PERGAMON

Vision Research 41 (2001) 2715–2732

Vision Research

www.elsevier.com/locate/visres

Segmentation in structure from motion: modeling and psychophysics

Corrado Caudek *, Nava Rubin

Center for Neural Science, New York University, 4 Washington Pl., New York, NY 10003, USA

Received 9 October 2000; received in revised form 23 April 2001

Abstract

Much work has been done on the question of how the visual system extracts the three-dimensional (3D) structure and motion of an object from two-dimensional (2D) motion information, a problem known as '*Structure from Motion*', or SFM. Much less is known, however, about the human ability to recover structure and motion when the optic flow field arises from multiple objects, although observations of this ability date as early as Ullman's well-known two-cylinders stimulus [The interpretation of visual motion (1979)]. In the presence of multiple objects, the SFM problem is further aggravated by the need to solve the segmentation problem, i.e. deciding which motion signal belongs to which object. Here, we present a model for how the human visual system solves the combined SFM and segmentation problems, which we term SSFM, concurrently. The model is based on computation of a simple scalar property of the optic flow field known as *def*, which was previously shown to be used by human observers in SFM. The *def* values of many triplets of moving dots are computed, and the identification of multiple objects the image is based on detecting multiple peaks in the histogram of *def* values. In five experiments, we show that human SSFM performance is consistent with the predictions of the model. We compare the predictions of our model to those of other theoretical approaches, in particular those that use a rigidity hypothesis, and discuss the validity of each approach as a model for human SSFM.

Keywords: Perceptual segmentation; Structure; Motion; Structure from motion

1. Introduction

The importance of motion information, or optic flow, for the perception of the three-dimensional (3D) layout and structure of surfaces and objects has been known for many years (cf. Gibson, 1950). Miles (1931) and, later, Wallach & O'Connell (1953) showed that a vivid perception of 3D objects undergoing rotational motion can be obtained from the two-dimensional (2D) projections of these moving objects on a flat screen. Wallach and O'Connell were careful to use unfamiliar objects (resembling randomly bent paperclips) to emphasize that the 3D structure of these objects could not be recovered from any of the static 2D images that comprised the moving stimulus (and indeed, observers do not report perceiving a 3D shape for the static images, unless they have seen the motion sequence from which they were taken before; Wallach, O'Connell, & Neisser, 1953; see also Sinha & Poggio, 1996). More recently, Braunstein (1962, 1976) and Ullman (1979) demonstrated that human observers are capable of extracting 3D structure information from moving images consisting only of dots rendered on the surface of an (invisible) volumetric shape, i.e. even when all other sources of shape information have been removed (e.g. shape outline/contours, texture density, disparity, shading). This phenomenon, termed *Structure from Motion* (SFM), has received much attention in experimental as well as theoretical and modeling work. Human SFM performance has been characterized quantitatively under many conditions, and much is known about the minimal condition needed to produce SFM (i.e. to

* Present address: Psychology Department, UniversityofTrieste, Via S. Anastasio 12, 34100 Trieste, Italy.
Direct correspondence to both authors. E-mail addresses: caudek@univ.trieste.it , nava@cns.nyu.edu

0042-6989/01/$ - see front matter
PII: S0042-6989(01)00163-8

create a perception of a moving 3D object from a 2D velocity field; see, for example, Lappin, Doner, & Kottas, 1980; Braunstein, Hoffman, Shapiro, & Andersen, 1987; Braunstein, Hoffman, & Pollick, 1990; Hildreth, Grzywacz, Adelson, & Inada, 1990; Todd & Bressan, 1990; Landy, Dosher, Sperling, & Perkins, 1991; Treue, Husain, & Andersen, 1991; Rubin, Hochstein, & Solomon, 1995a), the ability to discriminate between velocity fields generated by rigid and non-rigid motion (see, for example, Braunstein, 1976; Todd, 1982, 1984; Todd, Tittle, & Norman, 1995; Todd & Perotti, 1999; Lappin & Fuqua, 1983; Koenderink & van Doorn, 1986; Koenderink, Kappers, Todd, Norman, & Phillips, 1996; Norman & Todd, 1993; Perotti, Todd, & Norman, 1996; Sparrow & Stine, 1998), or between a single rigid rotation and visual noise (Turner, Braunstein & Andersen, 1995; Domini, Caudek, & Proffitt, 1997), and the relation between simulated and perceived slant (Loomis & Eby, 1988; Caudek & Proffitt, 1993; Domini & Caudek, 1999), tilt (Pollick, Nishida, Koike, & Kawato, 1994), orientation of the axis of rotation (Caudek & Domini, 1998), and angular velocity (Kaiser, 1990; Kaiser & Calderone, 1991; Domini, Caudek, Turner, & Favretto, 1998b). In addition, a rich body of theoretical work exists that relates these findings to models of human SFM (see, for example, Todd, 1982; Ullman, 1984; Grzywacz & Hildreth, 1987; Bennett, Hoffman, Nicola, & Prakash, 1989; Braunstein, 1989, 1994; Thompson, Kersten, & Knecht, 1992; Koenderink & van Doorn, 1991; Dijkstra, Snoeren, & Gielen, 1994; Eagle & Blake, 1995; Hildreth, Ando, Andersen, & Treue, 1995; Hogervorst, Kappers, & Koenderink, 1996).

