convenient description of the velocity field is in terms of deformation (*def*) calculated as follow:

$$def = \sqrt{\varphi_x^2 + \varphi_y^2} \quad (4)$$

[1] It can be shown that a rotation about an arbitrary axis can always be reduced to a rotation about an axis parallel to the image plane by removing the rotation component about the z-axis (see Todd & Bressan, 1990). When this is done, the y-axis can be chosen to be orthogonal to the direction of the velocity vectors.

[2] Since the field of view of our stimuli was only about 9° of visual angle, orthographic projection is a good approximation of the more "realistic" perspective projection. In fact, perspective information in motion displays becomes effective only for very large fields of view (e.g. Cornilleau-Peres et al., 2002).

The deformation component, however, does not allow a unique determination of the parameters of the projected surface and of its 3D motion. It has been shown, in fact, that *def* is the product of a one-parameter family of different combinations of σ and ω:

$$def = \sigma\omega \tag{5}$$

(see Domini & Caudek, 1999). Even though it is ambiguous, however, the *def* component has been shown to be the main determinant of perceived surface slant and angular velocity from orthographic projections of moving objects (e.g., Caudek & Domini, 1998; Caudek & Rubin, 2001; Domini & Braunstein, 1998; Domini & Caudek, 1999; Domini et al., 1997; Domini, Caudek, & Richman, 1998; Domini, Caudek, Turner, et al., 1998; Todd & Perotti, 1999). These findings, therefore, raise the problem of determining how the visual system is able to uniquely relate *def* to perceived surface orientation and angular velocity.

In order to address this issue, Domini and Caudek have shown that not all possible σ, ω pairs are equally likely to have produced a given *def* (if certain assumptions are introduced in the interpretation process). If uniform probability distributions are assumed for σ and ω, for example, then the conditional probability of a σ, ω pair given *def* is not uniform, but it has a maximum. The maximum value of $p(\sigma, \omega|def)$ is associated with the σ^*, ω^* pair for which: [3]

$$\begin{aligned} \sigma^* &= \frac{1}{k_\omega}\sqrt{def} \\ \omega^* &= k_\omega\sqrt{def} \end{aligned} \tag{6}$$

Consistent with this analysis, Domini and Caudek (1999) hypothesized that the visual system recovers surface orientation and angular velocity by determining the σ, ω pair to which is associated the maximum conditional probability for *def*. Their empirical data, together with those of other investigators (e.g., Todd & Perotti, 1999) support this hypothesis.

5. The revised model

The model previously described assumes that the visual system has access to the instantaneous properties of the optic flow. This assumption is implausible, however, since it has been shown that a biological system requires an extended time-window to measure the properties of a velocity field (e.g., Burr & Santoro, 2001; Caudek et al., 2002; Treue et al., 1995). In order to overcome this limitation, in the revision that we propose here for our model, we assume that the actual input for the perceptual analysis is provided by the displacement field (i.e., the measurement of the image features at the times t_0 and $t_0 + \Delta t$). On the basis of this information, then, we estimate the deformation component at the time $t_0 + \Delta t$. By using the estimated deformation, subsequently, we compute the local surface-slant and angular-velocity by means of the maximum-likelihood approach described by Domini and Caudek (1999).

Let us examine, therefore, the problem of how the deformation component may be estimated by using the displacement field. We propose that the visual system solves this problem through a process of statistical learning. We hypothesize that, by means of a statistical analysis of the visual input, the visual system learns to associate the expected value of *def* at the time $t_0 + \Delta t$ to the gradients of the displacement field defined by the time-interval Δt. We will not speculate here how this learning process takes place. Instead, we will show that, by making only a few assumptions about the properties of the parameters involved in this statistical analysis, it is possible to derive from Φ_x and Φ_y an estimate of the expected value of the deformation that is consistent with the qualitative pattern of the empirical data obtained in the present investigation.

In the Monte Carlo simulation that we have performed, the following assumptions were introduced:

(1) The slant (σ) and tilt (τ) specifying the orientation of a planar patch relative to the observer were uniformly distributed in the intervals [0°, 70.0°] and [0°, 90°], respectively (see Domini & Caudek, 1999).
(2) The average angular velocity $\bar{\omega}$ during the time interval Δt was uniformly distributed in the range [−2.1, 2.1] rad/s ($\Delta\alpha$ ranging from −120° to 120°, approximately).
(3) Due to the measurement noise, the gradients of the displacement field were defined within the uncertainty intervals $[\Phi_x, \Phi_x + \Delta\Phi_x]$ and $[\Phi_y, \Phi_y + \Delta\Phi_y]$.

