

Temporal Integration in Structure From Motion

Fulvio Domini and Quoc C. Vuong
Brown University

Corrado Caudek
University of Trieste

A temporal integration model is proposed that predicts the results reported in 4 psychophysical experiments. The main findings were (a) the initial part of a structure-from-motion (SFM) sequence influences the orientation evoked by the final part of that sequence (an effect lasting for more than 1 s), and (b) for oscillating SFM sequences, perceived slant is affected by the oscillation frequency and by the sign of the final gradient. For contracting optic flows (i.e., rotations away from the image plane), the sequence with the lowest oscillation frequency appeared more slanted; for expanding optic flows (i.e., rotations toward the image plane), the sequence with the highest oscillation frequency appeared more slanted.

The continuous change in the retinal projections produced by the relative motion between an observer and the environment is a very important source of information for the 3-D structure of the world. A way to characterize the dynamic properties of retinal projections is to describe them in terms of a pattern of moving features. We refer to this pattern as the *optic flow* (Gibson, 1950). Several investigations have revealed that perceptual analysis of the optic flow is not instantaneous but, rather, is performed over an extensive temporal window. Treue, Husain, and Andersen (1991) asked observers to discriminate between structured (cylinder) and unstructured (noise) random-dot displays with a limited dot lifetime. They found that performance improved over several dot lifetimes, thus indicating that some form of temporal integration occurs for the detection of a 3-D surface from the optic flow. In particular, they found a relatively constant point lifetime threshold (50–85 ms) for perceiving structure from motion (SFM) and long reaction times for detecting the structure (approximately 1 s).

Atchley, Andersen, and Wuestefeld (1998) showed observers optic-flow displays with an increasing (the displays began with few dots, and the number of dots increased as time went by) or decreasing (the number of dots decreased as time went by) texture density. Observers had to report when a 3-D surface was perceived (increasing condition) or when the surface disappeared (decreasing condition). A hysteresis effect was found: Detection thresholds were lower for the decreasing texture condition. These previous experiments indicate that the recovery of a 3-D surface from the optic flow builds up in time, thus revealing that human SFM is the product of a form of short-term temporal integration: A

3-D surface representation is generated gradually and is completed after about 200 ms (e.g., Atchley et al., 1998; Eby, 1992; Treue et al., 1991; van Damme & van de Grind, 1996). These experiments also show that a dynamic change in 3-D structure may require more than 1,000 ms to be detected (Treue et al., 1991).

These previous investigations describe a process of surface recovery from the optic flow. What has not been studied so far is whether the 3-D properties recovered from the optic flow are affected by some form of long-term temporal integration after an early stage of surface formation. The idea of a second integration stage following the stage of surface recovery is consistent with a recent proposal of Burr and Santoro (2001). By asking observers to discriminate upward versus downward translations, clockwise versus counterclockwise rotations, and expanding versus contracting radial optic-flow patterns, Burr and Santoro found evidence of two stages of analysis: an early local-motion analysis with a time constant of 200–300 ms and a later global-motion integration stage with a time constant of about 3,000 ms. In line with these findings, we propose that the 3-D properties that are recovered from the optic flow are the product of two stages of analysis: a first stage of about 200 ms having the purpose of creating an initial 3-D surface representation (*surface recovery*) and a second stage having the purpose of updating the perceived 3-D representation. In this second stage of analysis (*surface-orientation update*), we propose that the current 3-D surface orientation is computed by taking into account the 3-D motion that has been attributed to the same surface in previous moments of time.

It is important to note that the present investigation does not concern the process by which the optic flow is measured but, rather, concerns the temporal evolution of the 3-D information that has been recovered from the optic flow. Empirical evidence distinguishing temporal integration in optic-flow measurement from temporal integration in 3-D surface-orientation update is presented later in this article (see Experiment 4) and was shown previously in a study on the combination of stereo and motion information (Domini, Skirko, & Caudek, 2001). The model that we propose is analogous to the Kalman filtering method (Kalman, 1960) used by Hildreth, Ando, Andersen, and Treue (1995) and proposed by

Fulvio Domini and Quoc C. Vuong, Department of Cognitive and Linguistic Sciences, Brown University; Corrado Caudek, Psychology Department, University of Trieste, Trieste, Italy.

This research was supported by National Science Foundation Grant BCS-78441. Quoc C. Vuong was supported by a National Sciences and Engineering Research Council of Canada postgraduate scholarship.

Correspondence concerning this article should be addressed to Fulvio Domini, Department of Cognitive and Linguistic Sciences, P.O. Box 1978, Brown University, Providence, Rhode Island 02912. E-mail: Fulvio_Domini@brown.edu

Ando (1991) to estimate the depth map of a 3-D surface (see also Heel, 1990; Hildreth, Grzywacz, Adelson, & Inada, 1990; Hung & Ho, 1999; Matthies, Kanade, & Szeliski, 1989; Treue, Andersen, Ando, & Hildreth, 1995). Their model is a development of the incremental-rigidity scheme proposed by Ullman (1984). The method of Kalman filtering is an efficient technique used to estimate a noisy variable by taking into account its dynamic changes. Our model, however, differs from the algorithm of Hildreth et al. (1995) because it is based on a heuristic analysis of first-order optic flow (e.g., Braunstein, 1994; Domini & Caudek, 1999; Todd & Perotti, 1999), whereas Hildreth's model is based on a procedure of rigidity maximization aiming at improving the accuracy of the depth estimates. The characteristics of our model make it more suitable to account for both veridical performance and perceptual biases in long-term temporal integration of human SFM.

The present experiments investigated the process of surface-orientation update occurring after the early stage of surface recovery, thus extending the previous empirical evidence of temporal integration in SFM. In the following, we first illustrate some relevant properties of the optic flow, after which we describe our model of surface-orientation update and, finally, discuss four experiments in which the proposed model has been tested.

Empirical Research on the Optic Flow

A mathematical analysis of the properties of the optic flow reveals that only a few assumptions about the 3-D motion of the projected objects are needed to derive their veridical 3-D shape (Bennett, Hoffman, Nicola, & Prakash, 1989; Hoffman, 1982; Hoffman & Bennett, 1985, 1986; Koenderink, 1986; Koenderink & Van Doorn, 1975, 1976, 1990; Longuet-Higgins & Prazdny, 1980; Prazdny, 1980; Ullman, 1979). If 3-D rigid motion is assumed, for example, then three orthographic projections of four moving points are sufficient to derive their 3-D euclidean structure (Ullman, 1979). Whereas two views determine the first-order temporal properties of the optic flow (velocities), three views characterize the second-order temporal properties (accelerations).

Several theoretical studies have shown that second-order properties of the optic flow are needed to reconstruct the veridical 3-D shape of projected objects (e.g., Hoffman, 1982; Ullman, 1979), and several empirical studies have investigated whether human observers actually use this information (Braunstein, 1976; Braunstein, Hoffman, & Pollick, 1990; Braunstein, Hoffman, Shapiro, Andersen, & Bennett, 1987; Braunstein, Liter, & Tittle, 1993; Liter, Braunstein, & Hoffman, 1993; Loomis & Eby, 1988; Ono, Rivest, & Ono, 1986; Rogers & Graham, 1979). The majority of these studies established that only the first-order properties of the optic flow seem to be used by the human visual system (Caudek & Domini, 1998; Domini & Caudek, 1999; Domini, Caudek, & Proffitt, 1997; Domini, Caudek, & Richman, 1998; Liter & Braunstein, 1998; Liter et al., 1993; Norman & Todd, 1993, 1995; Todd & Bressan, 1990; Todd & Norman, 1991). In particular, two main findings lead to this conclusion: (a) Human observers do not derive a veridical 3-D euclidean structure from the optic flow (Domini & Braunstein, 1998; Norman & Todd, 1992), and (b) performance does not improve when the number of views is increased from two to many (Todd & Bressan, 1990; Todd & Norman, 1991).

Because a two-view sequence can be produced by the orthographic projection of an infinite number of different 3-D rigid structures, more recently researchers have tried to understand which structure among these is actually chosen by the visual system (Domini & Caudek, 1999; Norman & Todd, 1993, 1995; Todd & Bressan, 1990; Todd & Norman, 1991). Some light has been cast on this problem in studies of the properties of 3-D structures that are perceptually recovered from velocity fields generated by simple 3-D shapes such as dihedral angles (Braunstein et al., 1993; Liter & Braunstein, 1998; Todd & Perotti, 1999) or planar surfaces (Domini & Caudek, 1999). These studies have determined the relationships between the few parameters that characterize these velocity fields and the 3-D properties of the perceived structures. Whereas the laws that describe these relationships may differ from one study to another in terms of their mathematical form, the qualitative predictions are equivalent and stand on the common assumption that, after the early stage of surface recovery previously described, perceptual derivation of 3-D structure from the first-order optic flow is not affected by any form of long-term temporal integration. We propose here a different hypothesis, but first we need to summarize how some current models account for the perceptual derivation of surface orientation from the velocity field.

Perception of a Moving Planar Patch

For the present purposes, it is sufficient to consider a velocity field produced by the orthographic projection of a planar patch rotating about an axis contained in the image plane. It can be shown that such a linear velocity field can be characterized by only a few parameters and that these parameters can be used to predict the perceived 3-D orientation and motion of the planar patch (for a detailed discussion, see Domini & Caudek, 1999; Liter & Braunstein, 1998; Todd & Perotti, 1999).

The 3-D orientation of a planar surface can be described in terms of two parameters: slant (σ) and tilt (τ). Slant is the angle between the normal to the surface and the line of sight. Tilt is the angle between the (orthographic) projection of the normal to the surface on the image plane and the x -axis. The 3-D motion of a planar surface can be described by specifying the angular velocity and orientation of the axis of rotation. Because we consider here only rotations about an axis parallel to the image plane, it is convenient to choose a coordinate system in which the y -axis coincides with the axis of rotation. In this case, the rotation of the patch is fully specified by the magnitude (ω) of the angular velocity (see Figure 1a).

The instantaneous velocity field produced by the orthographic projection of this particular 3-D motion is characterized by parallel velocity vectors (Figure 1b). The intensity v of the velocity vectors is given by

$$v = \phi_x x + \phi_y y, \quad (1)$$

where ϕ_x and ϕ_y are the velocity-gradient components in the orthogonal and parallel directions relative to the axis of rotation, respectively. Three main empirical findings describe the relation between a linear velocity field and the perceived 3-D orientation and 3-D motion of a rotating planar surface: (a) The perceived tilt (τ') of the surface is equal to $\arctan(\phi_y/\phi_x)$, (b) the perceived slant (σ') of the surface is an increasing function of the deformation

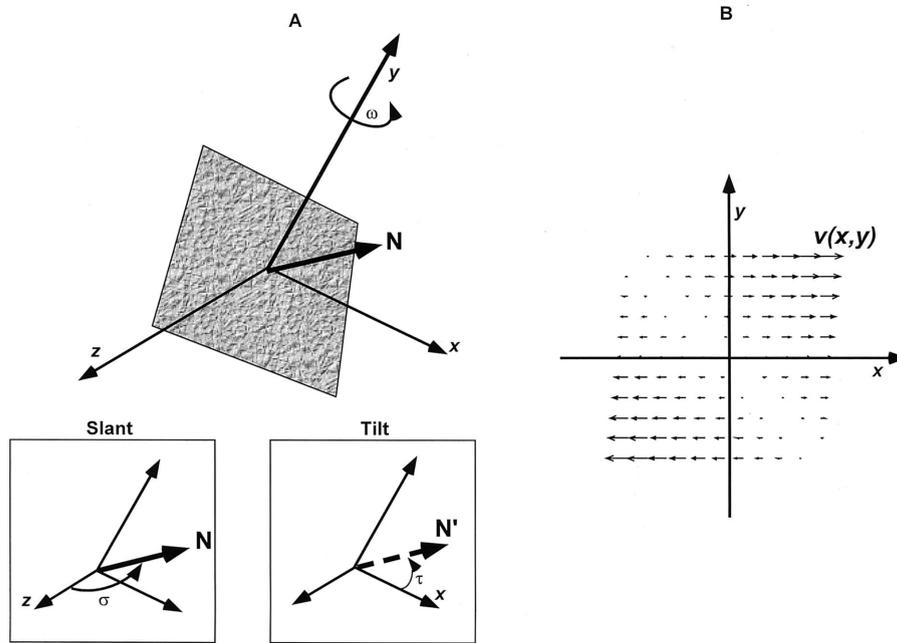


Figure 1. A: The 3-D orientation of a planar patch is defined by two parameters: slant, σ , and tilt, τ . B: Instantaneous velocity field produced by the orthographic projection of the 3-D angular motion, ω , about the y -axis. The field is characterized by parallel velocity vectors with two gradient components, one along the x -axis, ϕ_x , and the other along the y -axis, ϕ_y . N is the normal to the planar surface.

(def) component— $\sqrt{\phi_x^2 + \phi_y^2}$ —and a decreasing function of $\arctan(\phi_y/\phi_x)$, and (c) the perceived angular rotation (ω') is an increasing function of def (Domini & Caudek, 1999; Freeman, Harris, & Meese, 1995; LITER & Braunstein, 1998; Todd & Perotti, 1999).

In summary, it has been recently proposed that, in deriving the parameters (σ' , τ' , and ω') that describe the perceived 3-D orientation and 3-D motion of a planar patch, the visual system measures the two components ϕ_x and ϕ_y of the velocity gradient. Although it is known that the measurement of the velocity gradients occurs within an extended temporal window of about 200 ms, temporal integration has not yet been investigated beyond the early stage of surface recovery.

Long-Term Temporal Integration

Consider now the simplest case in which the gradient component ϕ_y is zero. An optic flow of this sort is produced by the orthographic projection of a rotating planar surface parallel to the axis of rotation (Figure 2a). A rigid flag rotating about its post is an example of this kind of 3-D motion. If the flag rotates away from the frontal-parallel plane, the resulting optic flow is characterized by a pure contraction (Figure 2b, bottom). If the flag approaches the frontal-parallel plane, the optic flow is characterized by a pure expansion (Figure 2b, top). Whereas the absolute value of the gradient $|\phi_x|$ represents the rate of compression or expansion, the sign of the gradient indicates whether the optic flow is expanding or contracting.

What it is important to note is that the SFM models described earlier, by predicting that the magnitudes of slant and angular rotation perceived at moment t_0 depend solely on the intensity of

the gradient ϕ_x , do not capture the full extent of observers' perceptions. This point can be illustrated by considering the case of a constant optic flow (see Perotti, Todd, & Norman, 1996). In the case of Figure 2b, for example, the models described earlier predict a constant surface slant, because the gradient ϕ_x at any moment is constant. Human observers, on the other hand, report that surface slant appears to be continuously increasing in the course of time, as a consequence of the perceived rotation (Domini, Caudek, Turner, & Favretto, 1998). Such a discrepancy between predicted and perceived slant, therefore, suggests that the models of human SFM described in the previous section are incomplete and must be supplemented by a component of long-term temporal integration serving the purpose of surface-orientation update.

