PERGAMON

Vision Research 42 (2002) 1213–1223

Vision Research

www.elsevier.com/locate/visres

Short-term temporal recruitment in structure from motion

Corrado Caudek [a,*], Fulvio Domini [b], Massimiliano Di Luca [b]

[a] *Department of Psychology, University of Trieste, Via S Anastasio 12, 34134 Trieste, Italy*
[b] *Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02192, USA*

Received 18 September 2001; received in revised form 12 December 2001

Abstract

Temporal integration was investigated in the minimal conditions necessary to perform a structure-from-motion (SFM) task. Observers were asked to discriminate three-dimensional (3D) surface orientations in conditions in which the stimulus displays simulated velocity fields providing, in each frame transition, either sufficient (3 moving dots) or insufficient information (1 or 2 moving dots) to perform the task. When only two moving dots were shown in each frame transition of the stimulus displays (Experiment 1), we found that performance decreased as dot-lifetime increased. A facilitation effect of the overall display duration was also found. The negative effect of dot-lifetime on performance contrasts with what found in Experiment 2 with three dots in each frame transition, where performance improved with increasing dot-lifetime up to 170 ms, and then reached a plateau. Finally, for an optimal dot-lifetime of 150 ms, we found that performance was still above chance when each frame transition specified the motion of only one dot (Experiment 3). These results indicate that temporal recruitment alone can support the recovery of 3D information from sparse motion signals, thus providing a strong indication for the importance of temporal integration in the perceptual analysis of the optic flow. Our results reveal, moreover, that temporal integration in SFM has different characteristics, depending on whether, in each frame transition, the stimulus displays provide either sufficient (3 or more moving dots) or insufficient information (1 or 2 moving dots) to specify the higher-order properties of the optic flow necessary for 3D surface recovery.

Keywords: Visual motion; Structure from motion; Temporal integration; Optic flow

1. Introduction

Spatial and temporal integration play an important role in human motion processing. *Spatial integration* is revealed by the fact that, in some stimulus conditions, perceptual performance benefits from an increase in the number motion signals. Some form of combination, in fact, is used to overcome ambiguities (Hildreth, 1984; Yuille & Grzywacz, 1988) and noise (van Doorn & Koenderink, 1983) of the local velocities. *Temporal integration* is revealed by the fact that, in some conditions, perceptual performance benefits from the increase of stimulus duration. Some form of sequential recruitment, in fact, enhances the signals which indicate similar speeds and directions (Adelson & Bergen, 1985; van Santen & Sperling, 1984).

In the present paper we will focus on temporal integration. Whereas a wealth of investigations has studied the process of temporal integration when observers are asked to detect the presence of motion (e.g., Todd & Norman, 1995), or to discriminate between different motion directions or speeds (e.g., Festa & Welch, 1997; Snowden & Braddick, 1991), very few investigations have studied the role of temporal integration in structure-from-motion (SFM), that is, when observers are asked to recover information about three-dimensional (3D) shape from the optic flow (e.g., Braunstein, Hoffman, & Pollick, 1990; Braunstein, Hoffman, Shapiro, & Andersen, 1987; Caudek & Domini, 1998; Domini & Caudek, 1999; Domini, Caudek, & Proffitt, 1997; Hildreth, Grzywacz, Adelson, & Inada, 1990; Landy, Dosher, Sperling, & Perkins, 1991; Lappin, Doner, & Kottas, 1980; Todd & Bressan, 1990). In the present investigation, temporal integration will be studied by considering the minimal conditions which allow the recovery of 3D information from a velocity field. At this level, it is difficult to disentangle the two alternative

* Corresponding author. Tel.: +39-040-6762739; fax: +39-040-4528022.
E-mail address: caudek@units.it (C. Caudek).

0042-6989/02/$ - see front matter
PII: S0042-6989(02)00052-4

hypotheses according to which: (1) temporal integration occurs at the level of the 3D representation ensuing from the perceptual analysis of the optic flow, or (2) temporal integration occurs so as to provide a better measurement of the optic flow from which, successively, a 3D representation can be formed. Even if this remains an open question, it is nevertheless worthwhile to investigate temporal integration in motion processing when the perceptual task requires an analysis of the higher-order properties of a velocity field (e.g., the ratio between the vertical and the horizontal velocity gradients), rather than the more basic motion detection or motion-direction discrimination tasks. The development of any psychologically plausible SFM model, in fact, must take into account the temporal properties of the process by which the perceptual system is able to recover 3D information from the optic flow. Before presenting the rationale that motivated the present experiments, however, it is necessary to introduce some concepts that describe the known relations between the optic flow and the perceived 3D structure. For the current purposes, it will be sufficient to consider the simplest case of the optic flow produced by the rotation of a planar surface.

2. Perceptual interpretation of a linear velocity field

The velocity field produced by the orthographic projection of a rigidly rotating *planar* surface is informative about the 3D orientation of the projected surface and about its 3D motion. The 3D 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 projection of the normal to the surface on the image plane and the x-axis (see Fig. 1). The 3D motion of the surface, conversely, can be described by specifying the angular velocity and the orientation of the axis of rotation. Since we will consider only rotations about a vertical axis parallel to the image plane, the rotation of the surface is fully specified by the intensity (ω) of the angular velocity.

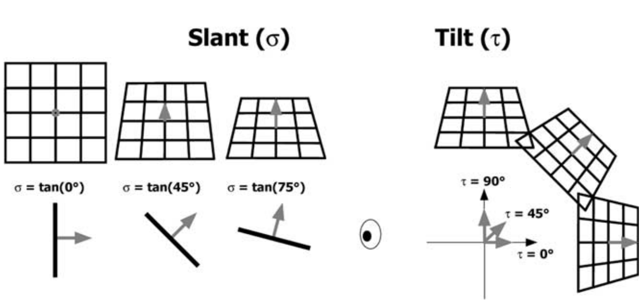

Fig. 1. *Left panel*: surface slant. A regular checkerboard is shown with slants of 0°, 45° and 75° (top panel). The side-view of the same checkerboard is shown in the bottom panel. The arrow indicates the normal to the surface. *Right panel*: surface tilt. A checkerboard with 75° slant is depicted with tilts of 0°, 45° and 90°.