On the basis of the previous assumption, the best estimate of *def* at the time $t_0 + \Delta t$ corresponds to the expected value of *def*, given that the gradients of the displacement field fall in the uncertainty intervals specified above:

$$\begin{aligned} \hat{def}(t_0 + \Delta t) = E[def(t_0 + \Delta t)|\Phi_{x0} \leqslant \Phi_x < \Phi_{x0} + \Delta\Phi_x, \Phi_{y0} \\ \leqslant \Phi_y < \Phi_{y0} + \Delta\Phi_y] \end{aligned} \tag{7}$$

In order to estimate the expected value of the deformation component, we ran a computer simulation in which we randomly selected 5,000,000 values for $g_x(t_0)$, $g_y(t_0)$ and $\bar{\omega}$, according to their hypothesized a priori probability distributions. For each triplet $[g_x(t_0), g_y(t_0), \bar{\omega}]$, we

[3] In the Monte Carlo simulation reported below, k_ω takes on the value of 0.68. The value given to k_ω was estimated empirically (on a different data-set) by Domini and Caudek (1999), who used a least square procedure to minimize the squared differences between the magnitudes of σ and ω reported by the observers and those predicted according to Eq. (6).

calculated the corresponding values Φ_x and Φ_y of the displacement field through Eq. (2) by specifying $\Delta\alpha = \bar{\omega}\Delta t$. We then sampled the Φ_x, Φ_y pairs falling within the intervals $[\Phi_{x0} \leqslant \Phi_x < \Phi_{x0} + \Delta\Phi_x]$ and $[\Phi_{y0} \leqslant \Phi_y < \Phi_{y0} + \Delta\Phi_y]$, where Φ_{x0}, Φ_{y0} are the values of the gradients of the displacement field used in the present experiments, and $\Delta\Phi_x$, $\Delta\Phi_y$ are the assumed levels of measurement noise. To each of these pairs corresponds a triplet $[g_x(t_0), g_y(t_0), \bar{\omega}]$, and each of these triplets allows us to compute an estimate of the instantaneous velocity gradients at the time $t_0 + \Delta t$:

$$\hat{\varphi}_x(t_0 + \Delta t) = \varpi \frac{g_{x0}\cos(\varpi(t_0 + \Delta t)) + \sin(\varpi(t_0 + \Delta t))}{\cos(\varpi(t_0 + \Delta t)) - g_x \sin(\varpi(t_0 + \Delta t))} \tag{8}$$

$$\hat{\varphi}_y(t_0 + \Delta t) = \varpi \frac{g_{y0}}{\cos(\varpi(t_0 + \Delta t)) - g_x \sin(\varpi(t_0 + \Delta t))} \tag{9}$$

From each pair $\hat{\varphi}_x(t_0 + \Delta t)$, $\hat{\varphi}_y(t_0 + \Delta t)$, then, we can obtain an estimate of the deformation component at the time $t_0 + \Delta t$:

$$\hat{def}(t_0 + \Delta t) = \sqrt{\hat{\varphi}_x^2(t_0 + \Delta t) + \hat{\varphi}_y^2(t_0 + \Delta t)} \tag{10}$$

In this manner, we define a distribution of *def* magnitudes. The average of this distribution, therefore, corresponds to an estimate of the expected value of the deformation component, given that g_x, g_y and $\bar{\omega}$ are distributed as indicated above, and given the uncertainty level $\Delta\Phi_x$, $\Delta\Phi_y$, for the measurement of the displacement field (see Appendix A). Before presenting the results of this simulation, however, we must distinguish between expanding and contracting flow fields.

6. Expanding and contracting flow fields

The stimuli used in the simulation were generated by the factorial combination of five magnitudes of the gradients of the displacement-field (Φ_x and Φ_y). The magnitudes of these gradients matched the magnitudes of the average velocity-fields used in the actual experiments: −0.36, −0.09, 0, 0.09, 0.36 s^{-1}. Nine of the 24 flow fields so generated are reported in a schematic form in Fig. 2 (the condition $\Phi_x = 0$, $\Phi_y = 0$ was obviously discarded).

Note that the flow is expanding when $\Phi_x \geqslant 0$, and contracting when $\Phi_x < 0$. An expanding flow is produced by a rotation of a planar surface toward the fronto-parallel plane; a contracting flow is produced by a rotation of a planar surface away from the fronto-parallel plane.