To describe the proposed model of surface-orientation update, we now consider the case described in the previous section (constant optic flow, $\phi_y = 0$), and assume for simplicity that discrete measurements of the optic flow can be obtained in successive moments t_0, t_1, \dots, t_n . Here the time interval $\Delta t = t_i - t_{i-1}$ is considered to be the short-term temporal integration window needed for the visual system to measure the gradient ϕ_{xi} . In the time span t_0 to t_n , the available gradient measurements are $\phi_{x0}, \phi_{x1}, \dots, \phi_{xn}$. According to the SFM models described earlier, the slant σ'_i and the angular rotation $\Delta\alpha'_i$ derived at moment t_i depend only on the gradient ϕ_{xi} (Figure 3, left). For example, according to Domini and Caudek (1999):

$$\sigma'_i = f_\sigma(\phi_{xi})$$

$$\Delta\alpha'_i = \omega'_i \Delta t = f_\omega(\phi_{xi}) \Delta t. \quad (2)$$

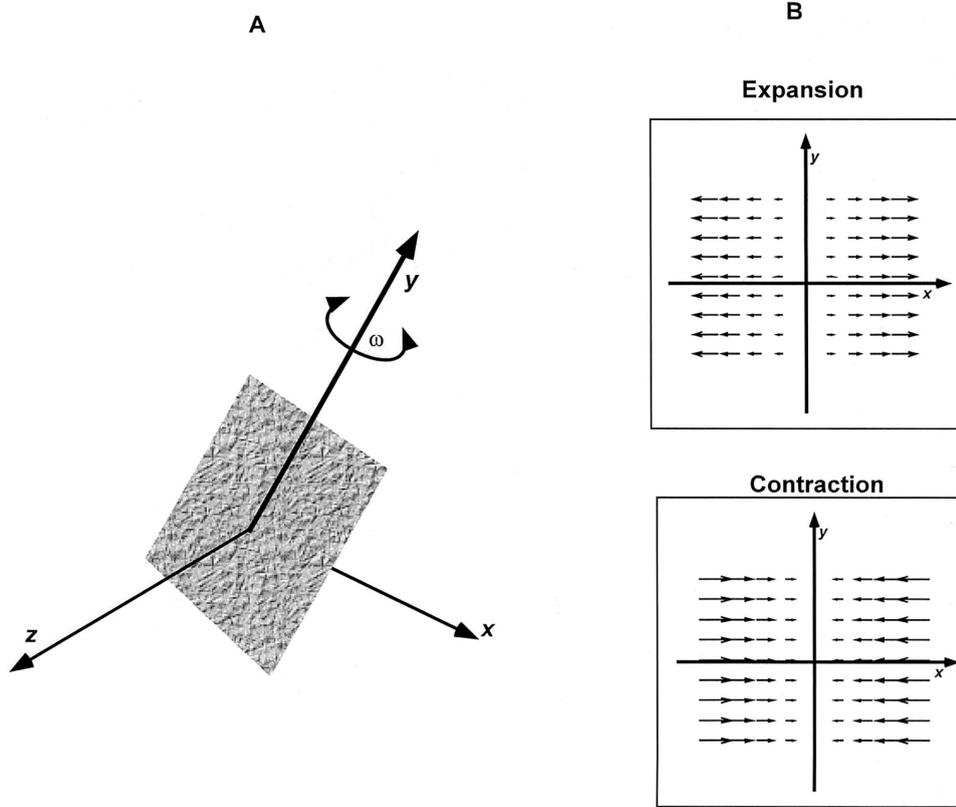


Figure 2. A: Rotating planar surface parallel to the axis of rotation (y), where ω is the 3-D angular velocity. B: Velocity field generated by a planar surface rotating about the y -axis. This velocity field is defined by only one gradient component in the horizontal direction, ϕ_x . The top shows an expanding optic flow (positive gradient). The bottom shows a contracting optic flow (negative gradient).

(The two functions f_σ and f_ω are specified in greater detail later, when we describe the implementation of the model in relation to the results of Experiment 1.) Given that the gradient ϕ_x is constant, the slant magnitude derived at any moment of the stimulus sequence will also be constant. Note, however, that over an extensive period of time, the predictions of such a model are internally inconsistent. Whereas the gradient ϕ_{xi} measured at time t_i implies the interpretation $\sigma'_i = f_\sigma(\phi_{xi})$, the previous measurement of the optic flow would require that, at time t_p , the slant be equal to $\sigma'_{i-1} + \Delta\alpha'_{i-1}$. The simplest way to solve this contradiction is to hypothesize that perceived slant is equal to the weighted average of the slant magnitudes derived at moments t_i and t_{i-1} :

$$\sigma'_i = w(\sigma'_0 + \omega'_0\Delta t) + (1 - w)f_\sigma(\phi_{xi}), \quad (3)$$

where w is any number in the range (0, 1). At the generic moment t_p , therefore, Equation 3 becomes

$$\sigma'_i = w[\sigma'_{i-1} + f_\omega(\phi_{xi-1})\Delta t] + (1 - w)f_\sigma(\phi_{xi}), \quad (4)$$

because $\omega'_{i-1} = f_\omega(\phi_{xi-1})$.¹

Equation 4 describes the surface-orientation-update model that we propose. In this equation, the weight w determines the strength of the integration. If $w = 0$, then the slant derived at time t_i depends solely on the gradient ϕ_{xi} . If $w = 1$, conversely, σ'_i is found by incrementing σ'_{i-1} by the full amount $\Delta\alpha'_{i-1}$.

It is important to note that the proposed integration model does not involve an analysis of second-order optic flow (acceleration); rather, it makes use of only first-order information (velocity). Even though the model is based only on first-order optic flow, derivation of surface orientation is not restricted to the information provided by two frames of an SFM apparent-motion sequence. Derivation of surface orientation at time t_p , in fact, adds to the (first-order) information provided by the gradient ϕ_{xi} the information relative to the 3-D orientation and rotation at time t_{i-1} (the magnitude of 3-D rotation being itself derived from the first-order optic flow).

¹ For simplicity, we have presented the model of Equation 4 in a discrete form. This model, however, is equivalent to a continuous model that assumes that, at an instant of time t , the derived slant is a function of (a) the derived slant at $t - \Delta t$, (b) the angular rotation derived from the gradient in the time interval Δt , and (c) the optic-flow gradient at time t . In symbolic form:

$$\sigma(t)' = w\{\sigma(t - \Delta t)' + \int_{t-\Delta t}^t f_\omega[\phi(u)]du\} + (1 - w)f_\sigma[\phi(t)].$$

If we assume that the optic flow gradient is constant during the interval Δt , the previous equation becomes

$$\sigma(t)' = w\{\sigma(t - \Delta t)' + f_\omega[\phi(t - \Delta t)]\Delta t\} + (1 - w)f_\sigma[\phi(t)].$$

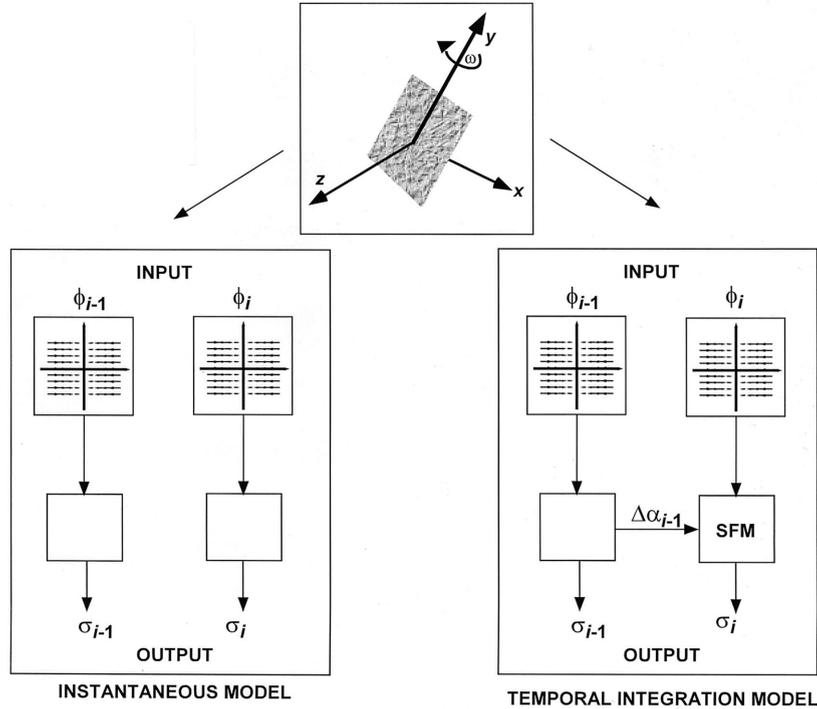


Figure 3. Top: Rotating planar surface parallel to the axis of rotation (y), where ω is the 3-D angular velocity. Bottom: Model described in Equation 4 when long-term temporal integration is assumed (right) and the temporal integration weight (w) is set to zero (left). The model represented on the left measures the optic flow generated by a surface rotating about the y -axis at each instant of time i . The slant of the surface is derived from the gradient ϕ_{xi} , and the information about ϕ_{xi} is gathered within a small temporal window of about 200 ms preceding moment t_i (short-term temporal integration). The model represented on the right, conversely, updates the existing representation of surface slant measured at time $i - 1$ (σ_{i-1}) by an amount equal to the angular rotation, $\Delta\alpha_{i-1}$, and combines this slant magnitude with that derived from the gradient ϕ_{xi} measured at time i . SFM = structure from motion.

The output of the model of Equation 4 depends on two parameters: the long-term integration weight w and the size Δt of the temporal window within which the optic flow is measured. If ϕ_x is constant, then Equation 4 becomes

$$\sigma_i = f_\sigma(\phi_x) + (w^0 + \dots + w^i) f_\omega(\phi_x) \Delta t. \quad (5)$$

Because the sum of w^i is a geometric series (from which 1 has been subtracted), we can rewrite Equation 5 as

$$\sigma_i = f_\sigma(\phi_x) + G_w(i) f_\omega(\phi_x) \Delta t, \quad (6)$$

where $G_w(i)$ is the outcome of the geometric series sum at the i th iteration, and it is equal to

$$G_w(i) = \frac{w^{i+1} - w}{w - 1}. \quad (7)$$

Given that $f_\sigma(\phi_x)$ and $f_\omega(\phi_x)$ are constant (because ϕ_x is constant), the time variation of the slant derived in Equation 6 is governed by the function $G_w(i)$. $G_w(i)$ is an increasing function of iteration step i and approaches a plateau, because w is a value between 0 and 1. With $n \rightarrow \infty$, the function $G_w(i)$ converges to the constant value $w/(1 - w)$. As a consequence, if ϕ_x is constant, then the iterative process of temporal integration is bounded. The slant

derived according to Equation 6, therefore, reaches a plateau after a certain number of iterations. This plateau represents an upper bound for (contracting) optic-flow sequences in which the derived slant is continuously increasing and a lower bound for (expanding) optic-flow sequences in which the derived slant is continuously decreasing. The speed with which the plateau is reached depends on w and Δt . $G_w(i)$ reaches a plateau as soon as w^{i+1} becomes negligible. The smaller the weight w , the faster the plateau is reached. The size of Δt is also important. It is obvious that if i_p iterations are necessary to reach a plateau, the time $t = i_p \Delta t$ will increase as Δt increases. The effects of w and Δt on the output of the long-term temporal integration model are illustrated in Figure 4.

In summary, with the present investigation, we intend to establish whether human SFM involves a process of long-term temporal integration serving the purpose of surface-orientation update. In Experiments 1 (surface rotation) and 2 (surface oscillation), we tested the surface-orientation-update model by using constant optic-flow fields. In Experiment 3, we tested the model by using flow fields that, in principle, provide second-order temporal information. Finally, in Experiment 4, we compared long-term temporal integration in SFM and speed-discrimination tasks.

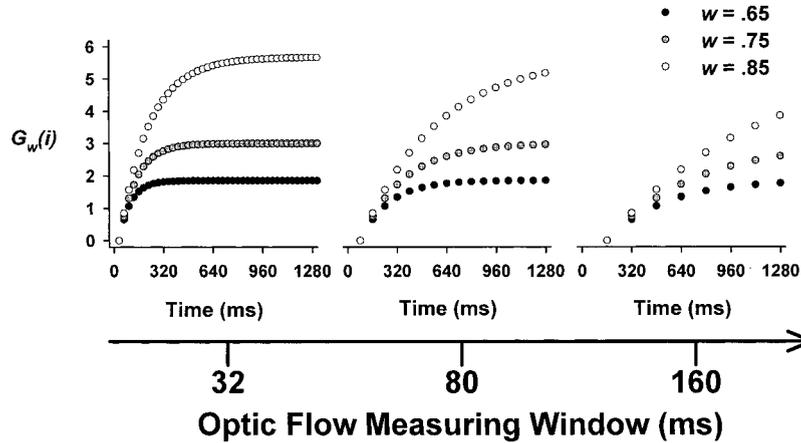


Figure 4. Output $[G_w(i)]$ of the surface-orientation-update model for different values of the parameters w and Δt . The values of w and Δt determine the time necessary for the convergence to either an upper bound (for contracting sequences) or a lower bound (for expanding sequences). The parameter w takes on the values of .65 (black circle), .75 (gray circle), and .85 (open circle). The parameter Δt takes on the values of 32, 80, and 160 ms in the left, central, and right graphs, respectively. Each point on the graphs represents the output of one iteration of the model.

Experiment 1

The purpose of Experiment 1 was to compare human performance in an opportunely devised SFM task with the outcomes of the model of Equation 4. Two versions of the model were implemented. In one version, the long-term temporal integration weight w was set to zero (no surface-orientation update), thus making the present model equivalent to the models proposed by Domini and Caudek (1999) and Todd and Perotti (1999). In the second version, the long-term temporal integration weight w took on a value larger than zero (surface-orientation update).

In each trial, participants were shown two constant optic-flow sequences presented side by side. The sequences had the same length and started at the same time. Each sequence was made up of two successive segments, with no interstimulus interval between them. The first segment was called *history*, and the following segment was called *test*. The optic-flow gradients used for the history and test segments are illustrated in Figure 5, in which the history segments are represented within shaded regions. In the same-gradient condition, the two history segments exhibited the same gradients ϕ_x (Figure 5, left). In the different-gradient condition, the two history segments exhibited different gradients (Figure 5, right). In the different-gradient condition, the largest gradient in the history segment was coupled with the smallest gradient in the test segment. The moment in which the transition between history and test occurs is indicated in Figure 5 with $t = 0$. Across trials, the lengths of the test segments for the two sequences took on different values. The task of the participants was to compare, at the end of stimulus sequences, the slants exhibited by the surfaces resulting from the two flow fields.

Simulations were run for the two versions of the model of Equation 4 (the simulations are described in detail below). For $w = 0$, performance of the model was accurate in both the same- and different-gradient conditions (i.e., the largest slant magnitude was attributed to the largest velocity gradient), regardless of the length of the test segments. For $w > 0$, in contrast, the model's perfor-

mance was very different in the two conditions. In the same-gradient condition, performance was accurate regardless of the length of the test segments; in the different-gradient condition, performance was accurate only for the long test sequences.

Method

Participants. Three volunteers, recruited from the Brown University community and naive as to the purpose of the experiment, and two of the authors (Fulvio Domini [F.D.] and Quoc C. Vuong [Q.V.]) participated in this experiment. All participants had normal or corrected-to-normal vision.