The studies mentioned above, as well as most of the other work in SFM, have concentrated on recovering 3D structure and motion from an optic flow field that was produced by a single moving object (but see Adiv, 1985; Thompson et al., 1992; Liter, Braunstein, & Hoffman, 1994). Much less work has been done, however, on the question of whether, and how, human observers are able to recover 3D structure from overlapping velocity fields produced by *multiple* objects' 3D motion. In the presence of multiple objects, the SFM problem is further aggravated by the need to solve the problem of *which motion signal belongs to which of the objects in the scene*. In other words, in order to solve the SFM problem, the visual system also needs to solve (prior to it or concurrently) the problem of segmenting the image into the distinct ecological sources of motion in the scene. We term this problem *Segmentation and Structure from Motion*, or SSFM.

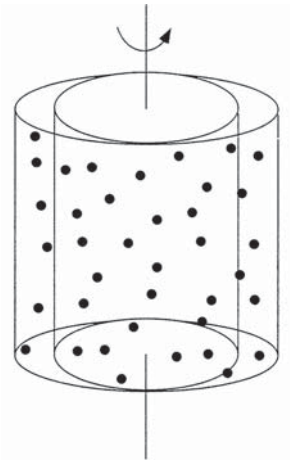

Fig. 1. Schematic representation of two coaxial random-dot cylinders used by Ullman (1979) to demonstrate the human ability to solve the combined segmentation and structure-from-motion (SSFM) problems.

Three questions can be posed about SSFM. First, do overlapping velocity fields indeed evoke the perception of multiple objects' 3D motion? Second, if the answer to the first question is positive, what is the computational approach that accounts for the perceptual outcomes? Third, what are the stimulus variables that affect the perception of SSFM, and how are they related to the underlying computations?

The first of these questions has been addressed by Ullman (1979). In his well-known two-cylinder display, two random-dot cylinders rotated with different velocities about their major axis (see illustration in Fig. 1). Ullman reported that, in these circumstances, 'each static view (…) appeared almost as a random collection of points. However, when the changing projection was viewed, the elements in motion across the screen were perceived as two rotating cylinders whose shapes and angles of rotation were easily determined' (pp. 134–135). Ullman's (1979) observations indicate that, at least under certain conditions, human observers are able to solve the SSFM problem, i.e. that overlapping velocity fields can evoke the impression of separate objects' 3D structure and motion, in the absence of any other shape or segmentation cues. More recently, Liter et al. (1994) examined the ability of observers to detect whether a SFM display depicted one or two rigid objects rotations and found that performance depended on the magnitude of the orientation difference between the two axes of rotation, and on the number of noise points added to the displays. In this study, we extend those findings to other random-dot motion displays.

The general computational approach at the basis of our model for human SSFM performance can be summarized as follows. Given an image made up of N dots (local motion signals), find a quantity, let us call it Q, which can be computed based on small subsets of p moving dots (e.g. 3 or 4), and whose value is constant (or approximately so) for all subsets that belong to a single, coherently moving object. Then, compute the value of Q (which, in principle, can be either a scalar or a vector) for all $\binom{N}{p}$ subsets of p dots in the image,

and examine the distribution of values. If the image were generated from the 3D motion of a single coherent object, we would expect the distribution of Q to contain a single cluster. If, in turn, the image were generated by two or more 3D objects, we would expect multiple clusters, as the number of multiple objects. In addition to these multiple clusters, the distribution will include many spurious Q-values, generated by subsets of p dots, some of which belong to one object and others to other objects. However, the values of Q generated by spurious dot clusters will generally have a flat (random) distribution. Therefore, if there are enough dots in the stimulus, the peaks in the distribution generated by the same-object dot clusters will be much higher than the values generated by the spurious clusters. Thus, by finding the number of clusters in the distribution of Q, one may find the number of underlying objects. Furthermore, by tracing back which subset of p dots generated which value of Q, one may label the dots as belonging to one object or the other. Note that this approach is a generalization of one that can be applied in many single-object SFM problems: by noting whether the distribution of Q contains a peak or not, one may discriminate, for example, between images that contain a coherent 3D moving object embedded in noise, and those that contain noise alone.

In its general form described above, our computational approach is similar to that put forward by Ullman (1979). However, we will propose a model that differs markedly in terms of the choice of the quantity Q and the predictions derived from that choice. Ullman (1979) proposed to adapt his single-object SFM algorithm, which can determine for any four points whose positions are known for three consecutive frames whether the quadruplets of dots have a unique interpretation as an underlying rigid body moving in 3D space. If such an interpretation exists, the algorithm recovers the rotation parameters (axis and angular speed) of the underlying object, as well as its 3D structure. If the motion of the N dots in the image arises from the rigid motion of two (or more) differently moving objects, the distribution of axes and angular rotations recovered for each of the $\binom{N}{4}$ quadruplets will contain two (or more) clusters at the values corresponding to the motion of each of the objects. The quantity Q suggested by Ullman is therefore the vector containing the axis of rotation and speed of quadruplets (i.e. $p = 4$). This has several implications for the computational complexity of the algorithm and the predictions it makes as a model of human SSFM performance. First, note that the algorithm makes use of a rigidity assumption to constrain the solution. Second, at least three views are needed to apply the algorithm, and the computation of the motion parameters requires the solution of transcendental equations (Ullman, 1979). Third, the peaks in the distribution need to be searched in a four-dimensional space (three for the axis of rotation, one for the speed). These characteristics make Ullman's (1979) algorithm computationally demanding, and there is evidence that human performance in SFM is not compatible with those computational demands. With regard to the rigidity assumption, several reports indicate that human observers can be strikingly poor in discriminating rigid from nonrigid motion (Caudek & Proffitt, 1993; Perotti et al., 1996; Domini et al., 1997; Hogervorst, Kappers, & Koenderink, 1997; Griffiths & Zaidi, 1998; Sparrow & Stine, 1998). Similarly, there is evidence that human SFM performance is based primarily on two-frame motion information and benefits only very little from higher than first-order temporal information, i.e. from more than two views (Lappin et al., 1980; Braunstein et al., 1990; Todd & Bressan, 1990; Liter, Braunstein, & Hoffman, 1993; Rubin et al., 1995a; Rubin, Solomon, & Hochstein, 1995b; Domini et al., 1997). These findings (as well as others, see below) therefore make it unlikely that the visual system makes use of an algorithm like that suggested by Ullman (1979) to solve the SSFM problem.