What is interesting to note for the present purposes is what happens when the output of our revised model is plotted against $|\Phi_x|$, for each of the five values of Φ_y used

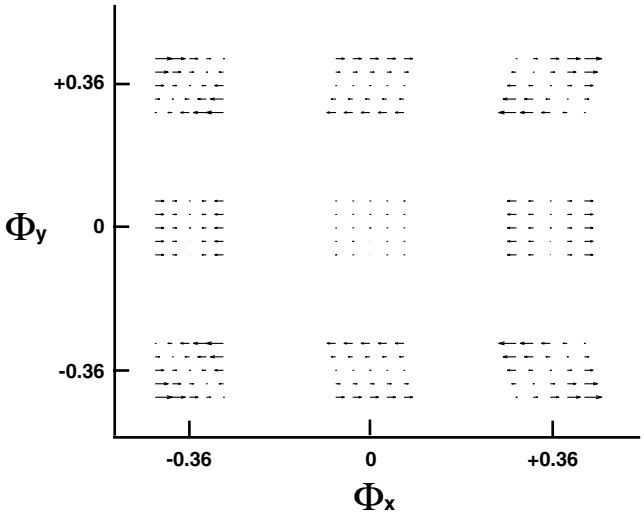

Fig. 2. Schematic representation of nine linear velocity fields generated by the factorial combination of the gradients −0.36, 0.0, 0.36 s^{-1}. Each arrow represents an instantaneous velocity vector V.

in the present experiments. Fig. 3 indicates that the relative variation of the estimated expected value of *def* is smaller for expanding than for contracting flow fields. If the visual system estimates the deformation component in a manner that is consistent with the model discussed above, therefore, we should expect that observers judge surface orientation and angular velocity differently when presented with displays that, in principle, convey the same information about surface orientation and angular velocity, but through expanding vs. contracting motion.

This prediction can be contrasted with what is expected according to a model which makes use of the instantaneous deformation. In this second case, in fact, the magnitude of *def* does not change for contracting and expanding flow-fields (see Eq. (4)). Consequently, the same surface orientation and angular velocity should be perceived for contracting and expanding flow fields.

Two more things must be highlighted in the results of our simulation. First, contracting and expanding fields produce similar estimated deformations when Δt (the time-window necessary for measuring the optic flow) is very small (50 ms—see Fig. 3, left panel). The relative-variation of the estimated *def* is larger for contracting than for expanding flow-fields, conversely, for Δt of at least 150 ms. This result is consistent with the psychophysical data indicating that the visual system needs at least 150 ms for obtaining reliable measures of the optic flow (e.g., Caudek et al., 2002). Second, the asymmetry of the model's output for contracting and expanding fields is very robust to noise. The qualitative pattern of results reported in Fig. 4, in fact, does not change even when the measurement uncertainty takes on enormous values (see Fig. 4, right most panel).

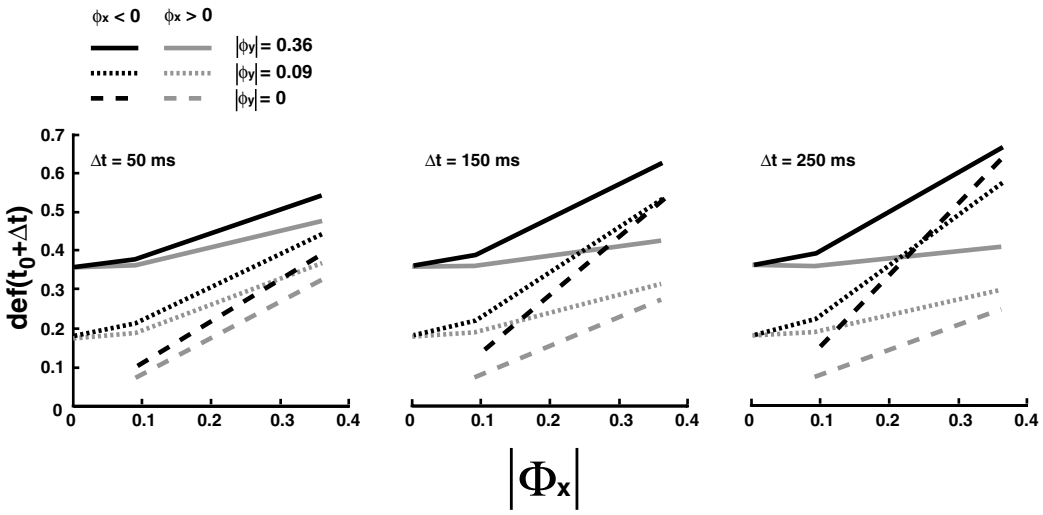

Fig. 3. Estimated deformation component as a function of the absolute value of ϕ_x generated by the factorial combination of the velocity gradients -0.36, 0.0, 0.36 s^{-1} for three time-windows: $\Delta t = 50$ ms (left panel), $\Delta t = 150$ ms (central panel), $\Delta t = 250$ ms (right panel). These gradients represent a subset of those used in Experiments 1 and 2.