For an orthographic projection, the projected velocity of any point belonging to the rotating planar surface is proportional to the distance of the point from the plane parallel to the image plane and containing the axis of rotation of the surface—this distance is called *depth*. If the surface is slanted about the horizontal axis (Fig. 2, top left panel), then all points along the same horizontal direction in the image plane have the same depth. As a consequence, they project the same 2D velocity (e.g., points 1 and 2 in the top left panel of Fig. 2). The points along the vertical direction, conversely, have different depths and, therefore, project different 2D velocities (e.g., points 1 and 3 in the top left panel of Fig. 2). In these conditions, the horizontal velocity gradient Φ_x is nil whereas the vertical gradient Φ_y is different from zero. The opposite happens if the surface is slanted about the vertical axis (Fig. 2, top central panel). In this case, all points along the same vertical direction in the

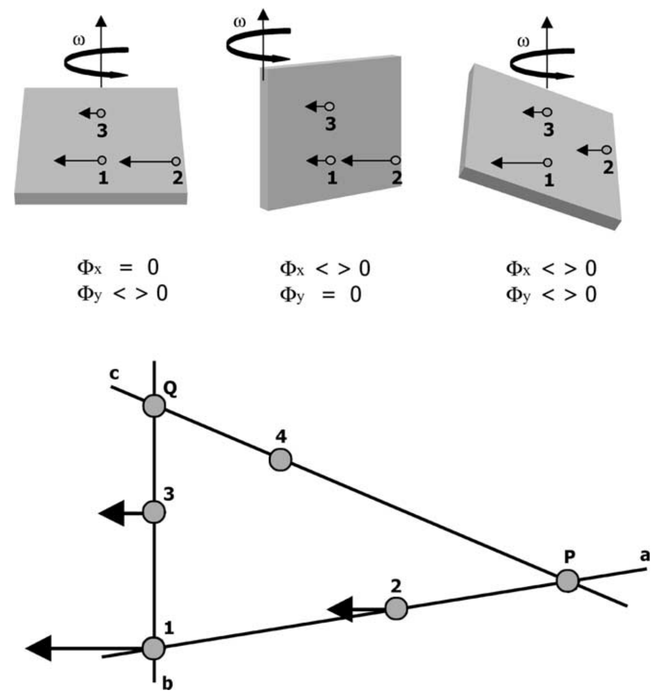

Fig. 2. *Top panel*. The planar surfaces depicted in the left, central and right panels are slanted about the horizontal axis ($\tau = 90°$), about the vertical axis ($\tau = 0°$) and about an oblique axis ($\tau = 45°$). ω represents the angular rotation about the vertical axis. Φ_x and Φ_y represent the horizontal and vertical velocity gradients, respectively. See text for details. *Bottom panel*. Schematic representation of the velocity field produced by the rotation of the surface in the top right panel. The arrows indicate the size of the velocity vectors associated with the points 1, 2, and 3. a is the straight lines connecting the points 1 and 2; b is the straight lines connecting the points 1 and 3; c is the straight line passing through the point 4 and intersecting a and b at points P and Q. The velocity at any arbitrary point 4 can be recovered from knowledge of the velocities at any other three non-collinear points 1, 2 and 3 (see text for details).

image plane have the same depth and, therefore, project the same 2D velocity (e.g., points 1 and 3). The points along the horizontal direction, conversely, have different depths and, therefore, project different 2D velocities (e.g., points 1 and 2). In these conditions, the vertical velocity gradient Φ_y is nil whereas the horizontal gradient Φ_x is different from zero. In the right panel of Fig. 2, finally, it is represented the general case in which the planar surface is slanted about both the vertical and the horizontal axes. In these circumstances, both the vertical and horizontal velocity gradients are different from zero. From the above considerations it follows that the relative intensities of the vertical and horizontal velocity gradients are related to the tilt of the projected planar surface. The surface tilt (τ), in fact, is equal to $\tau = \arctan(\Phi_y/\Phi_x)$.

The absolute values of the horizontal and vertical velocity gradients are related also to the slant of the planar surface (σ) and to its instantaneous 3D angular velocity (ω). The larger is the slant or the 3D angular velocity of the surface in the top left panel of Fig. 2, for example, the larger will be the difference between the 2D velocities of the points 1 and 3. In other words, as the slant or the 3D velocity of the surface of the top left panel of Fig. 2 grows, so it does the vertical gradient of the velocity field. Since for a generic tilt both the vertical and horizontal gradients are different from zero (see Fig. 2, top right panel), a global measure (called deformation) of the gradient intensities is adopted: $\mathrm{def} = \sqrt{\Phi_x^2 + \Phi_y^2}$. The relation between the deformation component of the velocity field, on the one hand, and surface slant (σ) and 3D angular velocity (ω), on the other, is: $\mathrm{def} = \sigma\omega$ (see Domini & Caudek, 1999).

Previous investigations about SFM have shown that perceived surface tilt (τ') is well predicted by $\arctan(\Phi_y/\Phi_x)$—in other words, τ' is strictly related to surface tilt (τ). Perceived surface slant (σ'), on the other hand, it is not (in general) well predicted by surface slant (σ), but it has been shown to be an increasing function of *def* and a decreasing function of $\arctan(\Phi_y/\Phi_x)$ (e.g., Domini & Caudek, 1999; Todd & Perotti, 1999; Liter & Braunstein, 1998; Freeman, Harris, & Meese, 1996).