The model presented here overcomes the problems outlined above by suggesting a different quantity, Q, with which to implement the general approach described before. We propose that, in resolving the SSFM problem, the visual system relies heavily on the distribution of values of a scalar quantity called *def* (Koenderink, 1986). This quantity is related to the deformations in the local velocity field, and can be computed for triplets of dots (i.e. $p = 3$) from the first-order (instantaneous, or two-frame) optic flow field. There is much evidence from previous studies that links the computation of *def* with human SFM performance. For example, Andersen (1996) studied observers' ability to discriminate SFM displays depicting corrugated surfaces from displays of points randomly positioned within a 3D volume. He found that an analysis of the deformations of the velocity field could predict the effect on performance of a variety of manipulations (smoothness of the velocity field, frequency and amplitude of corrugation, dot density). In contrast, approaches based on recovery of the Euclidean parameters for SFM do not predict that those manipulations should have an effect on performance (see also below, Experiments 2, 4–5). Subsequent work provided further evidence for the importance of the deformation components of the optic flow field in human SFM perception (Andersen & Atchley, 1997; Domini et al., 1997; Domini & Braunstein, 1998; Domini, Caudek, & Richman, 1998a; Domini et al., 1998b; Domini & Caudek, 1999; Atchley, Andersen, & Wuestefeld, 1998; Caudek & Domini, 1998). These studies demonstrated that a *def*-based analysis predicts human performance well when observers are asked to detect a single 3D surface (e.g. when it is embedded in noise). But the validity of the *def*-based approach when two overlapping surfaces are present remained unexamined. There are two

Note: eq (3) is corrected here; the pdf on the VR site has a typo

questions here: can the *def*-based approach be used to solve SSFM problems in principle, and, if so, is it used in practice, i.e. will a *def*-based SSFM model adequately predict human performance? Both questions are addressed here.

Importantly, the findings relating the *def*-based approach to single-surface perception capture cases not only when human observers succeed in SFM tasks, but also when they fail. This makes it attractive to apply *def* to our computational approach for solving the SSFM problem, i.e. to use it as the quantity Q. In a series of experiments, we present psychophysical evidence to support the predictions of the model about successes as well as failures of human SSFM and contrast them with the predictions of other models. It is well known that understanding where a system breaks down can tell us a lot about its underlying mechanisms (Gregory, 1998), and therefore experimental data about human failures in SSFM can provide strong support for our model (given that other models do not make similar predictions, see below). Nevertheless, before moving on to describe our model in detail, we should emphasize again that, in terms of the general theoretic approach, our model shares an important characteristic with that proposed by Ullman (1979). This is the fact that, by collecting information about the distribution of a quantity, Q, computed based on many samples of a small numbers of dots, p, the segmentation problem is solved *concurrently* with recovering 3D structure and motion information from the optic flow field. This is to be contrasted with approaches that attempted to compute a segmentation of the scene before computing SFM (e.g. Tomasi & Kanade, 1992).

1.1. Segmentation in SSFM based on the distribution of def

Given a two-dimensional velocity field, $\mathbf{V}(x,y) = \{V_x(x,y), V_y(x,y)\}$, the gradient matrix, $\mathbf{\Gamma}$, is defined as

$$\Gamma = \begin{bmatrix} \frac{\partial V_x}{\partial x} & \frac{\partial V_x}{\partial y} \\ \frac{\partial V_y}{\partial x} & \frac{\partial V_y}{\partial y} \end{bmatrix} \tag{1}$$

$\mathbf{\Gamma}$ can be decomposed into a sum of four constant matrices weighted by variable scalar coefficients (see Koenderink, 1986; Todorovic, 1993)[1]:

$$\Gamma = \frac{1}{2}\sum_{k=1}^{4} C_k \mathbf{M}_k \tag{2}$$

where

$$C_1 = C_{div} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}, \qquad \mathbf{M}_1 = \mathbf{M}_{div} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3}$$

$$C_2 = C_{curl} = \frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}, \qquad \mathbf{M}_2 = \mathbf{M}_{curl} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$C_3 = C_{def_1} = \frac{\partial V_x}{\partial x} - \frac{\partial V_y}{\partial y}, \qquad \mathbf{M}_3 = \mathbf{M}_{def_1} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$C_4 = C_{def_2} = \frac{\partial V_y}{\partial x} + \frac{\partial V_x}{\partial y}, \qquad \mathbf{M}_4 = \mathbf{M}_{def_2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Intuitively, this means that the velocity field can be described at each point as the weighted sum of four flow fields with distinct forms of spatial variation: in/outflow (‘*div*’), pure rotation (‘*curl*’), and two forms of shearing motion (‘def_1’ and ‘def_2’). This decomposition omits a possible global translation of the flow field. The coefficients $C_1 \ldots C_4$ are a function of (x, y), i.e. need to be computed locally; the computation of spatial derivatives is quite simple to implement in a neural network (cf. Marr, 1982; Koch & Segev, 1998), giving models that use their values biological plausibility.