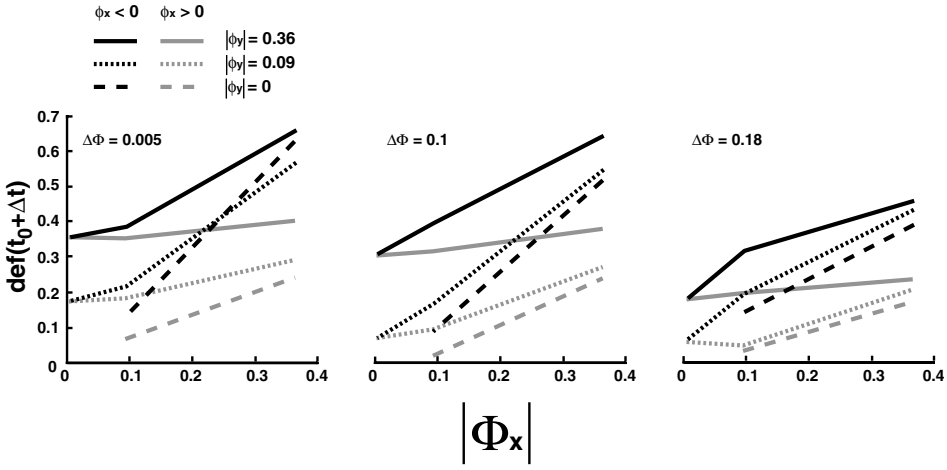

Fig. 4. Estimated deformation component for three levels of measurement uncertainty ($\Delta\Phi = 0.001$, 01, 0.18). Even when the interval within which the displacement gradients are measured is equal to half of the range used for generating the stimuli of the present experiments, the range of variation of the expected value of *def* remains larger for contracting flow fields (right panel).

In the following experiments we tested the predictions of our revised model by measuring perceived slant (Experiment 1) and perceived angular velocity (Experiment 2) for expanding or contracting flow fields. The same horizontal and vertical velocity gradients were used in both experiments.

7. Experiment 1

In Experiment 1 observers were asked to report the apparent slant of a simulated rotating surface. The stimuli consisted of constant optic-flow fields, and were generated by a factorial combination of five horizontal and vertical velocity gradients. According to the predictions of our revised model (see Fig. 3), we expected a larger sensitivity for contracting flow fields.

7.1. Method

Participants. Eight paid observers, recruited from the Brown University community and naive to the purpose of the experiment participated in this investigation. All participants had normal or corrected-to-normal vision.

Stimuli. The stimuli were moving high-luminance random-dots presented on a low-luminance background on a monitor. The motion of the dots defined a linear velocity field in which the dots were translated only in the horizontal direction, with the displacement Δx being equal to: $\Delta x = \phi_x^* x + \phi_y^* y$ (see Section 1), where the x, y

coordinate-system had the origin at the center of the monitor screen. The horizontal (ϕ_x^*) and vertical (ϕ_y^*) gradients of the average velocity field could take on five values: −0.36, −0.09, 0, 0.09 and 0.36. Each stimulus sequence, lasting 1336 ms, was repeated until an observer key press indicated the beginning of the next trial. There was a temporal gap of one second between each repetition of the sequence. Being constant over time, our stimuli do not represent a natural physical stimulus motion. Nevertheless, observers report perceiving them like surfaces whose slant increases over time (see Domini et al., 2002).

A probe was shown at the center of the screen within a circular gap of the stimulus display. The gap had a diameter of 2.9° of visual angle. The probe, similar to that used by Domini and Caudek (1999), depicted a hemisphere specified by 14 meridians and 7 parallels, and subtended 2.4° of visual angle. The slant and tilt of the base of the hemispherical probe could be adjusted by the participants by means of a mouse connected to the workstation.

Apparatus. The displays were presented on a high-resolution color monitor (1280×1024 addressable locations) under the control of a Hewlett-Packard Visualize X550 Workstation. The screen had a refresh rate of 60 Hz. The participants sat approximately 90 cm away from the screen and viewed the displays monocularly through a reduction screen placed approximately 2 cm away from the monitor. The circular aperture of the reduction screen limited the visible portion of the monitor to a region with a diameter of approximately 8.9° of visual angle. In the experiment, the moving pattern of the stimuli covered the full 8.9° visible area of the display. Approximately 500 dots were visible though this aperture. Dot density was kept constant, and dots were randomly removed or added during the sequence to fulfill this constraint. A chin-rest was used to restrict head movement. The experiment was run in a dark room.

Design. The two within-participants variables were the intensities (−0.36, −0.09, 0, 0.09 and 0.36) of the horizontal and vertical gradients of the velocity field. Each participant viewed five presentations of the 25 different experimental conditions in five different blocks.