Stimuli. The stimuli were moving high-luminance random dots presented on a low-luminance background. The motion of the dots defined

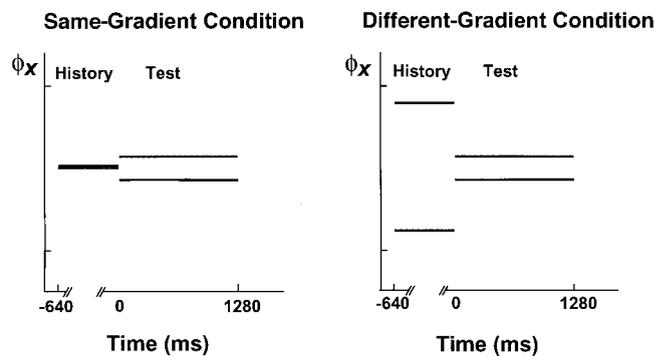


Figure 5. Time variation of the horizontal gradient (ϕ) of the two stimulus sequences used in Experiment 1 (black and gray lines). For both sequences, the history segment lasts 640 ms, and the test segment lasts 1,280 ms. In the same-gradient condition (left), both sequences exhibit the same gradient during the history segment. In the different-gradient condition (right), the gradient of one sequence during the history segment (black line) is greater than the other (gray line). In the test segment, the gradients of the two sequences do not vary across conditions. The sequence with the largest gradient in the history segment is coupled with the sequence with the smallest gradient in the test segment (black lines).

linear velocity fields (all velocity vectors were parallel to the vertical axis). In the test segment of the stimulus sequence, the instantaneous gradient components ϕ_{x1} and ϕ_{x2} of the two adjacent velocity fields took on the values of 0.1641 s^{-1} and 0.1484 s^{-1} (i.e., $\pm 0.0078 \text{ s}^{-1}$ with respect to an average gradient of 0.1563 s^{-1}), respectively, for both same-gradient and different-gradient conditions. In the case of one of the participants (D.R.), ϕ_{x1} and ϕ_{x2} were set to 0.1719 s^{-1} and 0.1406 s^{-1} . In the same-gradient condition, the history segment of the stimulus sequence had a gradient of 0.1563 s^{-1} (i.e., the average of the gradients of the two test segments). In the different-gradient condition, the gradients of the history segments of the stimulus sequence were 0.2188 s^{-1} and 0.0938 s^{-1} (or $\pm 0.0625 \text{ s}^{-1}$ with respect to an average gradient 0.1563 s^{-1}). In the different-gradient condition, the velocity field having the largest gradient during the history segment exhibited the smallest gradient during the test segment of the stimulus sequence (see Figure 5).

A nonlinear spatial component was added to the velocity gradients so that each velocity field gave rise to the perception of a squeezed cylinder rotating about its base. This nonlinearity was introduced to minimize the likelihood of depth reversals. No reversals were reported after its introduction (the displays appeared as convex surfaces; see Mamassian & Landy, 1998). The instantaneous optic flow can therefore be described by the following equation:

$$v(x, y, i) = \phi_x(i)x + \phi_y\sqrt{r^2 - y^2}, \quad (8)$$

where i is the frame number of the stimulus sequence.

The history segment of the stimulus sequence was made up of 40 frames, whereas the number of frames of the test segment was systematically manipulated in the 16–80 range, in 16-frame steps. Each frame was presented for 16 ms. The duration of the history segment was thus 640 ms, and the duration of the test segment ranged from 256 ms to 1,280 ms, in 256-ms steps. The two random-dot fields shown on each trial were contained in a region approximately 4 cm wide and 9 cm tall ($2.6^\circ \times 5.7^\circ$). They were separated by a blank region approximately 1 cm wide, with a distance of 5 cm between their centers. Each frame of the stimulus sequence displayed 500 dots, and dot density was kept constant (Sperling, Landy, Doshier, & Perkins, 1989).

Apparatus. The displays were presented on a high-resolution color monitor ($1,280 \times 1,024$ 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 approximately 76 cm from the monitor. The circular aperture of the reduction screen had a ray of 2 cm and limited the visible portion of the monitor to a region with a diameter of approximately 12.9 cm (8.1° of visual angle). A chin rest was used to restrict head movement. The experiment was conducted in a dark room.

Design. Three within-subjects variables were studied: type of history segment (same gradient or different gradient), duration of the test segment (256 ms, 512 ms, 768 ms, 1,024 ms, or 1,280 ms), and relative position of the velocity field with the largest gradient (left or right). The same-gradient and different-gradient conditions were blocked. In each block, participants viewed 20 trials for each of the 10 conditions, with the order of trials completely randomized. The sequence of the blocks was ABBA for 3 participants (1 expert and 2 naive) and BAAB for the remaining 2 participants.

Procedure. Participants were asked to report whether the velocity field supporting the largest perceived slant at the end of the stimulus sequence was located on the left or on the right. They responded by pressing the corresponding mouse buttons connected to the workstation. No feedback was provided for correct responses. Participants took part individually in two sessions. The first session served to determine the gradient difference ($\phi_{x1} - \phi_{x2}$) needed to reliably associate (in at least the 80% of the cases) the larger perceived slant with the velocity field having the largest gradient.

In the second session, each participant completed four blocks of 200 trials each. Participants took a break after each block.

Simulations

One simulation was run by using as input to the model the time series of gradient magnitudes measured every 160 ms (10 frames) of the stimulus sequence. The history segment was 640 ms (40 frames), whereas the test segment ranged from 256 ms to 1,280 ms (16 to 80 frames). To implement the model according to Equation 4, we must estimate the values of σ' and ω' . Because we assume that human SFM is based on a heuristic analysis of the first-order optic flow, we computed σ' and ω' according to the following equations (a discussion of the rationale underlying this choice was provided by Domini & Caudek, 1999):

$$\sigma' = \frac{1}{k_\omega} \sqrt{\phi_x}$$

$$\Delta\alpha' = \omega' \Delta t = k_\omega \sqrt{\phi_x} \Delta t.$$

The weight k_ω was estimated by Domini and Caudek (1999) with a least square procedure, and it took on the value of 0.638. This same value was used in the present simulation.

The simulation produced the time series of predicted slant magnitudes shown in Figure 6. In this simulation, the parameter w was fixed so as to represent the absence ($w = 0$) or the presence ($w = 1$) of long-term temporal integration. The parameter Δt was also fixed and took on the value of 160 ms. The output of this simulation is shown in Figure 6, where the gray lines represent the slant magnitudes recovered from the velocity field with the largest gradient in the test segment and the black lines represent the slant magnitudes recovered from the velocity field with the smallest gradient in the test segment. When $w = 0$, larger magnitudes of slant were assigned in the test segment to the velocity field with the largest gradient (compare Figures 5 and 6) in both the same-gradient and different-gradient conditions. This means that the outcome of the model was veridical. When $w = .6$, larger magni-

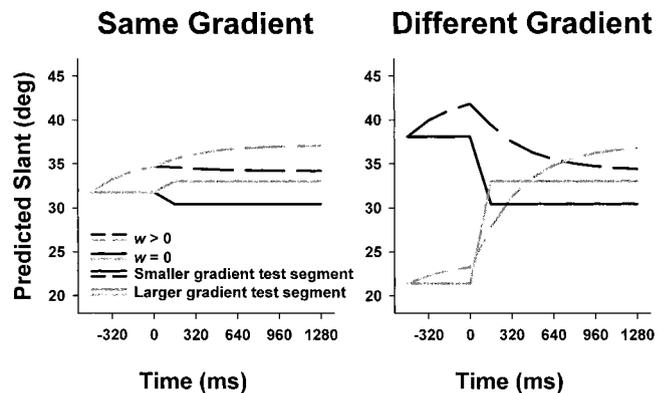


Figure 6. Time variation of slant magnitudes predicted by Equation 4 with $w = 0$ (solid lines) and $w = .6$ (dashed lines) for the same-gradient (left) and different-gradient (right) conditions. One data point is calculated for each 160-ms step of the stimulus sequence. The black lines (solid and dashed) represent the sequence with the smallest gradient during the test segment; the gray lines (solid and dashed) represent the sequence with the largest gradient during the test segment. deg = degrees.

tudes of slant were assigned to the velocity field with the largest gradient in the same-gradient condition, regardless of the length of the test segments. In the different-gradient condition, conversely, larger slant magnitudes are associated with the largest velocity gradient only after about 640 ms (approximately 40 frames) of the test segment. This means that, for these particular displays, long-term temporal integration biases the output of the model in such a manner that a larger slant magnitude is attributed to the velocity field with the smallest gradient in all test segments lasting less than 640 ms.

Figure 7 presents the results of the simulations in a different format. The figure shows the predicted slant difference, that is, the slant magnitude derived from the velocity field with the largest gradient minus the slant magnitude derived from the velocity field with the smallest gradient. When $w = 0$, the predicted slant difference had a positive value (meaning correct performance) during the entire test segment in both the same-gradient and different-gradient conditions. When $w = .6$, on the other hand, the predicted slant difference took on a positive value in the different-gradient condition only after about 640 ms of the test segment. As discussed earlier, this delay represents the time needed for the surface-orientation-update model to reach a plateau and increases with the sizes of both w and Δt .

In another set of simulations, the parameter w took on the values of 0, .5, and .85, and the parameter Δt took on the values of 32, 80, and 160 ms. The predicted slant differences for these simulations were transformed into predicted percentages of correct responses (i.e., judgments assigning the largest slant magnitude to the velocity field having the largest gradient). The uncertainty relative to the choice of the more slanted surface was modeled as Gaussian noise. The predicted percentage of correct responses was then determined by means of the term $CDF_0(s, \Delta\sigma')$, where CDF_0 is the normal cumulative-distribution function, with a mean equal to 0 and standard deviation s , and $\Delta\sigma'$ is the predicted slant difference. The

parameter s was estimated from pilot data and took on the value of 3.3°. The results of these simulations are shown in Figure 8.

Results and Discussion

Figure 9 reports the percentages of correct responses (i.e., the percentages of responses in which the largest perceived slant was attributed to the test segment having the largest gradient) as a function of the duration of the test segment. A mixed-design analysis of variance (ANOVA) was conducted on the frequencies of correct responses; condition (same gradient vs. different gradient), position (left vs. right), and duration of the test segment (256 ms, 512 ms, 768 ms, 1,024 ms, or 1,280 ms) were the within-subjects independent variables, and expertise (expert vs. naive) was the between-subjects independent variable. This analysis revealed a significant interaction between condition and duration, $F(4, 12) = 5.38, p < .01$, but no significant effect of expertise or significant interaction between expertise and other variables. The significant interaction between condition (same gradient vs. different gradient) and duration of the test sequence was consistent with the outcome of the simulation with $w > 0$ but not with the outcome of the simulation with $w = 0$. When $w = 0$, in fact, the model produced a veridical outcome in both the same-gradient and different-gradient conditions for all test segments. Moreover, no improvement in performance occurred as the duration of the test segment increased (see Figure 8, bottom row). When $w > 0$, on the other hand, better performance was observed in the same-gradient condition. Moreover, performance improved with the duration of the test segment.

For the short test segments of the different-gradient condition, finally, the model with $w > 0$ and a large temporal integration weight predicted that participants should attribute the largest perceived slant to the velocity field with the smallest velocity gradient (Figure 8, top row). It can be seen from Figure 9 that the predictions of the surface-orientation-update model ($w > 0$) were consistent with the qualitative trends of the psychophysical data. In the same-gradient condition, in fact, the percentages of correct responses were greater than 50% with the exception of participant M.H. in the shortest test segment. In the different-gradient condition, conversely, performance was much worse, and percentages of correct responses significantly above chance can be found only for 1 participant (Q.V.) and for the longest test segment. For shorter test sequences, the percentages of correct responses were significantly below chance, thus revealing a bias to report the test segment with the largest velocity gradient as having the smallest slant.

The predicted percentages of correct responses for the model with $w > 0$ were computed by fitting Equation 4 to the psychophysical data with three free parameters: the CDF_0 standard deviation s , the temporal integration weight w , and the time window Δt . The predicted percentages of correct responses for the model with $w = 0$ were computed by fitting Equation 4 to the data with only s and Δt as free parameters (the temporal integration weight was obviously fixed to zero). The unknown parameters were estimated by minimizing the root mean square differences (RMSDs) between the observed and the predicted percentages. Averaged across participants, the estimated values of the parameters that minimized the RMSDs were $s = 3.4, w = .82$, and $\Delta t = 0.099$ s for the surface-orientation-update model and $s = 3.4$ and $\Delta t = 0.096$ for

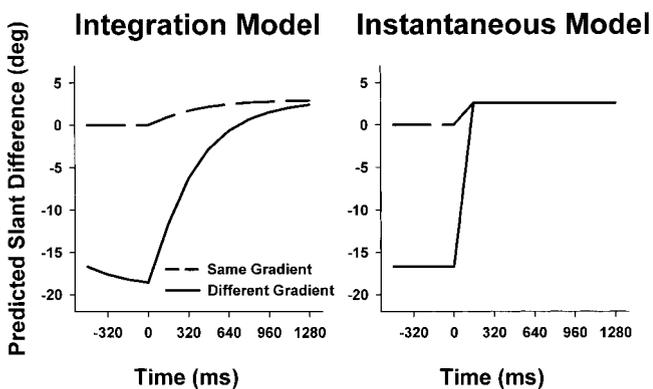


Figure 7. Predicted difference between the slant magnitudes of the two optic-flow sequences when long-term temporal integration is assumed (left) and when the temporal-integration weight (w) is set to zero (right). The predicted slant magnitude of the sequence with the smallest gradient during the test segment (gray lines in Figure 5) was subtracted from the slant magnitude of the sequence with the largest gradient (black lines in Figure 5). The slant differences for the same-gradient condition are represented by the dashed lines; the slant differences for the different-gradient condition are represented by the solid lines. deg = degrees.