The quantity *def* is defined as the following combination of the two shear coefficients (Koenderink, 1986; for an intuitive interpretation of *def*, see Domini & Caudek, 1999):

$$def = \sqrt{C_{def_1}^2 + C_{def_2}^2} \tag{4}$$

In the discrete case, when the instantaneous velocity field is approximated by the succession of two static frames, *def* can be computed for each triplet of points, and can be shown from Eqs. (3) and (4) to have the following form (Domini et al., 1997):

$$def = \frac{1}{|\sin\alpha|}\sqrt{\left(\frac{|V_1 - V_0|}{\rho_1}\right)^2 + \left(\frac{|V_2 - V_0|}{\rho_2}\right)^2 - 2\,|V_1 - V_0|\rho_1 \frac{|V_2 - V_0|}{\rho_2}\cos(\alpha - \alpha_\Delta)} \tag{5}$$

where $\{V_0, V_1, \text{and } V_2\}$ are the velocity vectors of the points $\{P_0, P_1, \text{and } P_2\}$, ρ_1 and ρ_2 are the distances of the points P_1, P_2 from the point P_0, α is the angle between the line segments P_0P_1 and P_0P_2, and α_Δ is the difference between the angles of the velocity vectors.

Each triplet of points defines a unique plane going through them. In the case of rigid motion, the *def* value of the triplet is related to the structure and motion properties of this plane. The orientation of a plane in 3D space can be described in terms of its slant (σ) and

[1] Our analysis and model focus on the instantaneous optic flow information or, equivalently, two successive frames of a time-discretized velocity filed. As mentioned above, human observers benefit only very little from higher than first-order temporal information, and therefore, our two-frame approach is a plausible basis for modeling human SSFM. In case it is desired to extend the model to longer motion sequences, such an extension follows immediately from our analysis by repeating the computations for every successive pair of frames.

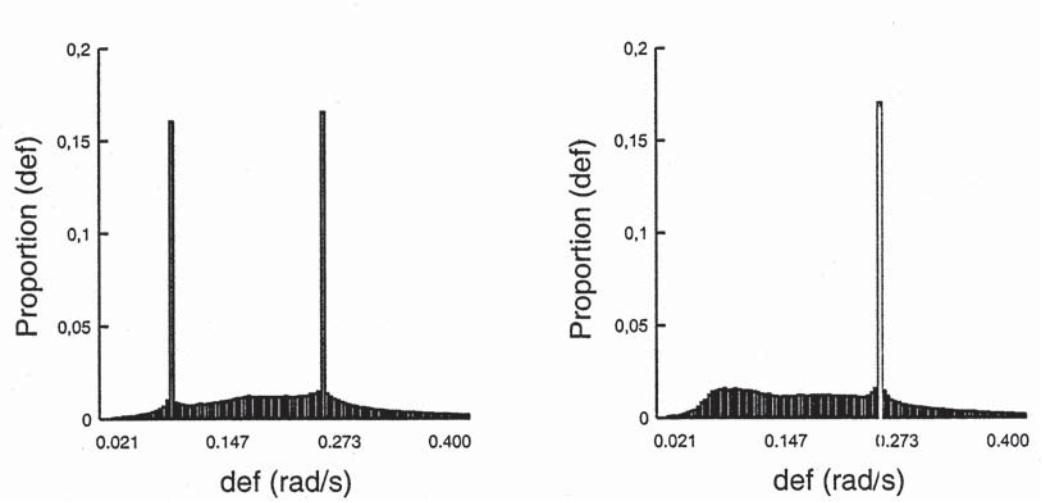

Fig. 2. Distributions of *def* values generated from all triplets drawn from 150 dots randomly located on planar surfaces rotating in 3D. Left, results for a velocity field produced by two transparent rigidly moving planes. Right, results for one plane embedded in noise. The stimulus parameters used for the simulations were identical to those used for the stimuli of Experiment 1.

tilt (τ). Slant is defined as the tangent of the angle between the line of sight (the z axis) and the normal to the plane, i.e. it measures how much the plane defined by the triplet deviates from the frontoparallel, or image plane. Tilt is defined as the angle between the projection of the normal to the triplet-plane on the image plane and the x axis, i.e. it measures the direction of the slant. Define ω as the projection of the triplets' rotation vector, $\mathbf{\Omega}$, on the image plane (i.e. the x–y plane). It can then be shown that:

$$def = \sigma\omega \tag{6}$$

In the case of a velocity field produced by the rotation of a planar surface, the plane going through all triplets is one and the same, as well as its motion, and therefore all the triplets will generate the same magnitude of *def*. This observation is the basis for our model. We propose that human performance can be well modeled by an algorithm that uses *def* as the quantity, Q, in the general computational approach for human SSFM described previously.