Procedure. Participants were asked to judge the perceived orientation of the rotating planar surface evoked by the linear velocity field. The task was performed by adjusting the slant of the base of the hemispherical probe until it was perceived to be parallel to the 3D surface. The tilt of the hemispherical probe was kept fixed, and it was determined by the ratio between the vertical and horizontal gradients of the velocity field. The slant of the base of the probe was varied by pressing the left and right buttons of a mouse connected to the workstation. Time was not restricted. No feedback was provided.

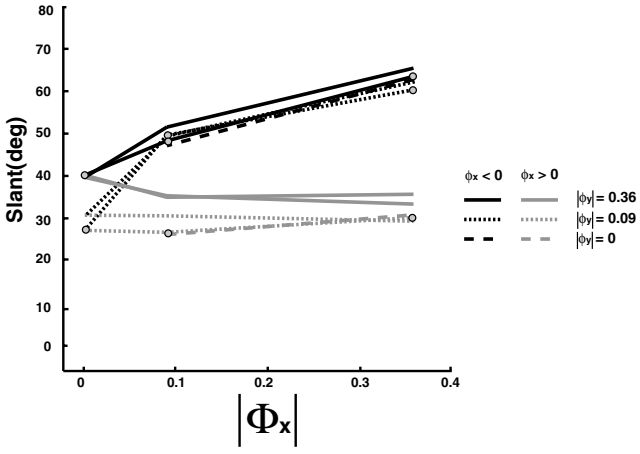

Fig. 5. The average judged slant in Experiment 1 as a function of the absolute value of ϕ_x, for each of the five values ϕ_y. Circles indicate negative values of ϕ_y.

7.2. Results and discussion

Fig. 5 reports the average judged slant magnitudes as a function of the absolute value of the horizontal gradient (ϕ_x^*) for the different values of the vertical gradient (ϕ_y^*). In order to test the hypothesis that observers are less sensitive to the velocity gradients for expanding flow fields, we performed a linear regression analysis on the data of Fig. 5 by using as the independent variables the horizontal velocity gradient, a dummy regressor coding for expanding vs. contracting flows, and an interaction regressor coding for the different slope in the two groups. R^2 for the fitted model was equal to 0.91 [$F(3,24) = 79.53$, $p < 0.001$]. Consistent with our hypothesis, the slope of the fitted regression line for contracting flow fields was significantly higher than the slope for the expanding flow fields [$F(1,24) = 63.046$, $p < 0.001$].

8. Experiment 2

Experiment 2 replicated the same design of the previous experiment, the only difference being that observers were asked to report the apparent angular velocity of the simulated rotating surface. Also in this case, we expected a larger sensitivity for contracting flow fields.

8.1. Method

Participants. Eight paid observers, recruited from the Brown University community and naive to the purpose of the experiment participated in this investigation. All participants had normal or corrected-to-normal vision.

Stimuli. The stimuli were identical to those of the first experiment, except for the fact that the probe used in the

previous experiment was replaced by a sphere. The sphere subtended 2.4° and was specified by a wire-frame depicting 14 meridians and 14 parallels. The angular rotation of the sphere could be varied.

Apparatus. The apparatus was identical to that used in the first experiment.

Design. The design was identical to that of the previous experiment.

Procedure. Participants were asked to judge the perceived angular velocity of a 3D random-dot rotating planar surface by adjusting the angular velocity of the central wire-frame sphere. Participants were asked to judge the perceived 3D angular velocity of the rotating planar surface evoked by the linear velocity field. The task was performed by adjusting the angular velocity of the spherical probe until it was perceived to rotate by the same angular velocity as the random-dot 3D surface. The angular velocity of the probe was varied by pressing the left and right buttons of a mouse connected to the workstation. Time was not restricted. No feedback was provided.

8.2. Results and discussion

Fig. 6 reports the average judged 3D angular velocity magnitudes as a function of the absolute value of the horizontal gradient (ϕ_x^*) for the different values of the vertical gradient (ϕ_y^*). As for Experiment 1, to test the hypothesis that observers are less sensitive to the velocity gradients for expanding flow fields, we performed a linear regression analysis on the data of Fig. 6 by using as the independent variables the horizontal velocity gradient, a dummy regressor coding for expanding vs. contracting flows, and an interaction regressor coding for the different slope in the two groups. R^2 for the fitted model was equal to 0.64 [$F(3,24) = 13.97$, $p < 0.001$]. Consistent with our hypothesis, the slope of the fitted regression line for contracting flow fields was significantly higher than the slope for the expanding flow fields [$F(1,24) = 5.822$, $p < 0.05$].