3. Minimal conditions for specifying a linear velocity field

For the purposes of the present investigation, it is important to point out the minimal conditions for specifying the properties of a linear velocity field. The first thing to note is that, by knowing the 2D velocities of any two points, it is possible to compute all the 2D velocities along the line defined by those two points—the 2D velocity difference between any couple of points, in fact, is related to the 2D inter-point distance by a proportionality constant. As a consequence, it is easy to show that *the knowledge of the 2D velocities of three non-collinear points is both necessary and sufficient to fully specify the velocity field produced by the orthographic projection of a rotating planar surface*. Let us examine, for example, the top right panel of Fig. 2. By knowing the 2D velocities of the points 1 and 2, in fact, it is possible to determine all the 2D velocities along the line a passing through these two points (Fig. 2, bottom panel); by knowing the 2D velocities of the points 1 and 3 it is likewise possible to determine the 2D velocities along the line b. If now we want to know the 2D velocity of any other point (for example, point 4 in the bottom panel of Fig. 2), then we need only to choose a third line c passing through this new point and intersecting a and b. The 2D velocities of the two points P and Q defined by the intersections between the lines c and a, and c and b, respectively, will be known (P belongs to a, in fact, and Q belongs to b) and, therefore, the 2D velocity of all points belonging to c will also be determined. In this manner, the whole velocity field can be fully specified.

4. Temporal integration in structure from motion

Having described how the relevant parameters of the optic flow relate to those of perceived 3D structure, it is necessary to point out that there are different forms of temporal integration in SFM. In the following, we will first distinguish between long- and short-term temporal integration in SFM. We will then point out that short-term temporal integration in SFM is characterized by different properties, depending on the number of motion signals that define the velocity field.

Long-term temporal integration is the process by which the 3D orientation of a surface representation is updated according to the slant and rotation magnitudes perceived in previous moments in time, and it is revealed by the fact that 3D SFM perceived at moment t_0 is affected by the properties exhibited by the optic flow several hundred milliseconds before the moment t_0 (e.g., Domini, Vuong, & Caudek, in press; Domini, Caudek, & Skirko, in press; Vuong, Domini, & Caudek, 2001). Temporal integration, however, plays an important role also in the early analysis of the optic flow, that is, when a 3D surface representation is initially derived from the velocity field. It is well established, in fact, that the perceptual analysis of the optic flow is not instantaneous, but rather it is performed over an extensive temporal window. The process by which motion signals that are not simultaneously present at any moment in time are accrued within a small temporal window in order to specify the velocity gradients from which 3D information can be recovered will be termed *short-term temporal integration* in SFM. Short-term temporal integration has been revealed by using SFM displays with limited dot-lifetime and by showing that performance improves over several dot-lifetimes. Treue, Husain, and Andersen (1991), for example, asked observers to discriminate between structured (cylinder) and unstructured (noise) random-dot displays with a small number of dots with limited dot-lifetime. They found that perceptual performance improved with the increase of stimulus duration, thus suggesting that information is integrated over several dot-lifetimes.

One particular aspect of short-term temporal integration in SFM has to do with the possibility that the visual system may be able to recover the higher-order properties of a velocity field (such as *def*, *div* or *curl*—see Koenderink, 1986) when different samples of a velocity field are presented in rapid succession, and when each sample contains a number of velocity signals that is insufficient to define these properties. We can therefore distinguish between two forms of short-term temporal integration, depending on the properties of the optic-flow samples that are integrated over time.

In one case, each sample is made up of a number of motion signals by itself sufficient to define the relevant higher-order properties used by the visual system to recover surface slant and tilt (that is, the horizontal and vertical velocity-gradients are defined within each sample). In these circumstances, the integration process serves the purpose of obtaining more robust and precise measurements of the gradients of the flow fields. For this reason, this form of temporal integration will be called *short-term consolidation*.

In a second case, each sample is made up of a number of motion signals that is too small to define those optic-flow properties that are necessary to recover surface orientation and motion. In these circumstances, the integration process serves the purpose of *defining*, in the course of time, those properties of the velocity field that are not specified instantaneously, but are necessary to recover 3D information from motion. This form of temporal integration will be termed *short-term recruitment*.

One example of short-term recruitment is provided by Treue et al. (1991). In one condition of their experiments, observers were able to discriminate between structured and unstructured displays *having only two dots* in each frame of the apparent-motion sequence. According to Treue et al., "this is to be expected [...], provided the viewing time is long enough, the spatial sampling will be sufficiently dense (due to temporal integration) to compute a surface representation" (p. 73).

To our knowledge, the study of Treue et al. (1991) is the only investigation on short-term recruitment in SFM. Because of the particular properties of the displays that had been used, however, this experiment does not provide conclusive evidence that perceived SFM is affected by short-term recruitment. Even if *the number of motion signals in each frame-transition was not sufficient for solving the SFM task*, the observers of the experiment of Treue et al. (1991) could have judged the spatial distribution of the local 2D velocity vectors without recovering information about 3D structure. The structured displays of Treue et al., in fact, simulated the y-axis rotation of a cylinder oriented with its axis parallel to the axis of rotation. As a consequence, the 2D velocities of the structured displays (but not of the unstructured displays) were bigger in the center of the displays than in the periphery. If observers had simply compared the local 2D-velocities in different spatial positions, then the results of this experiment could not be taken as evidence that 3D information can be recovered, because of short-term recruitment, from *only two motion signals in each frame transition* of a stimulus display. Since temporal integration plays an important role in perceived SFM, the limitations of this previous study warrant therefore further investigation on the phenomenon of short-term recruitment. In particular, we can ask whether 3D information can be recovered from a sequence of optic-flow samples providing only one motion signal at the time, and what are the temporal properties of such a short-term recruitment.