Let us consider the velocity field produced by two transparent, planar surfaces with slants (σ_1,σ_2) rotating with different angular velocities ($\mathbf{\Omega}_1$, $\mathbf{\Omega}_2$). Given N dots in the display, the $\binom{N}{3}$ possible triplets of dots can be divided into three subsets. First, there are the triplets made of points that all lie on one planar surface. Next, there are the triplets made of points that all lie on the other surface. Triplets from these first and second subsets will generate values of *def* equal to $\sigma_1\omega_1$, and $\sigma_2\omega_2$, respectively. Finally, there are the 'spurious' triplets—those in which two dots come from one surface, while the remaining dot comes from the other surface. The planes passing through these spurious triplets vary greatly in their slant, and the *def* values derived from them (via Eq. (6)) are entirely accidental and unrelated to the underlying motion of the two planar surfaces that generated the image. In other words, the *def* values generated by spurious triplets will vary greatly from triplet to triplet. What follows from this analysis is that, if we compute *def* for all triplets in the image, we will obtain a bi-modal distribution: two peaks at the values $\sigma_1\omega_1$ and $\sigma_2\omega_2$ plus a noisy distribution generated by the spurious triplets.

This situation is demonstrated in Fig. 2. The left panel shows the distribution of *def* values generated from all triplets drawn from 150 dots randomly located on two transparent, rigidly moving planes. The two peaks in the distribution, located at the *def* values of the two planes, are clearly differentiated from the distribution generated by the spurious triplets. The right panel shows the distribution of *def* obtained by simulating a single rigidly rotating plane embedded in noise. This distribution has only one peak. (In both cases, the planes were generated with the same orientation and angular rotation as the stimuli in Experiment 1, see below). Thus, computing the distribution of *def* values for a velocity field offers a natural way to determine the number of moving planes in it. Our model therefore solves the SSFM problem for multiple transparent planar surfaces by the following steps:

- Compute the distribution of *def* values for all triplets in the image.
- Determine the number of peaks in the distribution; this is the number of distinct surfaces in the scene.
- Determine which point belongs to which surface by labeling the points from all triplets that contributed to each observed peaks.

The detailed algorithm to perform steps 2, 3 is given in Section 2. Also, in Experiment 5 we describe a modification to the model, where, in step 1, the *def* distribution is computed only for a local neighborhood of each dot instead of the entire image.

1.2. Overview to experiments

In the next sections, we describe five experiments designed to test and compare the predictions of the model to human psychophysical performance. We presented observers with random-dot SFM displays that depicted one or two objects (with or without noise) and asked them to discriminate between those two cases. The observers' results were compared to those of the model. We were particularly interested in the *limitations* of the model, i.e. the circumstances under which it fails to perform the discrimination correctly. This occurs when the distributions of *def* for the cases of one and two objects become indistinguishable, which can happen as a result of added noise or specific manipulations of the geometry of the display, described below. Importantly, other models of SFM are unaffected by some of those manipulations, i.e. do not predict deterioration in performance, and therefore, testing human performance under those conditions is a strong test for our model.

The *def* distribution was systematically varied by perturbing planar surfaces into voluminous 'slabs' (Experiment 1) or corrugated surfaces (Experiment 4), changing the number of dots defining each surface (Experiment 2), and the amount of noise added to the stimulus displays (Experiment 3). Finally, in Experiment 5, we examined the role of local versus global computation of the *def* distribution. We modified the model to compute the *def* distribution only for a local neighborhood of each dot and compared the predictions of the modified ('local') and original ('global') models to human performance in SFM displays of curved surfaces.

2. General methods

2.1. Apparatus

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

2.2. Stimuli

The stimuli consisted of random-dot displays. We set the dots to the maximal electron-gun value; a homogenous region of that value had a 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 simulated an orthographic projection of either two surfaces or one surface embedded in nonrigid noise (extra noise dots were added to both kinds of stimuli in Experiment 3). The surfaces were oscillating in 3D about an axis of rotation contained in the image plane. Each oscillation cycle consisted of 78 frames (1083 ms). The stimulus remained on the screen until the observer gave his/her response.

Each stimulus display was contained within a circular 'window' with a diameter of 5.7° visual angle (420 pixels) to prevent changes in the projected contours of the simulated surfaces from being visible. The dots were randomly distributed with uniform probability density over the projection plane (not evenly distributed over the simulated surfaces).

2.3. Participants

The first author and naïve New York University undergraduate students participated in the experiments. The students received course credit for their participation. All observers had normal or corrected to normal vision.

2.4. Procedure

Participants were instructed that they would be viewing a series of random-dot displays and that the dots would appear to form surfaces moving in 3D. Observers were shown examples of surfaces like those that would be presented. Observers were told that their task was to determine whether, in each trial, the moving dots appeared to form two transparent surfaces oscillating with different angular velocities ('signal' trials), or one oscillating surface embedded in noise ('noise' trials). Observers provided their judgements with a keypress. 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 1.1 m. No restriction was placed on viewing time. Feedback was provided on each trial in the form of a beep for incorrect judgements. All participants were run individually in one session.

2.5. Design

In each experiment, the different experimental conditions were presented in different blocks (three conditions in Experiments 1, 2 and 4; four conditions in Experiments 3 and 5). In each block, each observer performed 80 trials (40 signal, 40 noise). The order of

presentation of the blocks, and the order of trials within each block, were randomized. Forty additional trials were presented at the beginning of each experimental session for practice.

2.6. Model simulations

We ran the model on the stimuli used in Experiments 1–5. Since the model operates on two-frame information, the analysis was based on the two frames where the planes reached the turning point in their oscillatory motion. (It is straightforward to extend the model to use more frames by accumulating the *def* distribution over time.) Since the dots' displacement between the frames was much smaller than the average distance between dots, we assumed that the solution to the correspondence problem was known.