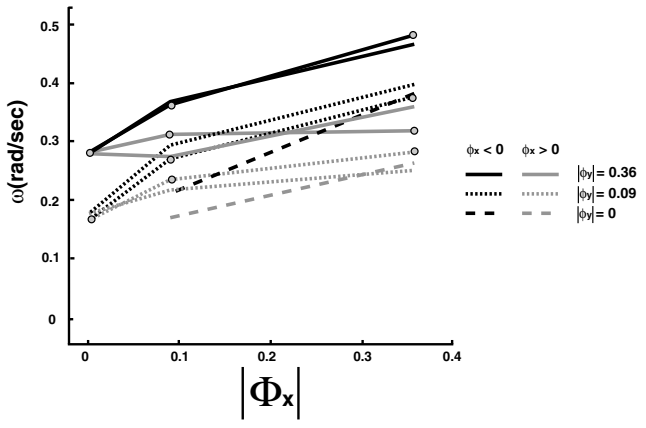

Fig. 6. The average judged angular velocity in Experiment 2 as a function of the absolute value of ϕ_x, for each of the five values ϕ_y. Circles indicate negative values of ϕ_y.

9. General discussion

A growing body of empirical evidence has recently shown that the perceptual derivation of 3D information from motion is based primarily on the information provided by the first-order velocity field (e.g., Braunstein et al., 1990; Domini et al., 1997; Lappin et al., 1980; Liter et al., 1993; Rubin, Hochstein, & Solomon, 1995; Todd & Bressan, 1990). It has also been shown mathematically that a veridical recovery of the slant (σ) and angular velocity (ω) of a projected surface requires the analysis of the second-order temporal properties of the velocity field (i.e., acceleration). Being a first-order property, in fact, *def* defines a one-parameter family of different combinations of surface orientations and angular velocities (see Eq. (5)). Regardless of this inherent ambiguity, however, also in the absence of second-order temporal properties, human observers provide consistent judgments of surface-slant and angular-velocity magnitudes from a velocity field (e.g., Domini & Caudek, 1999; Todd & Perotti, 1999). Domini and Caudek (1999) explained this phenomenon by hypothesizing that, in the absence of better information, the visual system interprets the optic flow by choosing the surface slant and angular velocity that have the maximum conditional probability, given *def*.

The model of perceived SFM proposed by Domini and Caudek (1999), however, is limited in two respects. First, it makes use of the properties of the instantaneous optic flow, whereas it is obvious that the visual system has access only to those properties that can be measured within an extended time-window. Second, it is inconsistent with the psychophysical results showing that different perceptual interpretations are obtained from contracting or expanding flow fields (Domini et al., 2003; Domini et al., 2002). In trying to overcome these limitations, in the present paper we revised our previous model by determining how *def* may be estimated from the displacement field. Having estimated *def*, surface slant and angular velocity are then derived as indicated by Domini and Caudek (1999).

One important consequence of our revised model is that a different interpretation is obtained from expanding and contracting flow fields: As is shown in Fig. 3, in fact, the relative-variation of the estimated *def* magnitudes is smaller for expanding than for contracting flow fields. This prediction of our revised model has been tested by asking observers to judge the slant (Experiment 1) and 3D angular velocity (Experiment 2) of expanding or contracting flow fields. Consistent with the predictions of our model, we found that observers were less sensitive to expanding motion (see Figs. 5 and 6).

It is important to clarify how, in our revised model, the notion of "statistical learning" relates to the estimation of the deformation component. Even though the Monte Carlo simulation described in the introduction makes use of the distal parameters g_x and g_y (defining the depth gradients of the projected surface) to estimate the expected value of *def*, this does not mean that we hypothesize that the visual system has access to these (unknown) parameters. If g_x and g_y were already known, in fact, there would no need to estimate *def* in order to compute surface orientation and angular velocity. With the simulation described in the introduction, instead, we intend to point out the fact that, if certain assumptions about the distributions of g_x and g_y are satisfied, then the expected value of *def* can be estimated on the basis of the displacement field alone. At the core of our revised model, therefore, we hypothesize that an estimate of *def* is associated with a given displacement field through a process of learning, as a consequence of repeated exposition in the course of perceptual experience.

What makes the previous hypothesis appealing is that, in order to compute estimates of *def* that are consistent with our empirical data, it is necessary to satisfy only two assumptions: (1) the orientation and average velocity of the projected surfaces are uniformly distributed, and (2) the gradients of the displacement field are measured with the uncertainty intervals $[\Phi_x, \Phi_x + \Delta\Phi_x]$ and $[\Phi_y, \Phi_y + \Delta\Phi_y]$.