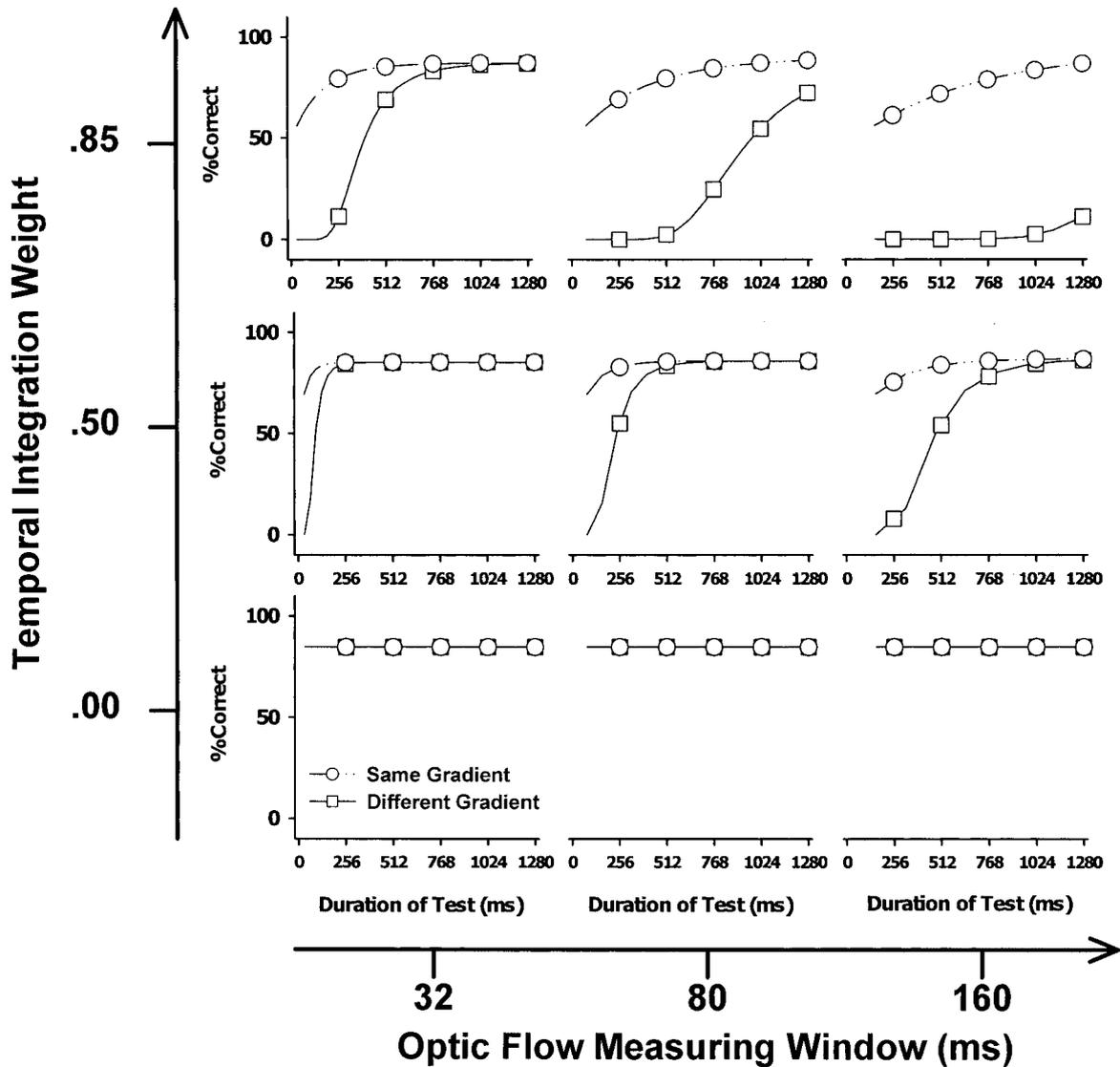


Figure 8. Predicted percentages of correct responses for the same-gradient (solid lines) and different-gradient (dashed lines) conditions, for different values of the parameters w (0, .50, and .85) and Δt (32, 80, and 160 ms).

the model assuming no long-term temporal integration. The averaged RMSDs for the two models were 7.98% and 34.29%, respectively. The estimated parameters and the RMSDs for the individual participants are shown in Table 1. The predictions of the two models are represented in Figure 9 with gray ($w > 0$) and black ($w = 0$) lines.

In conclusion, the results of the present experiment indicate that long-term temporal integration serving the purpose of surface-orientation update plays an important role in human SFM. In this experiment, in fact, we showed that identical velocity fields can give rise to the perception of different slant magnitudes, and we explained these differences through a model taking into account the time development of the velocity fields. These results cannot be accounted for by models lacking a long-term temporal integration component, as, for example, those proposed by Domini and Caudek (1999) and Todd and Perotti (1999).

Experiment 2

The purpose of Experiment 2 was to examine the perceptual interpretation of SFM sequences representing oscillating and rotating surfaces, because, for these stimuli, the two versions of the proposed model ($w = 0$ and $w > 0$) make very different predictions. A second purpose of Experiment 2 was to estimate the sizes of the parameters w and Δt that best predict human performance.

Method

Participants. Nine naive participants took part in this experiment. Two of them had participated in Experiment 1.

Stimuli. The stimuli were similar to those used in Experiment 1, except that some of the stimulus sequences were presented in a cyclic fashion. Figure 10 shows the time variation of the velocity gradients of the eight

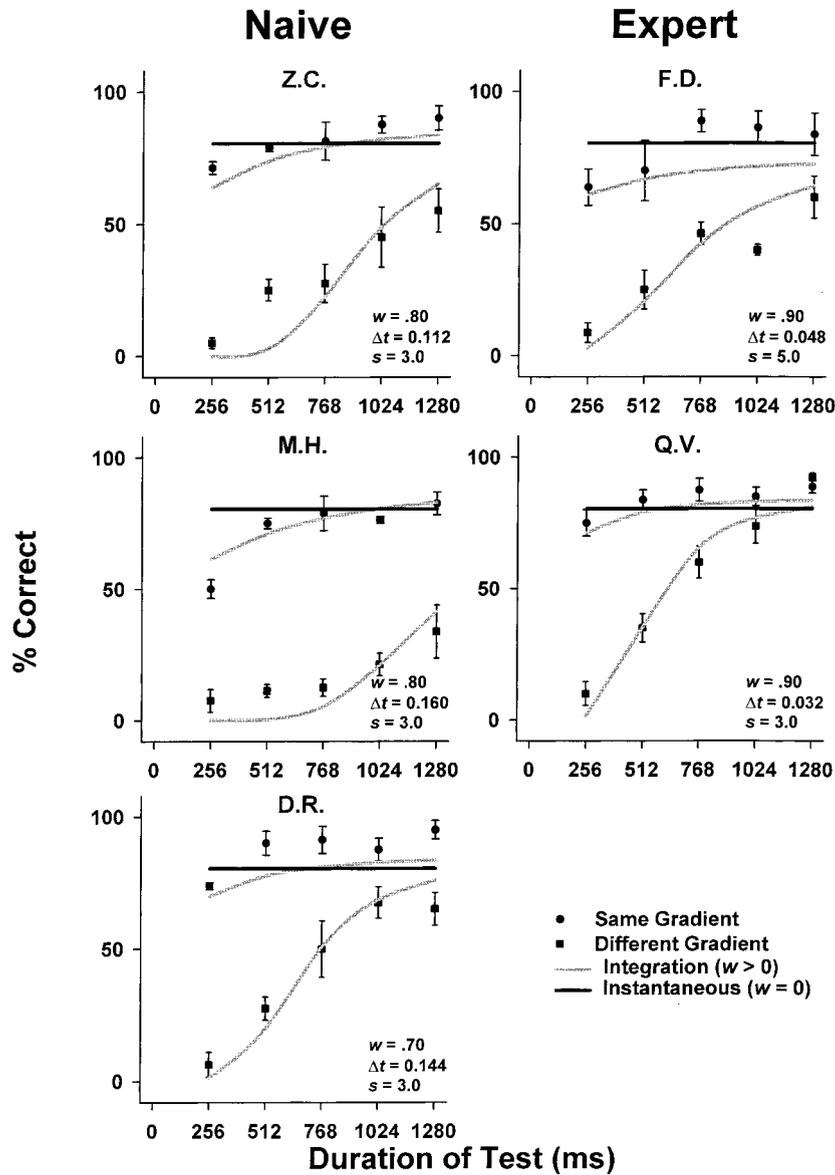


Figure 9. Mean percentages of correct responses for each participant (naive: Z.C., M.H., and D.R.; expert: F.D. and Q.V.) in Experiment 1, as a function of condition (same gradient [circles] or different gradient [squares]) and length of the test segments. The predictions of the model of Equation 4 are represented by the gray ($w > 0$) and black ($w = 0$) lines. The upper and lower gray lines represent the fits to the same- and different-gradient conditions, respectively. Also shown are the estimates of the three free parameters (w , Δt , and s) that minimize the root mean square difference between the model's predictions and the data. Error bars represent standard errors of the mean for each condition.

SFM sequences used in Experiment 2. The sequences represented in the top and bottom rows are identical except for the phase of the time variation of the gradients. Note that the sequences represented in the top row end with contracting flows, whereas those in the bottom row end with expanding flows. We therefore label the former sequences *end-contracting* sequences and the latter *end-expanding* sequences. Two optic-flow sequences that do not vary in a cyclic fashion, a contracting flow (top) and an expanding flow (bottom), are represented in the left portion of Figure 10. Optic-flow sequences with one oscillation cycle are represented immediately to the right: A flow that first expands and then contracts is represented on the top, and a flow that first contracts and then expands is represented on the

bottom. The other graphs represent sequences depicting two and four oscillation cycles.

In each trial, two optic-flow sequences were displayed side by side. The velocity gradients used for the contracting and expanding flows were -0.2500 s^{-1} and 0.2500 s^{-1} , respectively. A diagram of the time variations of the velocity gradients in the different experimental conditions is presented in Figure 10. The SFM sequences represented either constant or oscillatory optic flows. The oscillation frequency was 1, 2, and 4 cycles. All stimulus sequences had the same duration: 1,280 ms (80 frames).

Four end-contracting sequences and four end-expanding sequences were created, and thus six pairwise comparisons were possible within each

Table 1
Estimated Parameters for Each Participant in Experiment 1

Participant	s	w	Δt	Minimum RMSD
$w > 0$				
Q.V.	3.0	.90	0.0320	5.83
D.R.	3.0	.70	0.1440	7.96
M.H.	3.0	.80	0.1600	6.74
Z.C.	3.0	.80	0.1120	9.03
F.D.	5.0	.90	0.0480	10.35
M	3.4	.82	0.0992	7.98
$w = 0$				
Q.V.	3.0	0	0.096	27.95
D.R.	3.0	0	0.096	31.86
M.H.	3.0	0	0.096	46.23
Z.C.	3.0	0	0.096	37.00
F.D.	5.0	0	0.096	28.45
M	3.4	0	0.096	34.30

Note. RMSD = root mean square difference.

group. With c representing constant flow and 1, 2, and 4 representing the numbers of cycles, the possible comparisons were as follows: (c , 1), (c , 2), (c , 4), (1, 2), (1, 4), and (2, 4).

Apparatus. The apparatus was the same as that of Experiment 1.

Design. Three within-subjects variables were examined in Experiment 2: sign of the gradient ϕ_x at the end of the stimulus sequence, position of the cylinder with the highest oscillation frequency (left or right), and type of comparison or oscillation pair (six pairs). Participants completed trials individually in two blocks. In each block, they were shown 10 repetitions of each of the 24 experimental conditions, with the order of the trials completely randomized. Practice was provided before the experimental sessions to familiarize the participants with the stimulus displays.

Procedure. The procedure and task were the same as in Experiment 1.

Simulations

The predictions of the model assuming no long-term temporal integration ($w = 0$) are straightforward. It is obvious from Figure 10 that, at the end of the stimulus sequence, the model with $w = 0$ derives the same magnitude of the slant for all of the end-contracting and end-expanding sequences. The gradients at the end of the contracting and expanding test sequences, in fact, are identical in magnitude and differ only in their sign. If $w = 0$, therefore, the same slant magnitude is predicted for both constant and oscillatory sequences.

To explain the predictions of the model when $w > 0$, we note that the oscillatory sequences of the present experiment were made up of constant (contracting or expanding) optic-flow segments. As indicated in the previous discussion, for constant (contracting or expanding) optic-flow sequences, the model of Equation 4 derives a magnitude of slant that varies in the course of the sequence but eventually reaches a plateau if the stimulus sequence is long enough (see Figure 6). In the case of the present stimuli, the length

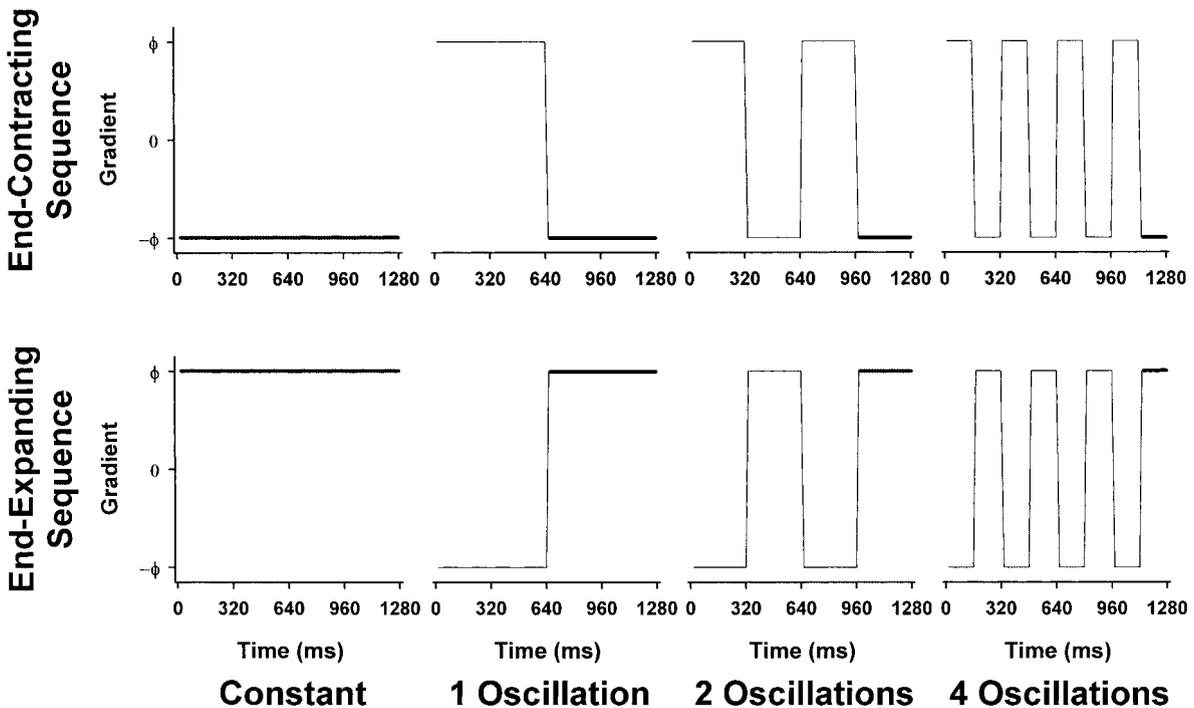


Figure 10. Time variations for the velocity gradients of Experiment 2. The end-contracting and end-expanding conditions are represented in the top and bottom rows, respectively. Thick black lines are used to emphasize the end part of each sequence. In both conditions, the test segment was made up of a constant optic flow or by a flow with 1, 2, or 4 oscillations.

of the constant optic-flow segments depends on the frequency of the oscillation cycle: The higher the frequency, the shorter the segments. If the frequency is too high, the constant optic-flow segments are too short, and the output of the long-term temporal integration model is prevented from reaching a plateau. It follows that manipulation of the oscillation frequency affects the slant magnitudes predicted by the long-term temporal integration model ($w > 0$).

To illustrate this point in greater detail, we now consider the stimulus sequences represented in the top of Figure 11: a constant contracting sequence (solid line) and a one-cycle end-contracting oscillatory sequence (dashed line). If $w = .65$ and $\Delta t = 32$ ms, then a short time is needed for the output of the long-term temporal integration model to reach a plateau. For the constant optic-flow sequence, in fact, a plateau is reached after only 160 ms (see Figure 11, bottom left). For the oscillatory sequence, the surface-orientation-update model derives a progressively decreasing slant magnitude during the initial expansion phase. When the contracting phase begins, therefore, the buildup of predicted slant begins from a smaller value than for the constant flow. Nevertheless, when both w and Δt are small, this relative disadvantage does not affect differently the slant magnitudes predicted at the end of either the constant or the oscillatory sequences (i.e., the solid and dashed lines converge).