We computed the value of *def* for all triplets of points in the display by using Eq. (5). To find the number of peaks in the resulting distribution of *def* values, it was first smoothed by convolving it with the Gaussian kernel:

$$f(t;h) = \frac{1}{nh}\sum_{1}^{n}\phi\left(\frac{t-x_i}{h}\right)$$

Here, $x_1,\ldots, x_n$ denote the *def* value of each bin in the histogram, $\phi(t)$ is the normal density function $(1/\sqrt{2\pi})\exp(-t^2/2)$, and h is the *window size*, which determines the amount of smoothing. The local maxima of the smoothed distribution were then found numerically. The value at each maximum was compared to a preset threshold, and if it was higher, that maximum was deemed a peak in the distribution. The values of the window size, h, and the threshold were chosen empirically in pilot trials so as to optimize performance in each experiment. The output was '1 object's rotation' if only one local maximum was above threshold, or '2 objects' rotations' if more than one local maximum was above threshold. This output was taken as the model's 'response', and subsequently, a d' measure was computed, as for the human observers.

3. Experiment 1

A main feature of our proposed model is that it works for surfaces but not for voluminous objects. Recall that the sharp peaks in the distributions of *def* in Fig. 1 were obtained because all the triplets on a given planar surface have the same slant, and therefore (Eq. (6)), the same *def* value. In contrast, if the dots are distributed inside a voluminous object, the planes defined by different triplets will have a wide range of slants, and therefore, the *def* distribution will be broad and flat. An example of this is shown in Fig. 3 for the stimuli used in Experiment 1. To generate those stimuli, we perturbed the depth value of the individual dots on planar surfaces, thus creating voluminous 'slabs'. Fig. 3 shows the *def* distribution of such slabs; the perturbation was smaller for the upper panels than for the lower panels. As the objects deviate from planarity, the peaks in the *def* distribution broaden and decrease in height, until at a certain point it is no longer possible to distinguish between distributions generated by two slabs (left) and those generated by one slab embedded in noise (right). A straightforward prediction of the model is therefore that human performance will degrade as the magnitude of perurbation of the planar surfaces (or, equivalently, the thickness of the slabs) grows. This prediction separates our model from that of Ullman (1979), or possible extensions to solve SSFM by other existing SFM models that are based on veridical recovery of Euclidean or affine 3D structure (e.g. Todd, 1982, 1984; Todd & Bressan, 1990; Longuet-Higgins & Pradzny, 1984; Bennett & Hoffman, 1986; Hoffman & Bennett, 1986; Koenderink & van Doorn, 1991), because these models work well for displays generated by voluminous objects. Thus, establishing human SSFM performance for flat versus voluminous objects is a strong test of our model. This was done in Experiment 1.

3.1. Method

3.1.1. Participants

The first author and four naïve observers participated in the experiment.

3.1.2. Stimuli

The displays simulated either the oscillation of two transparent overlapping perturbed planes ('slabs') or of one slab embedded in dynamic noise. To generate the slabs, at the beginning of each cycle, each dot on the plane was given a random value in the range of $\pm\zeta$, and its 3D position was moved by that amount perpendicular to the plane. In three experimental conditions, ζ took on the values of 0, 0.076, 0.152 times the radius of the stimulus window (210 pixels). The '*signal*' trials simulated two (perturbed) planes oscillating about a vertical axis that lied in the image plane. Their maximum angular displacements were $\theta_1 = 70°$ and $\theta_2 = 50°$. Each plane was defined by 150 dots. The '*noise*' trials simulated one (perturbed) plane embedded in nonrigid noise and oscillating about a vertical axis in the image plane ($\theta'_1 = 70°$). The simulated plane was defined by 150 dots. The additional 150 'noise' dots were assigned projective displacements incompatible with a rigid rotation in 3D. Each of these 150 dots simulated a different amount of 3D rotation randomly chosen in the range 40–60°.

3.1.3. Design

The amount of perturbation of the simulated planes was the only within-subjects independent variable. Three values were used in three separate blocks. Otherwise, the design was like that described in Section 2.

3.2. Results

The experimental data were analyzed by means of a signal detection paradigm in which the simulations of two rigid rotations were considered as signal trials (Green & Swets, 1966). A d' measure was computed for each observer based on the 80 trials in each experimental condition. Fig. 4 presents the results of the individual subjects in terms of their d' as a function of ζ (the maximum perturbation magnitude, or 'slab thickness'). All five subjects show deterioration in their performance for a larger ζ, as predicted by our model. A within-subjects analysis of variance (ANOVA) was conducted using the d's as the dependent variable. The effect of ζ, the magnitude of perturbation of the simulated planes, was significant [$F(2,8) = 13.45$, $P < 0.01$]. Post-hoc comparisons [Tukey's Honestly Significant Difference (HSD), $P < 0.05$] showed d's for planar surfaces to be significantly higher than those with the smallest amount of perturbation.