The extreme case of short-term recruitment is schematically shown in Fig. 3. Consider a stimulus display in which a velocity-field is sampled so that only one velocity vector is shown at each moment in time. The size of the temporal-integration window for short-term recruitment is indicated by Δt. In the left panel, three velocity signals are contained within the temporal-integration window. The hypothesis that we put forward is that the higher-order properties of the optic-flow can be computed by using all the motion signals that are contained within the integration window, *even if they are presented at different moments in time*. Being this the case, the conditions represented by the left panel of Fig. 3 would suffice for specifying surface orientation: three velocity vectors, in fact, provide the minimal conditions for computing tilt and *def* (the reasons for this are

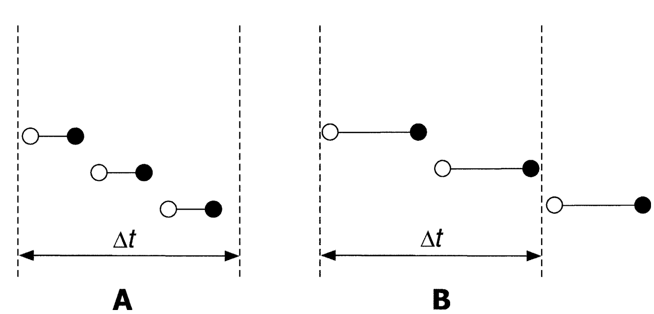

Fig. 3. Schematic representation of the position of three moving dots shown in succession. The initial and final position of each dot are represented by open and black circles, respectively. In the panel *A* dot-lifetime is smaller than in panel *B*, where the presentation of the motion of each dot in succession exceeds the dimensions of the temporal-integration window Δt. See text for details.

explained in the previous section). In the right panel, conversely, the temporal-integration window contains only two velocity vectors. Since the minimal conditions necessary for computing tilt and *def* are not met, in these conditions we propose that surface orientation is not perceptually specified. The hypothesis represented schematically in Fig. 3, therefore, implies the following prediction. If the short-term integration window has a fixed size, then *perceptual performance will be negatively affected by the increase of dot-lifetime in all SFM task in which the number of dots in each frame transition is not sufficient to specify surface orientation.*

In the following experiments, the prediction stated above was tested by examining the minimal conditions for short-term temporal integration in SFM. Observers were asked to perform a surface-orientation discrimination task when the number of dots, dot-lifetime and stimulus duration were manipulated. The stimulus displays were generated by first simulating a linear velocity field with a constant *def* and a variable tilt (+45° or −45°), and then by sampling a small number of motion signals from that velocity field. Observers were asked to discriminate between the two simulated surface orientations (+45° or −45° tilt). Because of the ambiguities of orthogonal projections, appropriate procedures were followed in the codification of the observers' responses, so as to take depth-reversals into account (see Method section).

It is important to note that, with the present experimental manipulations, we controlled the artifactual 2D cues described above. In the present experiments, in fact, the spatial distribution of 2D velocities remains constant across displays with different dot-lifetimes. A negative effect of dot-lifetime on perceptual performance, therefore, cannot be attributed to a *local* analysis of the 2D velocities in different locations of the image plane.

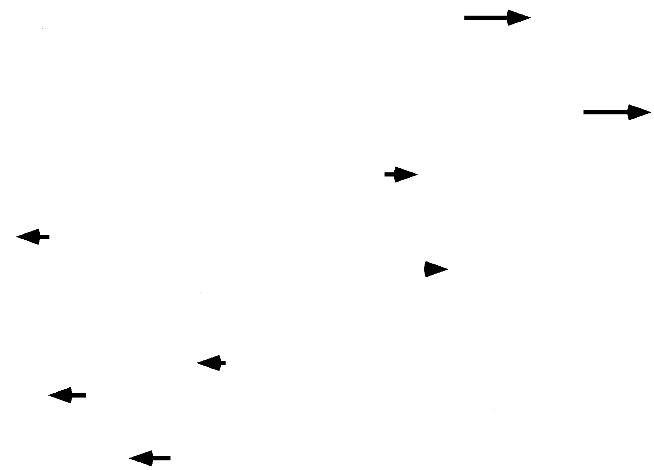

Fig. 4. Schematic representation of what the stimuli would have looked like on the screen if all the motion signals of the stimulus sequence had been presented simultaneously. The number of motion signals of the figure corresponds to that used in one of the experimental conditions of Experiment 1. The lengths of the arrows represent the velocities of the dots. Notice, however, that at any moment in time observers were shown only two (Experiments 1 and 2) or one (Experiment 3) of such motion signals.

5. General methods

5.1. Apparatus

The displays were presented on a high-resolution color monitor controlled by a Silicon Graphics Indigo II Impact Workstation. The 19″ screen had a resolution of 1280 × 1024 pixels, a refresh rate of 72 Hz, and was approximately photometrically linearized.

5.2. Stimuli

The stimuli consisted of random-dots displays. We set the dots to the maximal electron-gun value; a homogenous regions of that value had luminance of 82 cd/m^2. The luminance of the background was 3 cd/m^2. When drawing the dots, an anti-aliasing procedure was used: for locations falling on a pixel boundary, the pixel luminance was adjusted to an intermediate level of gray (256 levels) in proportion to the relative area falling on it.