Something different occurs when w and Δt are relatively large. If $w = .85$ and $\Delta t = 160$ ms, it takes longer for the output of the model to reach a plateau. As indicated in the bottom right of Figure 11 (for the constant contracting sequence), a plateau is not reached

before the end of the stimulus display (1,280 ms) and, as a consequence, the magnitudes of slant predicted at the end of the stimulus sequence differ for the constant and the oscillatory sequences (i.e., the solid and dashed lines do not converge).

The predictions of the surface-orientation-update model as a function of the frequency of the oscillation cycle and the sizes of the parameters w (.65, .75, and .85) and Δt (32, 80, and 160 ms) are shown in Figure 12 (end-contracting sequences) and Figure 13 (end-expanding sequences). Note that the difference in the slant magnitudes predicted for constant and oscillatory sequences increases with the size of w and Δt , for both end-contracting and end-expanding sequences. By means of the same procedure described in the previous experiment, the slant magnitudes predicted at the end of each stimulus sequence were transformed to percentages of responses in which the stimulus with the lowest frequency was selected as having the largest slant magnitude. The results of this simulation are shown in Figure 14.

Results and Discussion

The judgments of the participants were codified in terms of percentages of trials in which the SFM display with the lowest oscillation frequency was selected as having the largest slant at the end of the stimulus sequence. In the absence of any bias, the percentages of the lowest-frequency choices will not differ from 50%. The final parts of all end-contracting and end-expanding sequences, in fact, exhibit the same velocity gradient. The percentages of choices for the lowest-frequency stimulus, computed for each participant, all comparison pairs, and both end-contracting and end-expanding sequences, are shown in Figure 15, averaged across participants. It is immediately obvious from the figure that, at least for the pairs (c, 1), (c, 2), and (c, 4), the participants' judgments were strongly biased toward selecting the lowest-frequency sequence as more slanted for the end-contracting stimuli and toward selecting the lowest-frequency sequence as less slanted for the end-expanding stimuli.

Now we compare the psychophysical data with the predictions of the model. As in the previous experiment, the predictions of the surface-orientation-update model were computed by fitting Equation 4 to the psychophysical data with three free parameters: s , w , and Δt . The predictions of the model assuming no long-term temporal integration (i.e., $w = 0$) were computed by fitting Equation 4 to the data with only s and Δt as free parameters. The estimated values of the parameters that minimized the RMSDs were $s = 13.0$, $w = .9$, and $\Delta t = 0.176$ s and, when w was forced to zero, $s = 13.0$ and $\Delta t = 0.096$ s. The RMSDs were 9.15% and 31.23%, respectively.

The important point is that the response bias exhibited by the participants is predicted by the surface-orientation-update model, but it cannot be accounted for by the model in which the parameter w is set to zero (no long-term temporal integration). If we compare the psychophysical data of Figure 15 with the top-right corner of Figure 14 (where the simulation results are reported), we can conclude that the best fit for the participants' judgments is provided by the surface-orientation-update model with weights $w = .85$ and $\Delta t = 160$ ms.

One possible criticism that could be directed toward the present experiment is that the time window Δt necessary to measure the optic-flow gradients could be larger than half of the oscillation

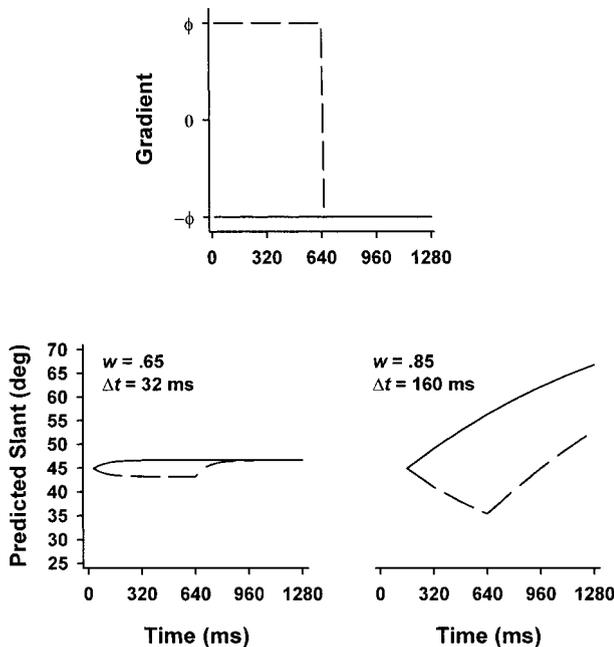


Figure 11. Top: Two sequences compared in Experiment 2: A constant contracting sequence (solid line) and a one-oscillation end-contracting sequence (dashed line). Bottom left: When w and Δt are small, the derived slant for both sequences is the same. Bottom right: When w and Δt are large, the derived slant for the constant optic flow is larger than the derived slant for the one-oscillation end-contracting sequence. deg = degrees.

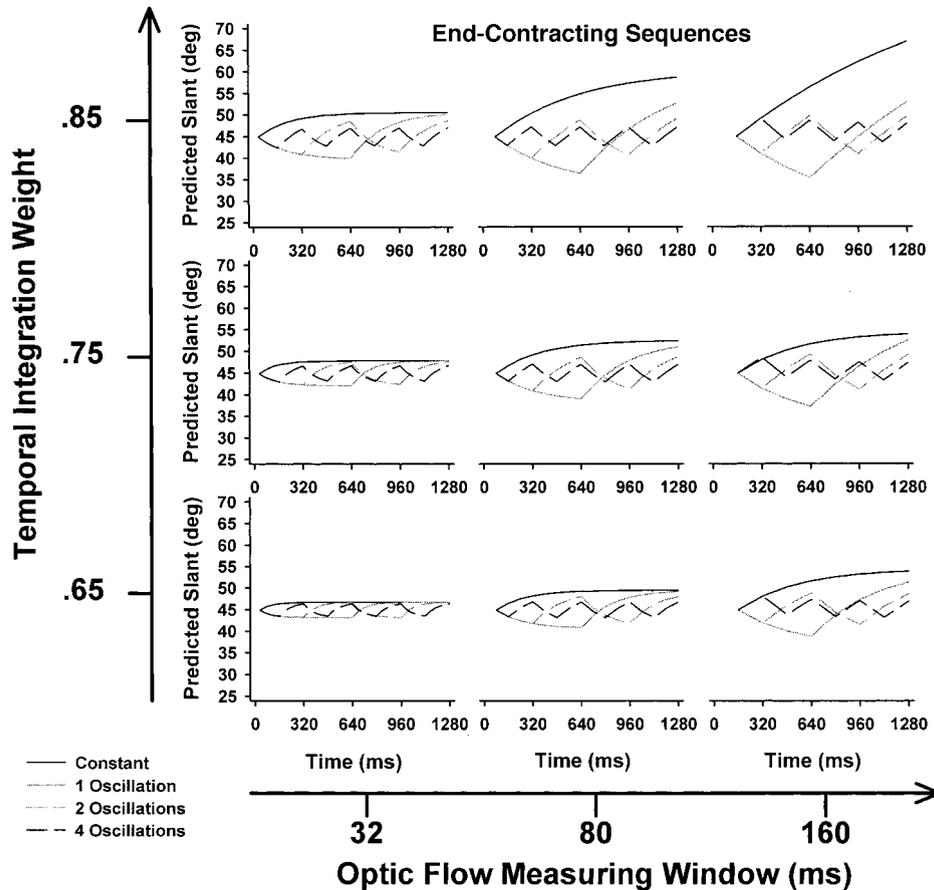


Figure 12. Time variations of the derived slant magnitudes for end-contracting sequences. Nine simulations were run with different values of the parameters w (.65, .75, and .85) and Δt (32, 80, and 160 ms). The curves within each graph refer to the number of oscillations of the stimulus sequences (0 [constant optic flow], 1, 2, or 4). deg = degrees.

period of the stimuli with the highest oscillation frequency (i.e., more than 160 ms). If this were the case, the gradient providing the input to the system would be computed by averaging over portions of constant optic-flow segments with velocity gradients having opposite signs, and it would be smaller (in absolute value) than the gradient ϕ at the end of the stimulus sequence (see Figure 10). As a consequence, a smaller value of slant would be attributed to the stimuli having the largest oscillation frequency, for both expanding and contracting sequences. This line of reasoning, however, can be rejected in that the opposite results were obtained for end-contracting and end-expanding stimuli: The sequence with the lowest frequency was selected as more slanted for the end-contracting stimuli and as less slanted for the end-expanding stimuli.

Experiment 3

In the introduction, we advanced the hypothesis that a constant flow field provides conflicting information. A contracting optic flow, for example, specifies an increasing slant (because the contraction of the flow indicates a rotation away from the frontal-parallel plane) and, at the same time, a constant slant (because the

gradient is constant). The surface-orientation-update model overcomes this contradiction by trading off the rotation specified by the contraction with the slant specified by the gradient (see Equation 4).

The purpose of the present experiment was to test the surface-orientation-update model in situations that maximize the potential conflict arising from the contraction–expansion of the flow (which specifies a rotation away or toward the image plane, i.e., an increasing or decreasing slant) and the properties of the gradient (which specify an increasing slant when the gradient increases and a decreasing slant when the gradient decreases). For this purpose, we simulated a contracting optic flow whose gradient decreases over time and an expanding optic flow whose gradient increases over time. In the first case, the contraction of the flow specifies an increasing slant, whereas the decreasing gradient specifies a decreasing slant. In the second case, the expansion of the flow specifies a decreasing slant, whereas the increasing gradient specifies an increasing slant. For completeness, we also tested two nonconflicting cases: a contracting optic flow with an increasing gradient and an expanding flow with a decreasing gradient. Note that, contrary to the previous experiments, the stimulus displays of

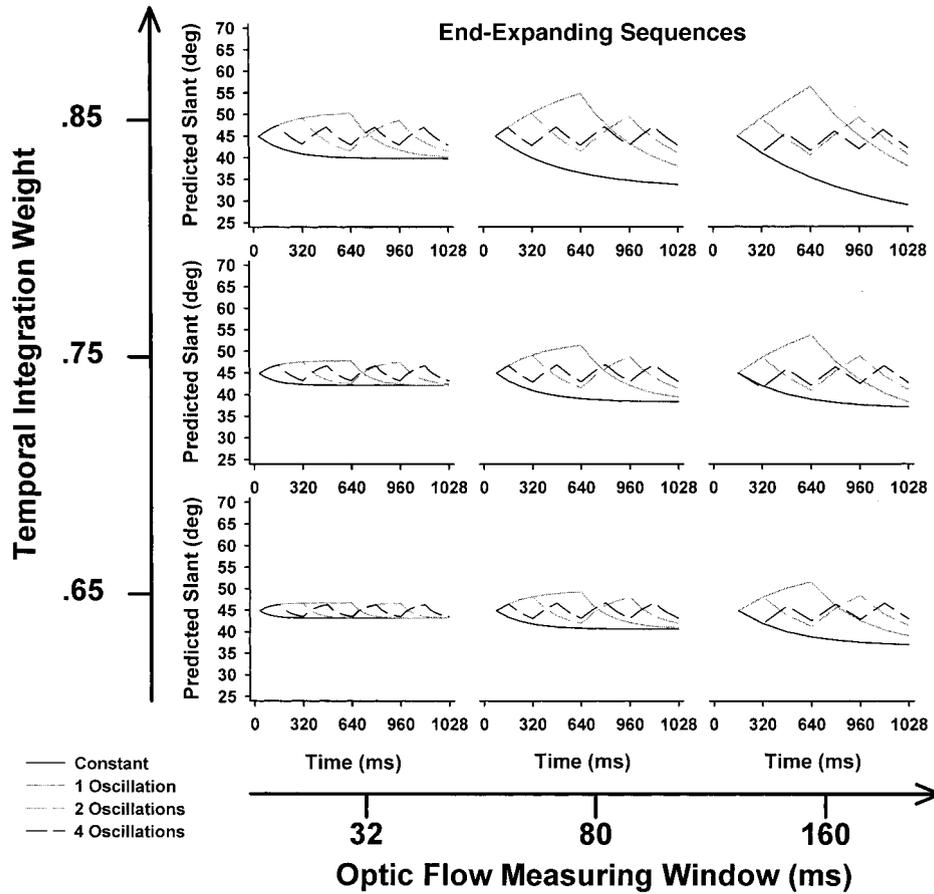


Figure 13. Time variations of the derived slant magnitudes for end-expanding sequences. Nine simulations were run with different values of the parameters w (.65, .75, and .85) and Δt (32, 80, and 160 ms). The curves within each graph refer to the number of oscillations of the stimulus sequences (0 [constant optic flow], 1, 2, or 4). deg = degrees.

Experiment 3 provided second-order temporal information, because the velocity gradients varied over time.

Method

Participants. Seven naive participants and one of the authors (Q.V.) volunteered their time. Four of the naive participants had participated in at least one of the previous experiments.

Stimuli. Two stimulus sequences (comparison and test) were presented side by side. Unlike in the previous experiments, however, the two sequences had different lengths: 2,880 ms (180 frames) for the comparison sequence and 160 ms (10 frames) for the test sequence. Although they had different lengths, the two sequences ended at the same time.

Four conditions were examined: contracting flow–decreasing gradient, contracting flow–increasing gradient, expanding flow–decreasing gradient, and expanding flow–increasing gradient. To illustrate these conditions, we now examine in detail the case of the contracting flow–increasing gradient and the contracting flow–decreasing gradient conditions. The absolute value of the velocity gradient produced by the projection of a planar surface rotating away from the frontal–parallel plane is given by

$$|\phi_x| = \sigma\omega, \tag{9}$$

where σ is the instantaneous value of the surface slant and ω is the 3-D angular velocity of the surface (Domini & Caudek, 1999). Given that the

optic flow undergoes a contraction, the velocity gradient has a negative sign. If the surface rotates with a constant angular velocity, then the velocity gradient increases during the rotation of the surface, because σ increases. This situation, therefore, produces a contracting flow with an increasing gradient. On the other hand, if the surface decelerates during the rotation, with ω decreasing at a faster rate than σ , the velocity gradient decreases as well. This produces a contracting flow with a decreasing gradient. It is easy to see how to extend such descriptions to the case of an increasing or decreasing expanding optic flow.

For the comparison sequences, in the increasing (decreasing) condition, the instantaneous gradient component ϕ_x varied from the value ϕ_{\min} (ϕ_{\max}) at the beginning of the sequence to the value ϕ_{\max} (ϕ_{\min}) at the end of the sequence. In the increasing condition, the instantaneous gradient $\phi_x(i)$ was governed by the following time law:

$$|\phi_x(i)| = \phi_{\min} + \frac{\phi_{\max} - \phi_{\min}}{n} i,$$

where i is the frame number (ranging from 0 to 180). In the decreasing condition, the time law was

$$|\phi_x(i)| = \phi_{\max} - \frac{\phi_{\max} - \phi_{\min}}{n} i;$$

ϕ_{\min} was 0.0625 s^{-1} , and ϕ_{\max} was 0.2500 s^{-1} .