3.3. Discussion

The results of Experiment 1 indicate that human SSFM is strongly affected by the amount of perturbation applied to the simulated planes: as the magnitude of perturbation increased, performance decreased. This confirms the prediction of our *def*-based model, that SSFM performance should be better for flat surfaces than for voluminous objects. To quantify the prediction, we ran the model on the stimuli of Experiment 1 and computed its d' values (as described in Section 2)

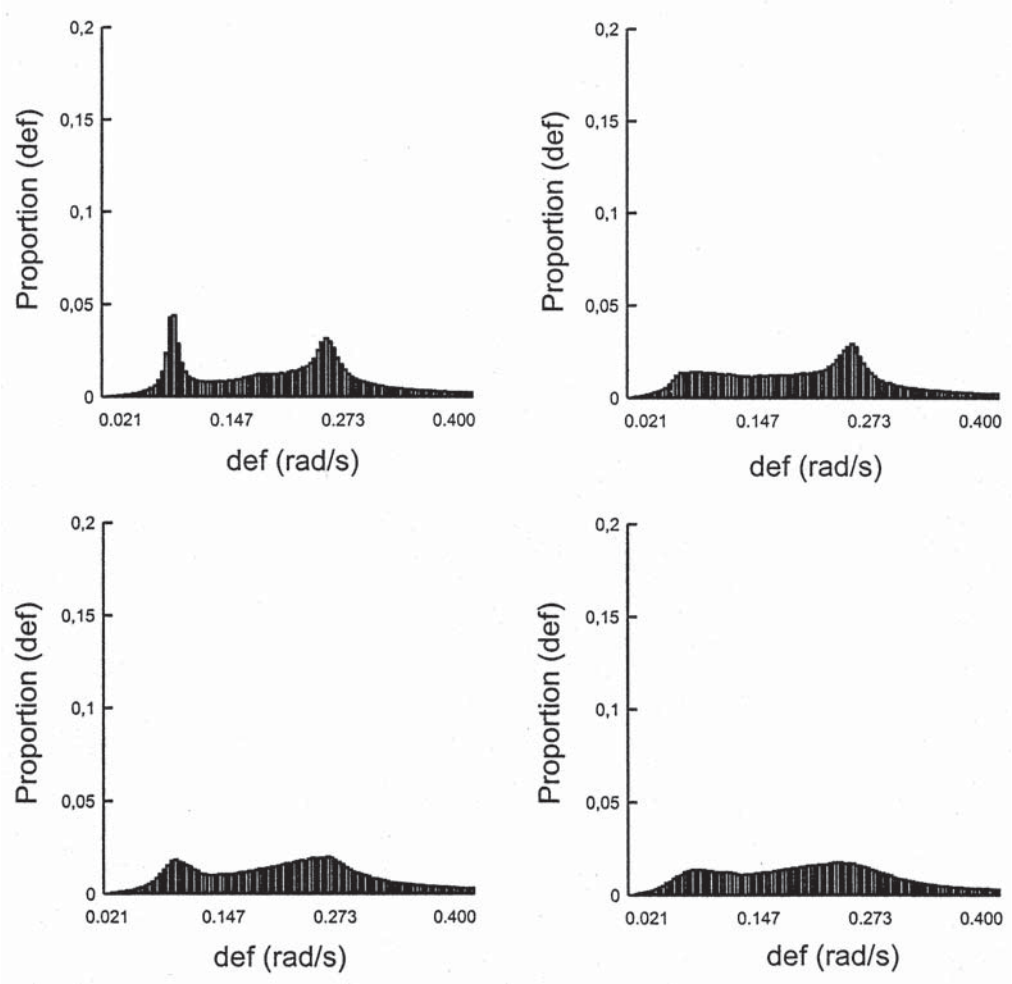

Fig. 3. Example distributions of *def* when the moving objects deviate from planarity. Here, the depth value of the dots on planar surfaces was perturbed so as to generate voluminous 'slabs', like those used as stimuli in Experiment 1. The magnitude of perturbation is 0.076 times the radius of the stimulus window for the upper panels and 0.228 times the radius of the stimulus window for the lower panels. The left panels show the distributions for two objects, and the right panels for one object embedded in noise.

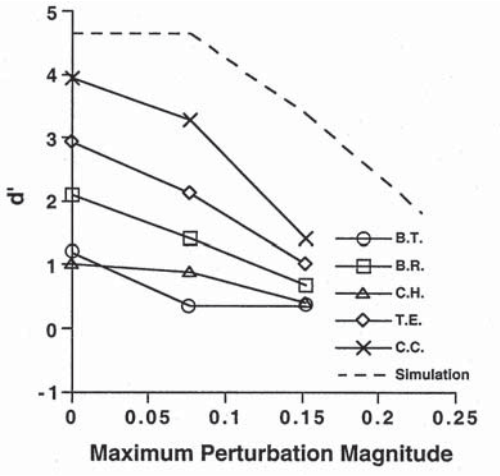

Fig. 4. d' scores as a function of perturbation magnitude (ζ) of planar surfaces in Experiment 1. ζ is expressed as a proportion of the radius of the stimulus window. The hatched line represents the performance of the *def*-based model.