The motion of the dots defined a linear velocity field with all velocity vectors parallel to the horizontal axis, as our stimuli only depicted surfaces rotating around the vertical axis (see Fig. 4). In such displays, the velocity of each image element remains constant for the whole stimulus duration and is determined uniquely by its 2D location (see Perotti, Todd, Lappin, & Phillips, 1998; Perotti, Todd, & Norman, 1996). A constant optic-flow field can be thought as the projection of a surface with an enormous constant slant and undergoing a minuscule amount of rotation. This is not, however, what the observers see. Observers report seeing a surface having a slant which can be very small or very large and undergoing a rotation of variable magnitude (very small or very large), with perceived slant and angular rotation magnitudes depending on the amount of simulated *def* (Domini & Caudek, 1999; Todd & Perotti, 1999).

Two tilt magnitudes were simulated: +45° or −45°. Since linear velocity fields were simulated, simulated tilt was constant in each trial. The *def* component took on the magnitude of 0.8 rad/s for all stimulus displays. The dots lived for a finite number of frames, and appeared and disappeared asynchronously (i.e., not all of the dots reached the end of their lifetime simultaneously). When a random dot disappeared, a new one was generated at a random position. Each dot had the size of one pixel.

Each stimulus display was contained within a circular "window" with a diameter of 5.70° visual angle (420 pixels). The dots were randomly distributed with uniform probability density over the projection plane.

5.3. Participants

The first author, one graduate student (M.D.L.) and three naive Trieste University undergraduate students participated in the experiments. All observers had normal or corrected to normal vision.

5.4. Procedure

Participants were instructed that they will be viewing random-dot kinematograms and that the dots would appear to form a surface moving in 3D space. Observers were shown examples of stimuli like those used in the experiments, but with each display defined by a larger number of dots. Observers were told that, in each trial, their task was to determine whether the moving dots appeared to form a surface with +45° tilt or −45° tilt. Because of the ambiguity of orthographic projections, prior to the experiment observers were carefully trained to follow the procedure described below. If the surface appeared to be oriented as a "floor" (i.e., with its lower portion on the x–y plane closer to the observer), then the left mouse-button connected to the workstation would signal the response "perceived tilt = +45°" (whereas the right button signaled the −45° response). If the surface appeared to be oriented as a "ceiling" (i.e., with its top portion on the x–y plane closer to the observer), then the right button would signal the response "perceived tilt = +45°" (whereas the left button signaled the −45° response).

Prior to the experimental sessions, observers were asked to complete several blocks of trials so as to familiarize them with the task and the stimulus displays. After training was completed, observers participated in a variable number of experimental sessions over several weeks.

Viewing was monocular. Head and eye motions were not restricted. The experimental room was dark during the experiment. The eye-to-screen distance was approximately 0.7 m. Feedback was provided on each trial in the form of a beep for incorrect judgements.

6. Experiment 1

The purpose of Experiment 1 was to determine whether it is possible to perform a surface-orientation discrimination task when the number of dots in each frame transition is not sufficient to specify surface orientation. In each frame transition, the stimulus displays provided only two moving dots, and dot-lifetime and display duration were manipulated. We formulated the following predictions. (1) The level of performance decreases with increasing dot-lifetime. According to our hypothesis, in fact, the increase of dot-lifetime would reduce the number of motion signals that are contained in the temporal-integration window for short-term recruitment. In turn, this would have a detrimental effect on performance, until the limiting case in which the short-term temporal-integration window contains less than three velocity vectors and, therefore, tilt and *def* can no longer be computed. (2) The level of performance increases with the increase of display duration. This effect has already been reported (e.g., Treue et al., 1991; van Damme & van de Grind, 1996), and represents a form of *consolidation* after the *recruitment* phase.

6.1. Method

6.1.1. Participants

Three volunteers, recruited from the Trieste University community and naive to the purpose of the experiment, and one of the authors (C.C.) participated in this experiment.

6.1.2. Stimuli

Constant velocity fields with tilt of +45° and −45° were simulated. Dot-lifetime took on the values of 10 (0.140 s) or 60 (0.840 s) frames. Each stimulus sequences was made up of 40, 80, 160 or 240 frames. Each frame of the stimulus sequence contained only two dots.

6.1.3. Design

Two within-participants variables were studied: dot-lifetime (10 or 60 frames), and number of frames of the stimulus sequence (40, 80, 160 or 240). Each of these variables was blocked. Within each block, the sign of the velocity field (indicating clockwise or counterclockwise rotation) and the simulated tilt magnitude (+45° and −45°) were completely randomized. Each observer completed an average of 275 trials for each combination of dot-lifetime and stimulus-sequence length.

6.1.4. Results

The proportions of correct judgments for each stimulus-sequence length and for each dot-lifetime are shown in Fig. 5 for each observer. A logit model was fitted to the data of each observer, with the dichotomy of correct/incorrect response as the dependent variable, and dot-lifetime and stimulus-sequence length as the two predictors. The fitted logit models for each observer, deleting the interaction, are given in Table 1, where the ratio of each coefficient to its standard error is shown in parenthesis. These analyses show that dot-lifetime has a greater effect than stimulus-duration on the log odds of a correct response ($Y = 1$), even if the range of both independent variables (dot-lifetime and stimulus-duration) was equated. The most important result, however, is that the dot-lifetime coefficient is negative: as expected, perceptual performance decreases if dot-lifetime increases.