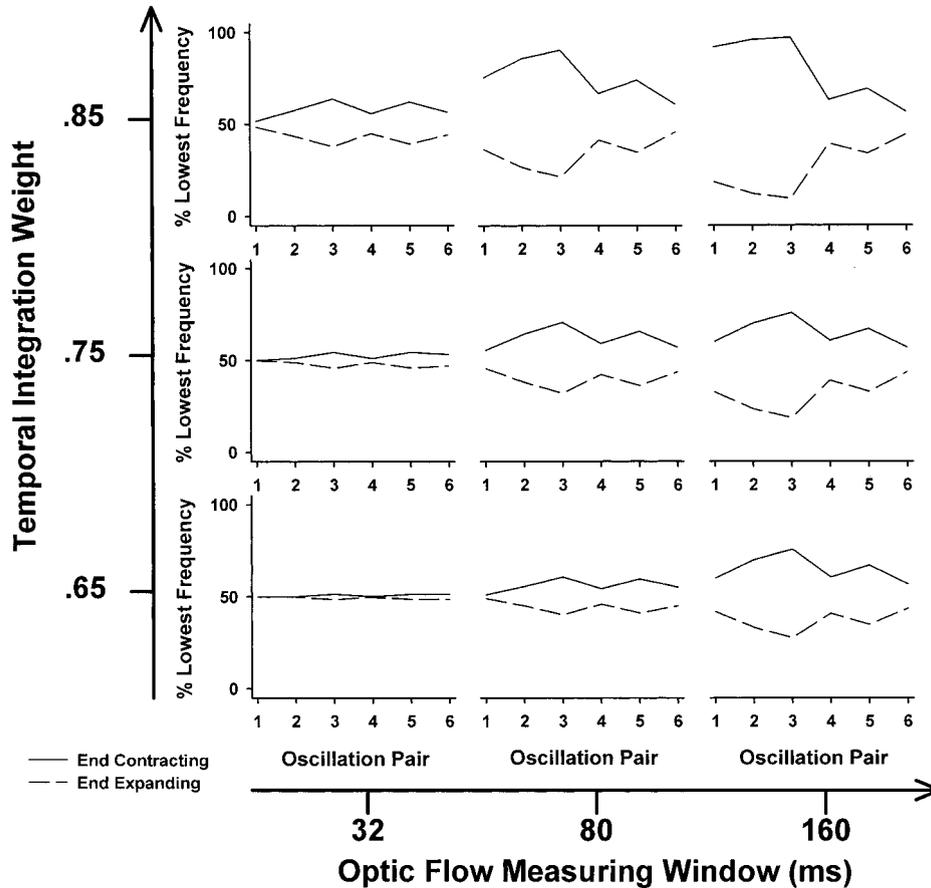


Figure 14. Predicted percentages for the lowest frequency sequence being selected as more slanted when two sequences with different oscillation frequencies are compared, for end-contracting (solid lines) and end-expanding (dashed lines) sequences. The six possible comparisons are labeled 1 through 6 on the x-axis and correspond to the pairs (c, 1), (c, 2), (c, 4), (1, 2), (1, 4), and (2, 4), respectively, where c indicates constant optic flow.

The test sequences were generated by extracting either the first or the last 10 frames from the comparison sequences. Frames 1–10 of the comparison sequences were labeled the *beginning-portion* (BP) test sequences; frames 170–180 of the comparison sequences were labeled the *end-portion* (EP) test sequences.

Each trial began with the comparison sequence appearing on the left or on the right of a central fixation mark (a 0.2-cm white square). After 170 frames (10 frames before the end of the comparison sequence), the test sequence appeared on the opposite side of the fixation mark. The comparison and test sequences ended at the same time.

Apparatus. The apparatus was the same as in Experiment 1.

Design. Each participant completed four blocks of 80 trials. Four variables were manipulated: sign of the gradient ϕ_x (contraction or expansion of the optic flow), time variation of the gradient ϕ_x (decreasing or increasing gradient), kind of test stimulus (BP or EP), and position of the test sequence (to the left or to the right of the comparison sequence). All of the variables were within-subjects variables. Participants viewed five presentations of the 16 experimental conditions, with the order of trials randomized within each block, for a total of 320 trials (20 trials for each condition).

Procedure. The procedure and the task were similar to those of the previous experiments. The participants' task was to decide whether the perceived slant of the test sequence was larger or smaller than the per-

ceived slant of the comparison sequence after both sequences had ended. The participants took part individually in either one session with four blocks or two sessions with two blocks per session. In both cases, participants took a brief break after each block. Practice was provided before the experimental sessions to familiarize the participants with the stimulus displays.

Simulations

To describe the predictions of the surface-orientation-update model, we represent an SFM display with a series of gradient values ($\phi_1, \phi_2, \dots, \phi_n$), where n represents the number of measurements of the optic flow obtained in the course of the stimulus sequence. If we assume that the optic flow is measured every 10 frames, then $\phi_1, \phi_2, \dots, \phi_{18}$ represents the gradient-measurement series that can be obtained from the comparison sequence. Figure 16 shows the absolute values of the gradients in the increasing-gradient condition (left) and the decreasing-gradient condition (right). The gradient of the BP test sequence is ϕ_1 (large open circle in Figure 16), and the gradient of the EP test sequence is ϕ_{18} (large open square in Figure 16). It is important to bear in mind

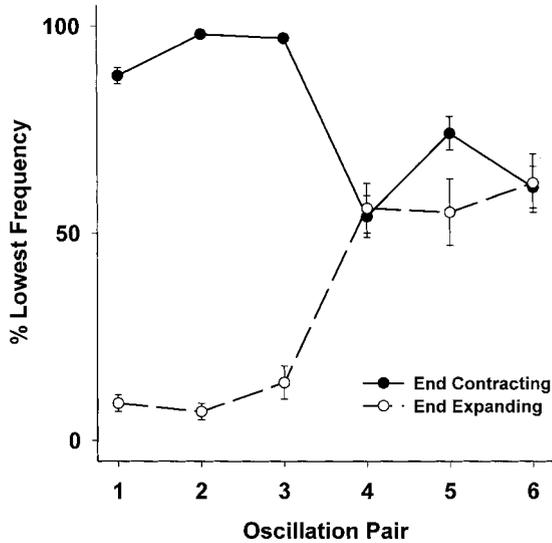


Figure 15. Mean percentages of the lowest frequency sequence being judged more slanted in Experiment 2, for end-contracting (black circles) and end-expanding (open circles) sequences, as a function of the six possible comparisons, labeled 1 through 6 and corresponding to (c, 1), (c, 2), (c, 4), (1, 2), (1, 4), and (2, 4), respectively. Error bars represent standard errors of the means.

that the graphs of Figure 16 do not specify whether the optic flow is contracting or expanding. This is specified by the sign of the gradient. The flow is expanding if the sign of the gradient is positive and contracting if the sign is negative.

We now examine in greater detail the contracting–decreasing condition. Because the optic flow is contracting, in this condition the velocity field specifies a rotation away from the frontal–parallel plane and, therefore, an increasing slant; the decrease of the gradient over time, on the other hand, specifies a decreasing slant. Given this conflicting information, the slant derived by the model in different moments of time depends on the magnitude of the temporal integration weight w (see Equation 4). Figure 17 (top left) shows the slant magnitudes derived by the model in the course of this stimulus sequence, for different temporal integration weights, with the parameter $\Delta t = 160$ ms (the size of the temporal integration window suggested by the results of Experiment 2). If $w = 0$ (no long-term temporal integration), the derived slant depends only on the current value of the gradient ϕ_p and therefore the magnitude of predicted slant decreases over time. These predicted slant magnitudes are represented in the figure by the solid black curves with dash marks. If $w = 1$ (solid black curve), the long-term temporal integration model sums up the rotation associated with each interval. Because the rotation is away from the image plane, the derived slant magnitude increases over time. Finally, if $1 > w > 0$, the magnitude of derived slant initially increases, then reaches a maximum, and finally decreases. The remaining parts of Figure 17 illustrate the predictions of the model in the other three stimulus conditions.

Because the participants were asked to decide whether the slant perceived at the end of the test sequence was larger or smaller than the slant perceived at the end of the comparison sequence, the predictions of the model can be formulated by comparing the

derived slant magnitudes at the end of the comparison sequence, on the one side, and the derived slant magnitudes of the (BP and EP) test sequences, on the other (see Figure 17). Bear in mind that the test sequences are identical to either the beginning or the end portions of the comparison sequences. It is important to note that the EP test sequence has the same gradient as the comparison sequence at the end of the stimulus display. The solid squares and circles of Figure 17 indicate the predicted slant magnitudes for the BP and EP sequences, respectively. Because we assume that only one velocity-gradient measurement can be obtained from these sequences (having set Δt to 160 ms), in this case the predicted slant magnitudes do not depend on the size of the temporal integration weight w .

It is especially interesting to consider the results of the simulations in which $w < .9$ because, in these circumstances, the predicted slant for the BP test sequence is larger than the predicted slant at the end of the comparison sequence. This result is paradoxical, because the contracting optic flow gives rise to the perception of a surface constantly rotating away from the image plane, whereas the model predicts that the beginning portion of this sequence should evoke a larger slant than its final portion. If $w > 0$, moreover, the predicted slant for the EP test sequence is smaller than the predicted slant at the end of the comparison sequence, even if the two corresponding velocity gradients are identical. The predictions of the surface-orientation-update model in the other three conditions (contracting–increasing, expanding–decreasing, and expanding–increasing) can be easily formulated by examining Figure 17.

Results and Discussion

The mean percentages of trials (out of 20 repetitions) in which the participants judged the test sequence to be more slanted than the comparison sequence, as a function of test sequence type (BP or EP) and time variation of the gradient (increasing or decreasing), are shown in Figure 18 (solid lines). We consider the BP test sequences first.

In the contracting–decreasing condition, the BP test sequence was judged to be more slanted than the comparison sequence in 98% of the cases. Such a result is paradoxical, because it means

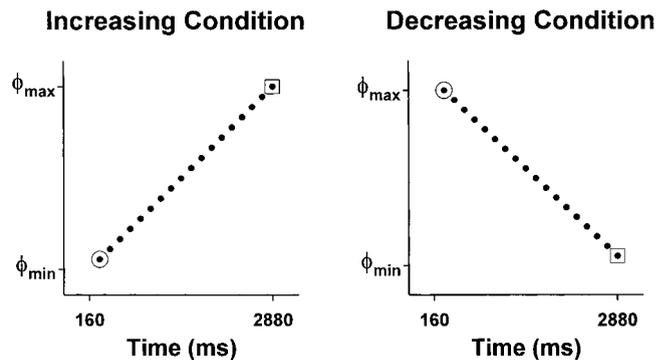


Figure 16. Time variations of the horizontal gradient component, ϕ_x , for the stimuli of Experiment 3, in the increasing-gradient (left) and decreasing-gradient (right) conditions. The large open circle represents the beginning-portion test sequence; the large open square represents the end-portion test sequence. max = maximum; min = minimum.

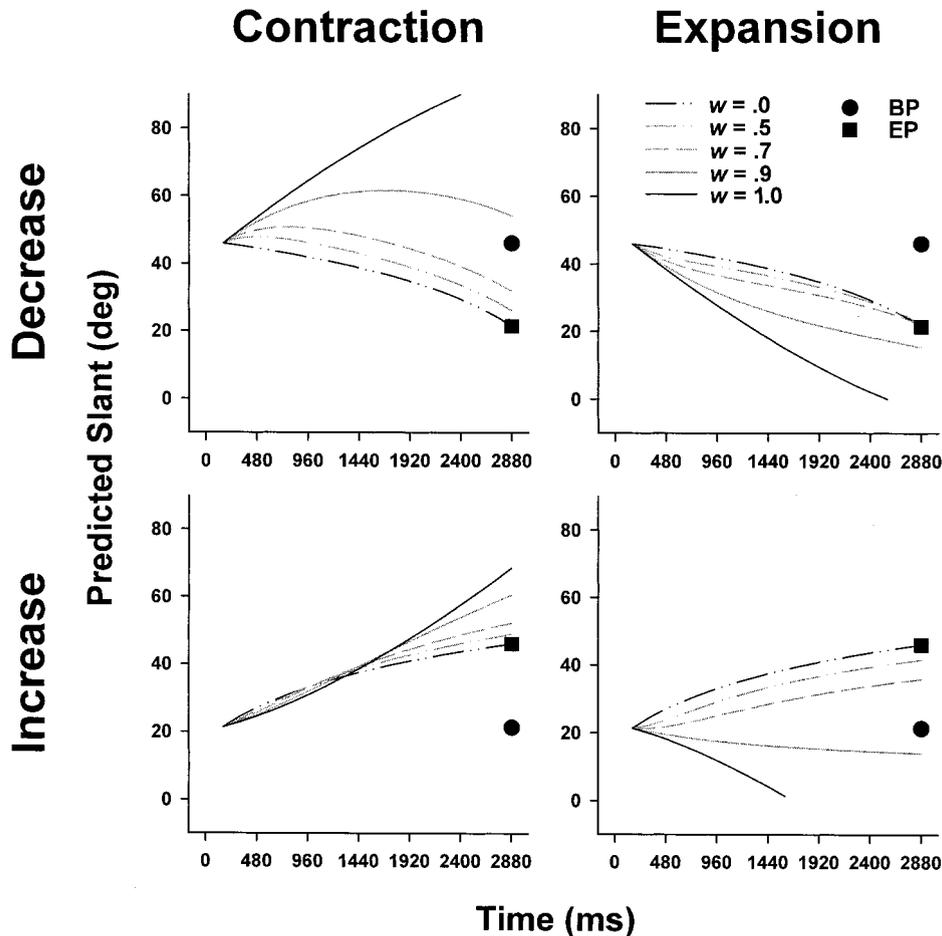


Figure 17. Time variation of the derived slant magnitudes for the comparison sequence of Experiment 3. The slant magnitudes were calculated for each 160-ms step and for different values of the parameter w (0, .5, .7, .9, and 1.0). The black circles and black squares represent the derived slants for the beginning-portion (BP) and end-portion (EP) test sequences. deg = degrees.

that the initial portion of a stimulus sequence depicting a surface rotating away from the image plane evokes a larger slant than the final portion of the same sequence. A similar result was found in the expanding-increasing condition. In that case, the BP test was judged to be less slanted than the comparison sequence in 79% of the cases. Again, such a result is paradoxical, because it means that the initial portion of a stimulus sequence depicting a surface rotating toward the image plane evokes a smaller slant than the final portion of the same sequence. These psychophysical data are consistent with the output of the surface-orientation-update model when $w > 0$ but cannot be accounted for if $w = 0$ (no long-term temporal integration).

In the contracting-increasing condition, the optic flow does not provide conflicting information, and therefore the BP test sequences should appear to be less slanted than the comparison sequences. In 99% of the cases, in fact, the participants' judgments were consistent with this prediction. In a similar manner, no conflicting information is provided in the expanding-decreasing condition, and, accordingly, in 77.5% of the cases the BP test sequences were judged to be more slanted than the comparison sequences.