In conclusion, in the present study we described a probabilistic model for deriving surface orientation and 3D motion from a linear displacement field. The proposed model addresses the problems of how the visual system may (1) estimate the properties of the optic flow, and (2) derive the orientation and the angular velocity of the projected surface from these estimated properties. The predictions of our probabilistic model are consistent with the present empirical results (i.e., different perceived surface orientations and angular velocities for expanding and contracting velocity fields), as well as with previous psychophysical findings (e.g., Caudek & Domini, 1998; Domini & Caudek, 1999; Domini et al., 1997; Domini, Caudek, & Richman, 1998; Domini, Caudek, Turner, et al., 1998; Todd & Perotti, 1999).

Acknowledgement

This research was supported by National Science Foundation grant BCS-78441.

Appendix A

If a planar surface $z(x, y) = g_x x + g_y y$ rotates about the vertical axis by an amount $\Delta\alpha$, the new x and z coordinates of a generic point $P(x, y, z)$ on the surface will be:

$$\begin{aligned} x' &= x\cos(\Delta\alpha) - z\sin(\Delta\alpha) \\ z' &= x\sin(\Delta\alpha) + z\cos(\Delta\alpha) \end{aligned} \tag{A.1}$$

or

$$\begin{aligned} x' &= x\cos(\Delta\alpha) - (g_x x + g_y y)\sin(\Delta\alpha) \\ z' &= x\sin(\Delta\alpha) + (g_x x + g_y y)\cos(\Delta\alpha) \end{aligned} \tag{A.2}$$

if z is replaced with the equation of a planar surface. It is then obvious that Eq. (3) can then be obtained from (A.2) by calculating $\Delta x = x' - x'$.

After the rotation through $\Delta\alpha$, the values of the gradients of the planar surface change. If x as described by (A.2) is substituted in the equation for z' in (A.2), the equation for z' becomes:

$$z' = \frac{g_x\cos(\Delta\alpha) + \sin(\Delta\alpha)}{\cos(\Delta\alpha) - g_x\sin(\Delta\alpha)}x' + \frac{g_y}{\cos(\Delta\alpha) - g_x\sin(\Delta\alpha)}y' \tag{A.3}$$

Note that y' is equal to y, since the rotation is about the vertical axis. The coefficients for x' and y' represent the values of the horizontal and vertical depth gradients after the rotation. If the depth gradients are multiplied by the instantaneous angular velocity (see Eq. (3)) the instantaneous velocity gradients are obtained (Eqs. (8) and (9)).

References

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Englewood Cliffs, NJ: Prentice-Hall.

Ando, H. (1991). Dynamic reconstruction of 3D structure and 3D motion. Proceedings of the IEEE Workshop on Visual Motion (pp. 101–110) Princeton, NJ, October.

Atchley, P., Andersen, G. J., & Wuestefeld, A. P. (1998). Cooperativity, priming, and 3-D surface detection from optic flow. *Perception and Psychophysics, 60*, 981–992.

Barron, J., Fleet, D., & Beauchemin, S. (1994). Performance of optic flow techniques. *International Journal of Computer Vision, 12*, 43–78.

Bennett, B. M., & Hoffman, D. D. (1985). The computation of structure from fixed-axis motion: Nonrigid structures. *Biological Cybernetics, 51*, 293–300.

Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1990). Discriminating rigid from non-rigid motion: minimum points and views. *Perception and Psychophysics, 47*, 205–214.

Braunstein, M. L., Hoffman, D. D., Shapiro, L. R., Andersen, G. J., & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 335–343.

Burr, D. C., & Santoro, L. (2001). Temporal integration of optic flow, measured by contrast and coherence thresholds. *Vision Research, 41*, 1891–1899.

Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 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.

Chiuso, A., Brockett, R., & Soatto, S. (2000). Optimal structure from motion: local ambiguities and global estimates. *International Journal of Computer Vision, 39*, 195–228.

Cornilleau-Peres, 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.

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*(2), 426–444.

Domini, F., & Caudek, C. (in press). Submitted chapter.

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*, 1164–1174.

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

Domini, F., Caudek, C., Turner, A., & Favretto, A. (1998). Discriminating constant from variable angular velocities in the kinetic depth effect. *Perception and Psychophysics, 60*, 747–760.

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

Eby, D. W. (1992). The spatial and temporal characteristics of perceiving 3-D structure from motion. *Perception and Psychophysics, 51*, 163–178.

Faugeras, O. (1993). *Three dimensional vision, a geometric viewpoint*. Cambridge, MA: MIT Press.

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature, 368*, 542–545.

Gelb, A. (Ed.). (1974). *Applied optimal estimation*. Cambridge, MA: MIT Press.