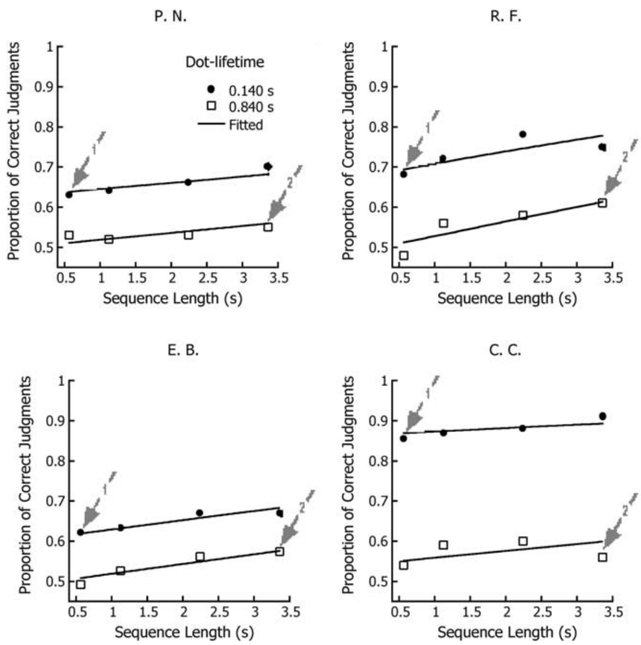

Fig. 5. Proportions of correct judgments for four observers as a function of sequence length and dot-lifetime in Experiment 1. The data point indicated by the gray arrow with the number 1 represents the experimental condition defined by a dot-lifetime of 10 frames (0.140 s) and the shortest sequence-length; the data point indicated by the gray arrow with the number 2 represents the experimental condition defined by a dot-lifetime of 60 frames (0.840 s) and the longest sequence-length. Notice that, for all observers, a better performance was observed in the case of the shortest dot-lifetime (gray arrow with the number 1), even if these two experimental conditions provided the same number of spatial samples of the velocity field.

Table 1
Experiment 1. Fitted logit models for each observer. The ratio of each coefficient to its standard error is shown in parenthesis

C.C.	$\Lambda^{-1}(P_c) = 2.209 + 0.070D - 2.481\mathrm{DL}$	$\chi^2_5 = 4.908$,
	(1.430) (−15.151)	$p = 0.427$
E.B.	$\Lambda^{-1}(P_c) = 0.523 + 0.102D - 0.668\mathrm{DL}$	$\chi^2_5 = 0.874$,
	(2.480) (−5.284)	$p = 0.972$
R.I.	$\Lambda^{-1}(P_c) = 0.893 + 0.155D - 1.140\mathrm{DL}$	$\chi^2_5 = 5.400$,
	(3.646) (−8.731)	$p = 0.369$
P.L.	$\Lambda^{-1}(P_c) = 0.633 + 0.070D - 0.764\mathrm{DL}$	$\chi^2_5 = 1.140$,
	(1.636) (−5.809)	$p = 0.950$

P_c is the odd of correct response, D is the stimulus duration, DL is the dot-lifetime.

It is important to compare the condition defined by a dot-lifetime of 10 frames and the shortest sequence-length (indicated in Fig. 5 by means of the gray arrow with the number 1) and the condition defined by a dot-lifetime of 60 frames and the longest sequence-length (indicated in Fig. 5 by means of the gray arrow with the number 2), since the information-content of these displays is identical. Both cases, in fact, provide *the same number of spatial samples of the velocity field*. Even if the information-content was equated, however, a better performance was observed for all observers in the case of the displays having the shortest dot-lifetime.

A second result of the logit analysis is the significant effect of the stimulus-sequence length (see Fig. 5). This result indicates that performance improves over several dot-lifetimes, and replicate the similar results reported by Treue et al. (1991) and van Damme and van de Grind (1996).

7. Experiment 2

In previous SFM studies, the increase of dot-lifetime has been found to have a positive effect on performance (e.g., van Damme & van de Grind, 1996). This finding is the opposite of what we have found in Experiment 1 and the difference between these results requires an explanation. To explain this difference, we propose that the increase of dot-lifetime has been found to have a beneficial effect on performance because, in those previous investigations, short-term recruitment was not necessary. We propose, in fact, that short-term temporal recruitment occurs only when the number of motion signals provided in each frame transition is *insufficient* to define surface orientation. In the presence of insufficient information in each frame transition, the decrease of dot-lifetime is beneficial because it increases the number of motion signals that are accrued within the short-term temporal-integration window. If the number of motion signals provided by each frame transition is *sufficient* to define surface orientation, on the other hand, short-term recruitment is not necessary. In these circumstances, performance benefits from the increase of dot-lifetime, since longer lifetimes allow a better measurement of the local motion signals (e.g., McKee & Welch, 1985). As a consequence, we expect that the effect of dot-lifetime will interact with the effect of the number of dots of the stimulus display. We expect that, with two dots, the increase of dot-lifetime will have a detrimental effect on performance but, with only three dots, performance will benefit from the increase of dot-lifetime. Three velocity vectors, in fact, are sufficient to specify *def* and tilt and, in these circumstance, short-term recruitment is not necessary.

7.1. Method

7.1.1. Participants

Two volunteers, recruited from the Trieste University community and naive to the purpose of the experiment, and one of the authors (C.C.) participated in this experiment.

7.1.2. Stimuli

The stimuli were similar to those of Experiment 1, except that the number of frames of each stimulus sequence was kept constant (80 frames), and dot-lifetime took on the values of 2, 4, 8, 12, 16, 32 and 64 frames. In each stimulus display, either two or three dots were shown.

7.1.3. Design

Two within-participants variables were studied: dot-lifetime (2, 4, 8, 12, 16, 32 or 64 frames), and number of dots (2 or 3). Each of these two variables was blocked. Within each block, the sign of the velocity field and the simulated tilt magnitude (+45° and −45°) were completely randomized. Each observer completed 240 trials in each combination of dot-lifetime and number of dots.

7.1.4. Results

The average proportions of correct judgments for each dot-lifetime and number of dots are shown in Fig. 6. As predicted, a different pattern of data is observed for displays with 2 and 3 dots. With 2 dots, performance peaked, on average, with dot-lifetimes of 8–12 frames, and then decreased. This result is consistent with the results of Experiment 1. With 3 dots, instead, performance increases until the 12-frame lifetime, and then it reached a plateau. The dot-lifetime corresponding to plateau performance with 3 dots in the present experiment (about 170 ms) is consistent with the values reported previously: 130 ms in the experiment of Treue et al. (1991), and 180–200 ms in the experiment of van Damme and van de Grind (1996).