Consider now the EP sequences, that is, the cases in which the test sequences were identical to the final part of the comparison sequences. Even when the velocity gradients were the same in both cases, in the contracting condition participants did not perceive the test and comparison sequences as having the same slant magnitude. In the contracting-decreasing condition, participants judged the test sequence to be less slanted than the comparison sequence in 72% of the cases, $t(7) = -8.35$, $p < .001$. In the contracting-increasing condition, participants judged the test sequence to be less slanted than the comparison sequence in 67% of the cases, $t(7) = -3.95$, $p < .01$. Both of these results are consistent with the predictions of the surface-orientation-update model with $w > .5$ (see Figure 17). In the expanding-decreasing condition and in the expanding-increasing condition, finally, the participants' judgments were not significantly different from chance, $t(7) = -1.21$, ns , and $t(7) = -1.22$, ns , respectively.

As in the previous experiments, we computed the predictions of the surface-orientation-update model by fitting Equation 4 to the data with s , w , and Δt as free parameters. The estimated values of the parameters that minimized the RMSDs were $s = 19$, $w = .6$,

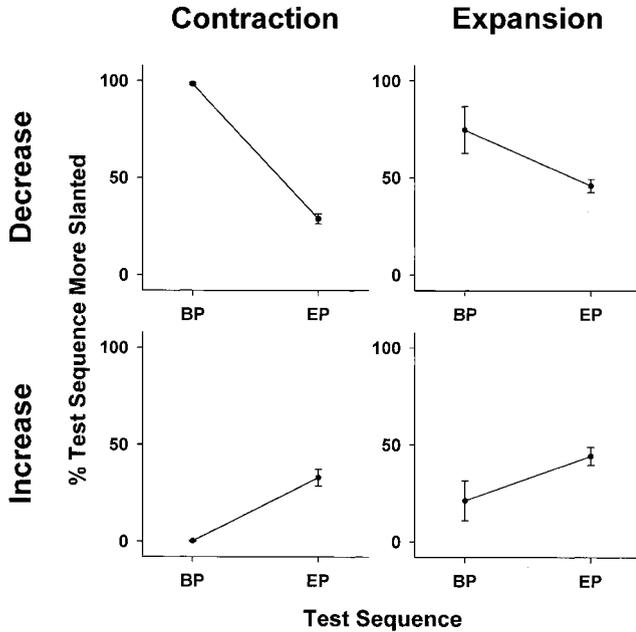


Figure 18. Mean percentages of the test stimulus being judged more slanted in Experiment 3, for the beginning-portion (BP) and end-portion (EP) sequences, in the contracting–decreasing, contracting–increasing, expanding–decreasing, and expanding–increasing conditions. Error bars represent standard errors of the means.

and $\Delta t = 0.176$ s and, when w was fixed to zero, $s = 19$ and $\Delta t = 0.016$ s. The RMSDs for the two models were, respectively, 9.96% and 12.91%.

In conclusion, the results of the present experiment reveal that the contracting–decreasing and expanding–increasing conditions give rise to paradoxical perceptions. These results can be predicted by the surface-orientation-update model, which also accounts for the results obtained in the contracting–increasing and expanding–decreasing conditions (in which no paradox was observed in the observers’ responses).

Experiment 4

The results of the previous experiments are consistent with the hypothesis that observers update a surface representation over time; that is, they combine the current motion gradients with the slants and angular velocities previously perceived. One alternative explanation of the results of the previous experiments, however, is that observers integrate the motion gradients over successive intervals without using an intervening computation of slant and angular velocity; that is, long-term temporal integration may act not at the level of a 3-D representation but at the lower level of 2-D motion processing. Motion detectors with different temporal resolutions, for example, could pool their activation over an extended period of time (e.g., Festa & Welch, 1997). The purpose of Experiment 4 was to test this alternative hypothesis. Using the same methodology as in the previous experiments, we compared performance on a 2-D speed-discrimination task with performance on an SFM task. To the extent that speed judgments do not rely on a 3-D representation, any qualitative differences in the results

obtained with the two tasks would suggest that different mechanisms are involved.

Method

Participants. Two naive participants and two of the authors (F.D. and Q.V.) took part in the present experiment.

Stimuli. Each stimulus sequence consisted of a history segment followed by a test segment. In the SFM task, planar surfaces were simulated as rotating counterclockwise about the vertical axis during the history segment and rotating downward about the horizontal axis during the test segment. In a similar manner, in the speed-discrimination task, a uniform velocity field representing a leftward translation was simulated during the history segment, followed by a downward translation during the test segment.

The optic-flow gradients for the two rotating surfaces are illustrated in Figure 19. As shown in the figure, during the test segment, one gradient was larger than the other. There were two possible history conditions. In the *more* condition, the optic-flow field having the largest gradient in the test segment exhibited the largest gradient in the history segment as well. In the *less* condition, the optic-flow field with the largest gradient in the test segment had the smallest gradient in the history segment. In the history segment, the gradients were 0.0625 s^{-1} and 0.375 s^{-1} (see Figure 19). The gradients for the test segment were determined for each participant in a pilot experiment (see Table 1) so as to obtain a rate of at least 80% correct responses (i.e., the largest slant attributed to the largest gradient).

The stimulus displays for the speed-discrimination task were similar to those used for the SFM task, the only difference being that uniform velocity fields simulating a pure translation were used. The 2-D velocities for these displays were equal to half of the mean velocities of the optic-flow gradients used in different conditions of the SFM task (Figure 19).

For both the SFM and speed-discrimination tasks, the history segment was made up of 50 frames (16 ms per frame). The length of the test segment was systematically varied, in 10-frame steps, from 10 to 50 frames. The two random-dot fields shown on each trial were contained in a circular region with a radius of 8.8 cm (5.6° of visual angle) and were separated by a blank region approximately 1 cm wide, with a distance of approximately 5 cm between their centers. Each circular region contained 250 dots.

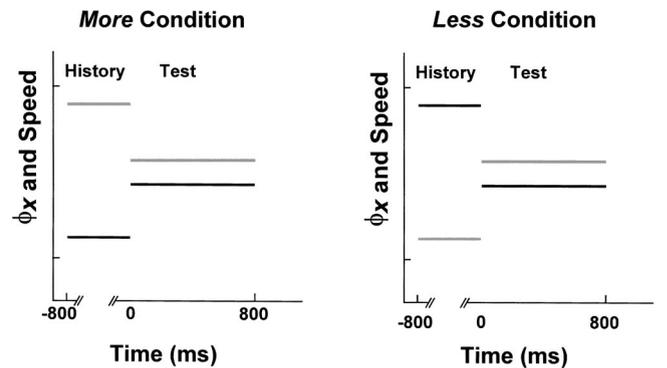


Figure 19. Time variations of the horizontal gradient (ϕ) or of the speed for the two stimulus sequences used in Experiment 4 (black and gray lines). For both sequences, the history segment lasted 800 ms, and the test segment lasted up to 800 ms. In the *more* condition (left), the sequence having the largest gradient or speed in the history segment also had the largest gradient or speed in the test segment. In the *less* condition (right), the sequence having the smallest gradient or speed in the history segment had the largest gradient or speed in the test segment.

Apparatus. The apparatus and experimental setup were the same as in the previous experiments, except that participants wore an eye patch over their nondominant eye rather than viewing the monitor through an aperture.

Design. There were four within-subjects variables: task (SFM vs. speed discrimination), position of the velocity field with the larger gradient or mean velocity (left vs. right), history condition (more vs. less), and number of frames of the test segment (10, 20, 30, 40, or 50). One naive participant (K.C.) and 1 expert participant (F.D.) completed the speed-discrimination task first, followed by the SFM task. The remaining 2 participants completed the tasks in the reverse order. For each task, participants completed four blocks of 100 trials. In each block, there were five repetitions of each of 20 conditions (2 types of history \times 5 durations \times 2 positions).

Procedure. The procedure and the task were similar to those of the previous experiments. In the SFM task, participants were asked to report whether the velocity field that evoked the largest perceived slant at the end of the stimulus sequence was located to the right or to the left. In the speed-discrimination task, they were asked to report whether the field having the largest speed was located to the right or to the left. Before the experimental sessions, each participant completed two additional blocks of 100 trials to determine the gradient difference needed to reliably associate a larger perceived slant with the largest velocity gradient in the SFM task or to correctly identify the largest speed in the speed-discrimination task (at least 80% correct judgments).

Results and Discussion

Figure 20 shows mean percentages of correct responses (i.e., percentages of trials in which the field with the largest gradient was correctly selected as having the largest slant in the SFM task or percentages of trials in which the field with the largest speed was correctly selected in the speed-discrimination task) as a function of the duration of the test segment. A mixed-design ANOVA with expertise as a between-subjects variable and position (left vs. right), task (SFM vs. speed discrimination), history condition (more vs. less), and number of frames (10–50) as the within-subjects variables was conducted; percentage of correct judgments was the dependent variable. This analysis did not show a significant main effect of expertise, nor did expertise significantly interact with any of the other variables. The main effect of the number of frames of the test segment was significant, $F(4, 8) = 26.62, p <$

.001. This effect, however, must be interpreted as a by-product of the strong three-way interaction among task, history condition, and number of frames, $F(4, 8) = 8.72, p < .01$. The meaning of this interaction is illustrated in Figure 20.

Figure 20, left, shows the results for the *less* stimulus condition. When the displays were shown with no history segment, in the SFM task participants attributed the largest slant to the optic-flow field with the largest gradient in at least 80% of the cases. In a manner consistent with the surface-orientation-update model, when the same flow fields followed a velocity field having a much smaller gradient (*less* condition), a smaller magnitude of slant was perceived. So, for the 10-frame test segments, on average, participants judged the optic-flow field with the largest gradient as having the largest slant in only 46% of the trials. This percentage steadily increased as the length of the test segment increased, up to 86% for 50-frame test segments. In the *more* stimulus condition, the test sequences were preceded by a flow field having a much larger gradient. According to the surface-orientation-update model, in these conditions a larger magnitude of slant should be perceived, thus producing a ceiling effect. Consistent with this prediction, in the *more* condition, participants' judgments were not affected by the length of the test segment. On average, participants attributed the largest slant to the optic-flow field with the largest gradient in 88% of the trials. Overall, the results of the SFM task were consistent with those of the previous experiments.

We examine now the results of the speed-discrimination task, starting with the *less* condition. In this case, unlike the SFM task, the participants' judgments were not affected by the length of the test segment. On average, participants attributed the largest speed to the velocity field with the largest velocity in 94% of the trials. In the *more* condition, conversely, speed-discrimination judgments were affected by the length of the test segment. The percentage of correct responses was as low as 42% for the 10-frame test and increased up to 85% for the 50-frame test. Again, this result contrasts sharply with the nil effect of the length of the test sequence in the *more* condition of the SFM task. Even if the history effect in the speed-discrimination task can be related to temporal integration in motion perception (e.g., Burr & Santoro, 2001; Fredericksen, Verstraten, & van de Grind, 1994a, 1994b; Raymond & Isaak, 1998; Watamaniuk & Sekuler, 1992), the important point for present purposes is that the results of the speed-discrimination task followed a pattern completely different from those of the SFM task. This suggests, therefore, that the long-term temporal integration effects revealed in the present investigation do not pertain to the lower level of motion measurements; rather, they are specific to the perceptual recovery of the 3-D properties from the optic flow.

General Discussion

In four experiments, participants were asked to indicate which of two adjacent SFM displays appeared to be more slanted in depth at the end of the stimulus sequence. According to current SFM models, the slant perceived at one moment in time is not influenced by the properties that the velocity gradients take on outside a small temporal window of about 200 ms that is necessary to measure the properties of the optic flow (e.g., Treue et al., 1991; van Damme & van de Grind, 1996). Contrary to this assumption, we found that judgments of surface slant covaried with the values

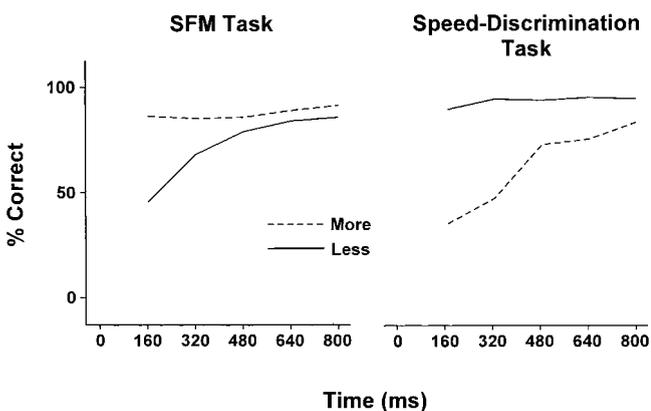


Figure 20. Mean percentages of correct responses, as a function of the duration of the test segment in Experiment 4, for the structure-from-motion (SFM) task (left) and the speed-discrimination task (right) in the two experimental conditions (*more* and *less*).

that the velocity gradients of the stimulus displays took on up to 1 s before the judgments were made, thus suggesting that human SFM is modulated by a process of long-term temporal integration.

In the first experiment, we investigated how the initial part of an SFM display (history segment) influences the perceived slant of the final part of the same sequence (test segment). Two SFM sequences were presented side by side in each trial. The gradients of the two velocity fields during the test phase were slightly different, and thus if participants were shown the test sequences only, they reliably judged the sequence with the largest gradient as having the largest slant. When the same test sequences followed opportunely devised history sequences, however, the opposite result was obtained: The test sequence with the smallest gradient was judged to be more slanted than the test sequence with the largest gradient. Such a dramatic influence of the history sequence was found to last up to 1,280 ms.

In the second experiment, these findings were extended to oscillatory sequences (i.e., sequences in which the optic-flow gradient changed in a cyclic fashion). Even if the two SFM displays presented in each trial exhibited the same velocity gradients at the end of the oscillation sequence, we found that perceived slant was affected both by the frequency of oscillation and by the sign of the gradient. For contracting optic flows (i.e., rotations away from the image plane), the sequence with the lowest oscillation frequency appeared more slanted; for expanding optic flows (i.e., rotations toward the image plane), the sequence with the highest oscillation frequency appeared more slanted.

In the third experiment, we explored the seemingly paradoxical consequences of long-term temporal integration for perceived SFM. For a contracting optic flow with a gradient decreasing over time, we found that the initial segment of the sequence evoked larger magnitudes of perceived slant than the final segment of the same sequence. This result is paradoxical, because a contracting flow represents a surface that is rotating away from the image plane: in other words, a surface whose slant continuously increases over time.

Finally, in the fourth experiment, we compared performance on an SFM task with performance on a speed-discrimination task. The results of this experiment support the hypothesis that the temporal integration effects described here are specific to the perceptual recovery of the 3-D properties from the optic flow and rule out the alternative hypothesis ascribing them to the lower level of 2-D motion measurements.

To account for these results, we propose a model that combines the information provided by two potentially conflicting sources: the representation of surface orientation in an immediately preceding moment in time, on the one hand, and the current optic-flow gradient, on the other. The long-term temporal integration model assumes that (a) the optic flow is measured during the time interval Δt (this process has been studied, for example, by Treue et al., 1991, and by van Damme & van de Grind, 1996, and we label it *short-term temporal integration*), (b) the perceptual analysis of the optic flow is based only on the first-order properties of the velocity field (extensive evidence for this has been provided; e.g., Domini & Caudek, 1999), and (c) perceived slant is computed as the weighted average between the current slant value and the slant and angular-rotation values associated with the same surface location in previous moments in time (we label this process *long-term temporal integration*).