Griffiths, A. F., & Zaidi, Q. (1998). Rigid objects that appear to bend. *Perception, 27*, 799–802.

Hildreth, E. G., Ando, H., Andersen, R. A., & Treue, S. (1995). Recovering three-dimensional structure from motion with surface reconstruction. *Vision Research, 35*, 117–137.

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

Hoffman, D. D., & Bennett, B. M. (1985). Inferring the relative 3-D position of two moving points. *Journal of the Optical Society of America, A2*, 350–353.

Hoffman, D. D., & Bennett, B. M. (1986). The computation of structure from fixed-axis motion: rigid structures. *Biological Cybernetics, 54*, 71–83.

Hoffman, D. D., & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics, 42*, 195–204.

Hogervorst, M. A., Kappers, A. M., & Koenderink, J. J. (1997). Monocular discrimination of rigidly and nonrigidly moving objects. *Perception and Psychophysics, 59*, 1266–1279.

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

Koenderink, J. J., & Van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America, A8*, 377–385.

Landy, M. S., Dosher, B. A., Sperling, G., & Perkins, M. E. (1991). The kinetic depth effect and optic flow—II. First- and second-order motion. *Vision Research, 31*, 859–876.

Lappin, J. S., Doner, J. F., & Kottas, B. L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science, 209*, 717–719.

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

Longuet Higgins, H. C. (1981). Artificial intelligence—a new theoretical psychology? *Cognition, 10*, 197–200.

Luong, Q. T., & Faugeras, O. (1996). The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision, 17*, 43–76.

Negahdaripour, S., & Lee, S. (1992). Motion recovery from images sequences using only first order optical flow information. *International Journal of Computer Vision, 9*, 163–184.

Norman, J. F., & Todd, J. T. (1993). The Perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception and Psychophysics, 53*, 279–291.

Oliensis, J. (1996). Provably correct algorithms for multi-frame structure from motion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Perotti, V. J., Todd, J. T., & Norman, J. F. (1996). The visual perception of rigid motion from constant flow fields. *Perception and Psychophysics, 58*, 666–679.

Proffitt, D. R., & Caudek, C. (2002). Depth perception and the perception of events. In A. F. Healy, R. W. Proctor, & I. E. Weiner (Eds.), *Handbook of psychology: experimental psychology*. Wiley.

Rubin, N., Hochstein, S., & Solomon, S. (1995). Restricted ability to recover three-dimensional global motion from one-dimensional motion signals: psychophysical observations. *Vision Research, 35*, 463–476.

Sparrow, J. E., & Stine, W. W. (1998). The perceived rigidity of rotating eight-vertex geometric forms: extracting nonrigid structure from rigid motion. *Vision Research, 38*, 541–556.

Subbarao, M. (1988). Interpretation of image flow: rigid curved surfaces in motion. *International Journal of Computer Vision, 2*, 77–96.

Subbarao, M. (1989). Interpretation of image flow: a spatio-temporal approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*, 266–278.

Subbarao, M., & Waxman, A. M. (1986). Closed form solutions to image flow equations for planar surfaces in motion. *Computer Vision Graphics, and Image Processing, 36*, 208–228.

Todd, J. T., Akerstrom, R. A., Reichel, F. D., & Hayes, W. (1988). Apparent rotation in three-dimensional space: effects of temporal, spatial, and structural factors. *Perception and Psychophysics, 43*, 179–188.

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

Todd, J. T., Oomes, A. H. J., Koenderink, J. J., & Kappers, A. M. L. (2001). On the affine structure of perceptual space. *Psychological Science, 12*, 191–196.

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

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

Tomasi, C., & Kanade, T. (1993). Shape and motion from image streams: a factorization method. *Proceedings of the National Academy of Sciences USA, 90*, 9795–9802.

Treue, S., Andersen, R. A., Ando, H., & Hildreth, E. C. (1995). Structure-from-motion: perceptual evidence for surface interpolation. *Vision Research, 35*, 139–148.

Treue, S., Husain, M., & Andersen, R. A. (1991). Human perception of structure from motion. *Vision Research, 31*, 59–75.

Tsai, R. Y., & Huang, T. S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6*, 13–26.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London, B, Biological Sciences, 203*, 405–426.

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

Ullman, S. (1986). The optical flow of planar surfaces. *Spatial Vision, 1*, 263–276.

van Damme, W. J., & van de Grind, W. A. (1996). Non-visual information in structure-from-motion. *Vision Research, 36*, 3119–3127.

Waxman, A. M., Kamgar-Parsi, B., & Subbarao, M. (1987). Closed-form solutions to image flow equations for 3-D structure and motion. *International Journal of Computer Vision, 1*, 239–258.