A repeated measures analysis of variance (ANOVA) revealed that the interaction between the variable dot-lifetime and the variable number of points was significant [$F(6,12) = 6.522, p < 0.01, \eta^2 = 0.765$], as expected. Significant, but marginal to the interaction, were also the variables dot-lifetime [$F(6,12) = 15.286, p < 0.001, \eta^2 = 0.884$] and number of dots [$F(1,2) = 39.942, p < 0.05, \eta^2 = 0.951$].

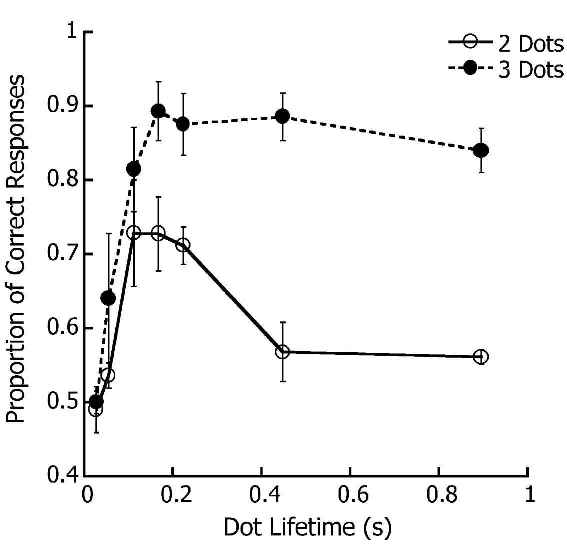

Fig. 6. Proportions of correct judgments averaged across observers as a function dot-lifetime and number of dots in Experiment 2.

The fact that poor performance (with 2 and 3 dots) was observed with dot lifetimes of 2 and 4 frames (27 and 55 ms) is consistent with previous reports indicating a temporal requirement of 80–100 ms for both accurate velocity-discriminations (e.g., McKee & Welch, 1985) and accurate performance in SFM tasks (Treue et al., 1991). In this regard, however, the individual differences are interesting to note. The most experienced observer (C.C.) was able to perform 80% correct with 55 ms dot-lifetime in the case of 3 dots whereas, with 2 dots, his performance was at chance (52% of correct responses). The other not-experienced observers, conversely, did not show any meaningful pattern of results for the two shortest lifetimes.

8. Experiment 3

Experiment 3 had a twofold purpose. First, to obtain a direct measurement of the size of the temporal-integration window for short-term recruitment. Second, to test the hypothesis that the minimal stimulus-conditions described in Fig. 3 suffice for short-term recruitment in perceived SFM.

8.1. Method

8.1.1. Participants

One volunteer, recruited from the Trieste University community and naive to the purpose of the experiment, and one of the authors (C.C.) participated in this experiment.

8.1.2. Stimuli

The stimuli were similar to those of Experiment 1, except that only one dot was shown in each frame of the stimulus display. The number of frames of each stimulus sequence was equal to 3 times the dot-lifetime (for a dot-lifetime of 8 frames, for example, the stimulus sequence was made up of 24 frames). Dot-lifetime took on the values of 8, 10, 11, 12, 13, 14 and 20 frames for observer M.D.L., and 8, 9, 10, 11, 12, 15, 20 for observer C.C.

8.1.3. Procedure

Since no restriction was placed on the generation of the stimulus displays, given the difficulty of the task, observers were allowed to skip those trials in which the random placement of the dots on the screen did not support the perception of a 3D surface. The program cycled through until a sufficient number of response was provided in each condition by each observer. Otherwise, the procedure was the same as in the previous experiments.

8.1.4. Design

The within-participants variables dot-lifetime was blocked. Within each block of trials, the sign of the velocity filed and the simulated tilt magnitude (+45° and −45°) were completely randomized. Each observer completed an average of 180 trials for each dot-lifetime.

8.1.5. Results

The proportions of correct judgments for each dot-lifetime are shown in Fig. 7. The asterisks in the figure indicate the data points in which performance is significantly above chance at the 95% confidence level. The first thing to note is that at least a crude representation of surface orientation can be obtained even in these minimal conditions: three dots presented sequentially, one at the time. For the optimal dot-lifetime, an average percentage of 65% of correct responses was observed. This result can be taken as the most direct demonstration of short-term recruitment in perceived SFM. It should be noticed, however, that the level of 65% represents a sort of upper limit for performance in the present stimulus conditions. Observers were allowed, in fact, to skip the trials which did not support a clear 3D percept, and these trials were then generated again with a different random placement of the dots on the screen (see Method section). Because observers would have presumably performed at guess rate on those trials, if these trials had been counted in the figure above (65%), then the level of performance would have been reduced considerably. The second thing to note is that performance is above chance for dot-lifetimes up to 168 ms (observer C.C.) and 182 ms (observer M.D.L.). Accordingly, we can conclude that the size of the temporal window for short-term recruitment can measure up to 550 ms.

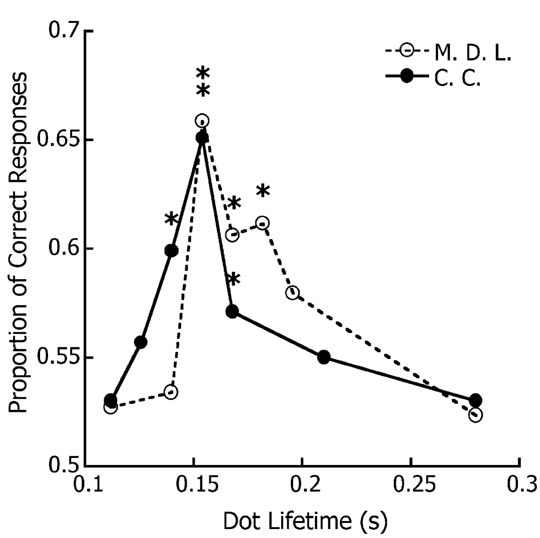

Fig. 7. Proportions of correct judgments for two observers as a function dot-lifetime in Experiment 3. The asterisks indicate the data points in which performance is significantly above chance at the 95% confidence level.