The output of the proposed model is consistent with all of the results of the present experiments. In this regard, it should be noted that our aim was not to obtain the best quantitative agreement between the model's predictions and the psychophysical data. The performance of individual observers, in fact, depends on factors such as their level of expertise in the task at hand, their overall alertness, or, generally, the level of internal noise (Pelli & Farell, 1999), all factors that we did not wish to include in the model (for a similar approach, see Caudek & Rubin, 2001). Instead, we focus on the strong qualitative agreement between the trends shown by the psychophysical data and those predicted by the surface-orientation-update model ($w > 0$), and we note that this agreement is in stark contrast to the predictions that can be made when long-term temporal integration is ruled out ($w = 0$).

Two factors determine the output of our model: the size of the temporal window Δt during which the optic flow is measured and the temporal integration weight w (see Equation 4). Consider the temporal window first. Even though the quantification of the time needed for visual measurement of 2-D motion has been much debated, it is clear that this process requires at least several milliseconds (e.g., Festa & Welch, 1997). It has been shown, for example, that the minimum physiological delay for a bilocal correlator (Reichardt, 1961) is approximately 50 ms (Koenderink, Van Doorn, & van de Grind, 1985; Todd & Norman, 1995). Psychophysical studies have revealed, moreover, that visual motion processes may integrate multiple events to determine specific characteristics (e.g., motion direction) and that this integration process may take up to 200–300 ms (Welch, MacLeod, & McKee, 1997).

In an earlier investigation, Todd, Akerstrom, Reichel, and Hayes (1988) investigated the optimal spatial-temporal parameters that allow a reliable perception of rigid structure in dynamic displays. In one experiment directly relevant to the present discussion, they showed that a two-frame random-dot sequence depicting a smooth velocity field gives rise to a rigid percept if each frame is displayed for at least 50–100 ms. This and other findings (e.g., Treue et al., 1991; van Damme & van de Grind, 1996), therefore, suggest that a processing time of up to 200 ms may be required to derive a rigid 3-D structure from the optic flow. Consistent with these results, we found that our data are best fit by assuming that the measurement of the optic-flow gradient requires between 80 ms and 160 ms.

The second factor affecting the model's output is the temporal integration weight w . This weight encodes the long-term memory of the system. If $w = 0$, the current slant estimates do not depend on those obtained in previous moments (i.e., no long-term temporal integration is involved). If $w = 1$, each slant value currently computed is incremented by the entire amount of angular rotation recovered from the previous optic-flow gradients. In all four experiments, the best fits were obtained with temporal integration weights in the range of .6–.9. This suggests that the perceptual recovery of 3-D information from the optic flow is affected by long-term temporal integration but also that temporal integration occurs within a moving temporal window with a fixed size.

The hypothesis that human SFM updates a 3-D representation over time has already been proposed in Ullman's (1984) incremental-rigidity scheme. Ullman's algorithm updates a 3-D model by considering the new image positions of the projected object's features. Initially, the object representation is flat. As new frames of the motion sequence become available, the algorithm computes

the 3-D coordinates of the object representation so as to maximize the rigidity of the transformations from the previous to the new object representation. It is important to note that, according to this algorithm, it is the accuracy of a 3-D representation that builds up over time.

Hildreth et al. (1990) presented empirical evidence suggesting that the accuracy of human performance in an SFM task improves over time, in agreement with Ullman's incremental-rigidity scheme. Alternative interpretations of Hildreth et al.'s data have been proposed (Todd & Bressan, 1990), but this study represents an important attempt at establishing the role of temporal integration in human SFM. More recently, Hildreth et al. (1995) and Treue et al. (1995) proposed another model motivated by Ullman's incremental rigidity scheme. Within a closed-loop architecture, this model is characterized by four processing stages. In one stage, the 2-D velocities are extracted. In a second stage, the 3-D velocities of the projected features are computed so as to maximize the rigidity of the 3-D configuration, and the depths of the projected features are computed on the basis of the 3-D and 2-D velocities. In the third stage, the depth estimates are averaged over an extended time period. In the fourth stage, a smooth 3-D surface is fitted to the estimated depth values of sparse image positions.

The model of Hildreth et al. (1995), although successful from a computational standpoint, is less suited to account for the long-term process of surface-orientation update investigated in the present study. This point can be illustrated in greater detail, for example, by considering the predictions of the model of Hildreth et al. (1995) for the contracting optic-flow stimuli of Experiment 3. For a contracting optic flow, the model of Hildreth et al. predicts a continuously increasing slant magnitude (see Figure 21, as compared with the predictions of our model shown in the top left of Figure 17). As a consequence, such a model cannot account for, in 98% of the cases, observers judging the beginning portion of the test sequence to be more slanted than the comparison sequence. This result means that the initial portion of a stimulus sequence depicting a surface rotating away from the image plane is judged as having a larger slant than the final portion of the same sequence, and it is obviously incompatible with any veridical analysis of the optic flow.

Further evidence for a process of temporal update of a 3-D representation comes from a study conducted by Domini et al. (2001). In this study, Domini et al. showed that the perceptual derivation of the slant magnitudes from the optic flow can be influenced by the disparity information that had been presented in a previous moment in time. In these experiments, two stereograms specifying planar surfaces slanted about the horizontal axis (0° or 45°) were presented during the first half of the total display duration (1 s). After 500 ms, the random dots presented to one eye were replaced by a blank field, and the random dots presented to the other eye were animated so as to produce two constant optic-flow fields specifying planar surfaces (test vs. comparison) rotating in depth about the horizontal axis. Observers were asked to compare the slant specified by the test and comparison motion fields. When the two motion stimuli were preceded by the same disparity field (either 0° or 45° for both stimuli), observers reliably associated the largest magnitude of perceived slant with the largest velocity gradient. Conversely, when one motion field followed the 0° stereogram and the other motion field followed the 45° stereogram, a bias was observed. In these conditions, perceived slant was

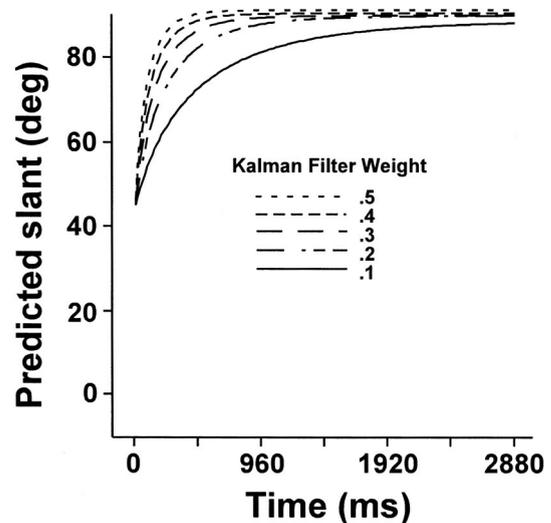


Figure 21. Time variations of the derived slant magnitudes for the comparison sequence of Experiment 3, according to a simplified version of the model of Hildreth et al. (1995). The simulation assumes that, after 160 ms, a 3-D representation of a planar surface has been recovered. To allow a better comparison of the two models, we also assumed that the model of Hildreth et al. initially recovers the same slant magnitude as our model. In the simulation, the predicted slant is computed by using the Kalman filter procedure described by Hildreth et al. The different curves refer to different values of the Kalman filter weight. Although this simulation is purely qualitative, it is important to note that the predicted slant always increases with time. The predictions of the model, therefore, are inconsistent with the results of Experiment 3. deg = degrees.

enhanced for the motion gradient following the 45° stereogram and reduced for the motion gradient following the 0° stereogram. The influence of the disparity gradients on the perceptual analysis of the optic flow was weakened by increasing the length of the motion sequences but persisted for at least 800 ms.

Conclusion

The results of the present experiments indicate that, for both constant optic-flow fields and displays providing second-order temporal information, surface orientation perceived at one moment in time is affected by the optic-flow properties at previous moments in time. The present findings can be accounted for by a temporal integration model assuming that (a) a 3-D representation is derived heuristically from the first-order velocity field and (b) perceived local surface orientation is updated by averaging the slant magnitudes specified by the current optic flow with the slant and angular rotation magnitudes perceived at previous moments in time. Unlike previous models, the temporal integration model proposed here is consistent with both veridical and nonveridical human performance.

References

- Ando, H. (1991). Dynamic reconstruction of 3D structure and 3D motion. In *Proceedings of the IEEE Workshop on Visual Motion* (pp. 101–110). Washington, DC: IEEE Computer Society Press.
- Atchley, P., Andersen, G. J., & Wuestefeld, A. P. (1998). Cooperativity,

- priming, and 3-D surface detection from optic flow. *Perception & Psychophysics*, *60*, 981–992.
- Bennett, B. M., Hoffman, D. D., Nicola, J. E., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America*, *6*(A), 1052–1069.
- Braunstein, M. L. (1976). *Depth perception through motion*. New York: Academic Press.
- Braunstein, M. L. (1994). Decoding principles, heuristics and inference in visual perception. In G. Johansson, S. S. Bergstrom, & W. Epstein (Eds.), *Perceiving events and objects* (pp. 436–446). Hillsdale, NJ: Erlbaum.
- Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1990). Discriminating rigid from non-rigid motion: Minimum points and views. *Perception & 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.
- Braunstein, M. L., Liter, C. J., & Tittle, J. S. (1993). Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 598–614.
- 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., & Rubin, N. (2001). Segmentation in structure from motion: Modeling and psychophysics. *Vision Research*, *41*, 2715–2732.
- 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*, 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 & Psychophysics*, *60*, 1164–1174.
- Domini, F., Caudek, C., Turner, J., & Favretto, A. (1998). Discriminating constant from variable angular velocities in the kinetic depth effect. *Perception & Psychophysics*, *60*, 747–760.
- Domini, F., Skirko, P., & Caudek, C. (2001, May). *Temporal integration of stereo and motion information*. Paper presented at the meeting of the Vision Sciences Society, Sarasota, FL.
- Eby, D. W. (1992). The spatial and temporal characteristics of perceiving 3-D structure from motion. *Perception & Psychophysics*, *51*, 163–178.
- Festa, E. K., & Welch, L. (1997). Recruitment mechanisms in speed and fine-direction discrimination tasks. *Vision Research*, *37*, 3129–3143.
- Fredericksen, R. E., Verstraten, F. A., & van de Grind, W. A. (1994a). An analysis of the temporal integration mechanism in human motion perception. *Vision Research*, *34*, 3153–3170.
- Fredericksen, R. E., Verstraten, F. A., & van de Grind, W. A. (1994b). Spatial summation and its interaction with the temporal integration mechanism in human motion perception. *Vision Research*, *34*, 3171–3188.
- Freeman, T. C. A., Harris, M. G., & Meese, T. S. (1995). On the relationship between deformation and perceived surface slant. *Vision Research*, *35*, 317–322.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Heel, J. (1990). Direct dynamic motion vision. In *Proceedings of the IEEE Conference on Robotics and Automation* (pp. 1142–1147). Washington, DC: IEEE Robotics and Automation Society Press.
- Hildreth, E. C., 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 & Psychophysics*, *48*, 19–36.
- Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America*, *72*(A), 888–892.
- Hoffman, D. D., & Bennett, B. M. (1985). Inferring the relative 3-D positions of two moving points. *Journal of the Optical Society of America*, *2*(A), 350–353.
- Hoffman, D. D., & Bennett, B. M. (1986). The computation of structure from fixed-axis motion: Rigid structures. *Biological Cybernetics*, *54*, 71–83.
- Hung, Y. S., & Ho, H. T. (1999). A Kalman filter approach to direct depth estimation incorporating surface structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*, 570–575.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, *82*, 35–45.
- Koenderink, J. J. (1986). Optic flow. *Vision Research*, *26*, 161–180.
- Koenderink, J. J., & Van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, *22*, 773–791.
- Koenderink, J. J., & Van Doorn, A. J. (1976). Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, *66*(A), 717–723.
- Koenderink, J. J., & Van Doorn, A. J. (1990). Affine structure from motion. *Journal of the Optical Society of America*, *8*(A), 377–385.
- Koenderink, J. J., Van Doorn, A. J., & van de Grind, W. A. (1985). Spatial and temporal parameters of motion detection in the peripheral visual field. *Journal of the Optical Society of America*, *2*(A), 252–259.
- Liter, J. C., & Braunstein, M. L. (1998). The relationship of vertical and horizontal velocity gradients in the perception of shape, rotation, and rigidity. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1257–1272.
- 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., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London, Series B*, *208*, 385–397.
- Loomis, J. M., & Eby, D. E. (1988). Perceiving structure from motion: Failure of shape constancy. In *Proceedings of the Second International Conference on Computer Vision* (pp. 383–391). Washington, DC: IEEE Computer Society Press.
- Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research*, *38*, 2817–2832.
- Matthies, L., Kanade, T., & Szeliski, R. (1989). Kalman filter-based algorithm for estimating depth from image sequences. *International Journal of Computer Vision*, *3*, 209–236.
- 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: MIT Press.
- Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotary objects undergoing affine stretching transformations. *Perception & Psychophysics*, *53*, 279–291.
- Norman, J. F., & Todd, J. T. (1995). The perception of 3-D structure from

- contradictory optical patterns. *Perception & Psychophysics*, 57, 826–834.
- Ono, M. E., Rivest, J., & Ono, H. (1986). Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 331–337.
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America*, 16(A), 647–653.
- Perotti, V. J., Todd, J. T., & Norman, J. F. (1996). The visual perception of rigid motion from constant flow fields. *Perception & Psychophysics*, 58, 666–679.
- Prazdny, K. (1980). Egomotion and relative depth map from optical flow. *Biological Cybernetics*, 36, 87–102.
- Raymond, J. E., & Isaak, M. (1998). Successive episodes produce direction contrast effects in motion perception. *Vision Research*, 38, 579–589.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 303–317). New York: Wiley.
- Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, 8, 125–134.
- Sperling, G., Landy, M. S., Doshier, B. A., & Perkins, M. E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 826–840.
- 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 & Psychophysics*, 43, 179–188.
- Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, 48, 419–430.
- Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics*, 50, 509–523.
- Todd, J. T., & Norman, J. F. (1995). The effects of spatiotemporal integration on maximum displacement thresholds for the detection of coherent motion. *Vision Research*, 35, 2287–2302.
- Todd, J. T., & Perotti, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception & Psychophysics*, 61, 1577–1589.
- 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.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, 13, 255–274.
- van Damme, W. J., & van de Grind, W. A. (1996). Non-visual information in structure-from-motion. *Vision Research*, 36, 3119–3127.
- Watamaniuk, S. N., & Sekuler, R. (1992). Temporal and spatial integration in dynamic random-dot stimuli. *Vision Research*, 32, 2341–2347.
- Welch, L., MacLeod, D. I. A., & McKee, S. P. (1997). Motion interference: Perturbing perceived direction. *Vision Research*, 37, 2725–2736.

Received January 3, 2001

Revision received August 10, 2001

Accepted November 27, 2001 ■

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.