9. General discussion

The results of the present study provide evidence for the role of short-term recruitment in SFM and supports the claim of Treue et al. (1991) that perceived 3D shape can be defined by motion signals that are by themselves insufficient, in each frame transition, for solving the SFM problem. In Experiment 1, we found that observers were able to discriminate between 3D surface orientations also when the apparent-motion sequences provided only 2 dots in each frame. Consistently with our proposal of a limited-size temporal-integration window for short-term recruitment, moreover, a better performance was observed with shorter dot-lifetimes, even when the information-content of displays with different dot-lifetimes (140 vs. 840 ms) was equated. In Experiment 2, we found that dot-lifetime had a different effect on performance depending on the number of dots. For both displays made up of 2 or 3 dots, an increase of dot-lifetime up to 140–170 ms was beneficial for performance. At that point, however, the effect of dot-lifetime on performance became dependent on the necessity of short-term recruitment for 3D shape recovery. An increase of dot-lifetime was detrimental for performance if the displays were made up of 2 dots (short-term recruitment being necessary). Performance reached a plateau as dot-lifetime increased, conversely, if the displays were made up of 3 dots (short-term recruitment not being necessary). In Experiment 3, we estimated the size of the temporalintegration window for short-term recruitment by showing one single dot in each frame transition. As it can be seen in Fig. 6, performance decreased for dot-lifetimes longer than 154 ms, even though performance was still significantly above chance when dot-lifetime was equal to 182 ms. This result suggests, therefore, that short-term recruitment in perceived SFM occurs within a temporal window that can be as long as 550 ms. The results of this experiment indicate, furthermore, that the process of short-term temporal integration is counterbalanced by the need of obtaining accurate measurements of the velocity field. As in Experiment 2, in fact, we found that performance was not above chance when dot-lifetime was shorter than 55 ms.

It is interesting to relate the present findings on short-term recruitment in SFM to those relative to temporal integration of 2D motion signals—when there is no need of recovering 3D information. An effect similar to the one described here had been reported, for example, by van Doorn, Koenderink, and van de Grind (1985). They showed that motion detection increased as the number of separate motion impulses presented per unit time increased. This is to be expected according to any model of temporal summation of motion. Nakayama and Silverman (1984), moreover, presented evidences for a fast and a slow form of motion integration. The fast process was revealed by the fact that integration declined if the delays between impulses were longer than 40 ms. The slow process was revealed by the fact that some form of integration still occured even over delays longer than 300 ms. Similar data had been reported by Regan and Beverley (1984), indicating that the longer integration interval lasted for about 1 s. Two stages of analysis had been found also by measuring the contrast and coherence thresholds for direction discrimination in optic-flow stimuli. Burr and Santoro (2001) asked observers to discriminate upward vs. downward translations, clockwise vs. counterclockwise rotation and expanding vs. contracting radial optic-flow patterns. They found evidence for 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 3000 ms. Whereas the methodology used in the present investigation allows one to rule out the hypothesis that observers performed the SFM task at the level of 2D velocities (i.e., without recovering 3D structure), the finding of (at least) two stages of analysis—short- and long-term processing—for both the measurement of 2D velocities and the recovery of 3D information from the optic flow represents an interesting empirical result that warrants further future research.

Acknowledgements

This research was supported by National Science Foundation grant BCS-78441. We would like to thank Randolph Blake and two anonymous reviewers for their valuable comments on a previous version of this manuscript.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, 2*, 284–299.

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

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

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

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

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., & Skirko, P. (in press). Temporal integration of motion and stereo cues to depth. *Perception and Psychophysics*.

Domini, F., Vuong, Q., & Caudek, C. (in press). Temporal integration in structure from motion. *Journal of Experimental Psychology: Human Preception and Performance*.

Festa, E. K., & Welch, L. (1997). Recruitment mechanisms in speed and fine-direction discrimination tasks. *Vision Research, 37*, 3129–3143.

Freeman, T. C., Harris, M. G., & Meese, T. S. (1996). On the relationship between deformation and perceived surface slant. *Vision Research, 36*, 317–322.

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

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.

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

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

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

Liter, J. C., & Braunstein, M. L. (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.

McKee, S. P., & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America, 2*, 243–251.

Nakayama, K., & Silverman, G. H. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research, 24*, 293–299.

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception & Psychophysics, 60*, 377–388.

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.

Regan, D., & Beverley, K. I. (1984). Figure-ground segregation by motion contrast and by luminance contrast. *Journal of the Optical Society of America, 1*, 433–442.

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research, 31*, 907–914.

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

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

van Doorn, A. J., & Koenderink, J. J. (1983). Detectability of velocity gradients in moving random-dot patterns. *Vision Research, 23*, 799–804.

van Doorn, A. J., Koenderink, J. J., & van de Grind, W. A. (1985). Perception of movement and correlation in stroboscopically presented noise patterns. *Perception, 14*, 209–224.

van Santen, J. P., & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America, 1*, 451–473.

Vuong, Q., Domini, F., & Caudek, C. (2001). Temporal integration in structure from motion. Poster presented at the 1st Annual Meeting of the Vision Sciences Society, Sarasota, FL, May 2001.

Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature, 333*, 71–74.