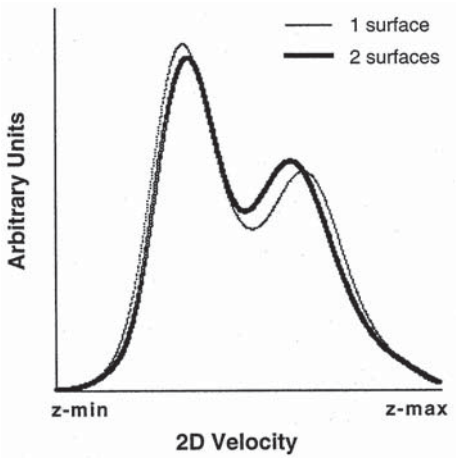

Fig. 5. Smoothened frequency plot of the distribution of 2D velocities for signal (two surfaces) and noise (one surface) trials of Experiment 1. These data were generated by normalizing the velocity vectors associated to each possible neighborhood of 15 points within the stimulus display.

for the different levels' perturbation. The results are shown in Fig. 4, together with those of the human observers. We find a marked drop in the performance of the model as the magnitude of perturbation (i.e. the thickness of the simulated slabs) is increased. Furthermore, although the absolute performance of the model is higher than that of our observers, the rate at which it deteriorates is similar. We should note that our aim here was not to obtain a good 'fit', or quantitative agreement between the model and human performance. The absolute level of performance of individual observers depends on factors such as their level of alertness and experience in SFM tasks or, generally, the level of internal noise (Pelli & Farell, 1999), which we did not wish to include in the model. Instead, we focus on the strong *qualitative agreement* between the trends shown by human observers and those predicted by the model. This agreement is in stark contrast to the expected prediction of any SSFM model based on SFM approaches that recover the Euclidean or affine 3D structure and motion parameters. This approach recovers the veridical solution for both planar and voluminous objects. (In fact, Ullman's (1979) algorithm requires non-planar quadruples of points and, therefore, would work for the 'slabs' but not for planar surfaces. However, given the subsequent work in single-object SFM, it would be straightforward to extend Ullman's approach to SSFM to work for planar surfaces as well.) Thus, such models predict that human performance will not be affected by the amount of perturbation of the surfaces in Experiment 1, as opposed to the marked deterioration found in reality.

On first impression, the poor SSFM performance in the case of voluminous objects may appear as a serious limitation of the human visual system. However, note that in fact, ecologically, most of the visual information is obtained from the surface of objects (Gibson, 1950). Thus, a computational approach that works only for flat (or piecewise-flat; see Experiment 5) surfaces makes sense, especially if it has advantages in other aspects. One such advantage of the *def*-based approach is that it is computationally simple: it is based on spatial derivatives of the optic flow field (Eq. (1)), which is a natural quantity for biological systems to compute (cf. Marr, 1982; Koch & Segev, 1998). In contrast, solving SSFM by recovering the full 3D structure and motion parameters involves the solution of complex nonlinear equations (Ullman, 1979).

If information about the number of objects in SSFM display is contained in the distribution of *def*, which is computed from the 2D velocity field, perhaps it can be found more directly in the velocity field itself. Specifically, the following simple hypothesis may be put forward: two objects are perceived if the distribution of 2D velocities is bimodal; one object is perceived if this distribution is unimodal. To test it, we calculated the distribution of 2D velocities for the stimuli of Experiment 1. To overcome variations of speed across the display, all the possible sets of 5, 10, 15, and 20 neighboring dots were selected within it. For each of these sets, the velocity magnitudes of each dot were computed. The distributions of all groups were normalized and converted to z-values, and those were pooled to create one histogram for the entire display. Examples of normalized distributions of 2D velocities for both 'signal' and 'noise' stimuli are shown in Fig. 5.

The results indicate that the distributions of 2D velocities for two surfaces and one surface in noise are similar, and therefore do not provide information sufficient for discriminating 'signal' and 'noise' trials. The nonrigid motion in the 'noise' trials, in fact, was gener-

ated so as to have the same mean velocity as the corresponding rigid motion in the 'signal' trials (see Section 3.1). Thus, the good performance of human observers in the case of flat surfaces cannot be accounted for by an analysis of 2D velocities.

4. Experiment 2

Experiment 2 manipulated the number of dots in the stimulus displays. Effects of dot numerosity have been found in several SFM tasks. Turner et al. (1995), for example, found that the ability of observers to discriminate quadratic surfaces from points randomly placed in a volume improves as the number of dots increases. Liter et al. (1994) found that the detection of one versus two objects in SFM increases as each object is defined by more dots. In the context of our model, we should expect a decrease in performance as the number of dots decrease, since the height of the peaks in the *def* distributions decreases with dot number. For example, if there are 12 points defining each plane, there are 220 triplets that belong to each plane and have the same *def* value, leading to high peaks in the distribution. Although the number of the spurious triplets is also high (1584), their *def* values will be distributed randomly, so that the two 220-high peaks can be detected easily. In contrast, if there are only four points defining each plane, the number of triplets belonging to each plane—and therefore the height of the two peaks—is just 4, making their detection mong the noisy distribution generated by the 48 spurious triplets difficult. The purpose of Experiment 2 was therefore to measure the effect of dot numerosity on psychophysical performance and to compare it to the effect on the *def*-based model.

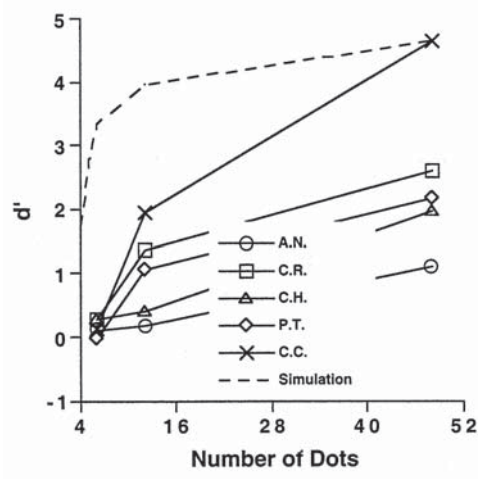

Fig. 6. d' scores as a function of the number of dots defining the surfaces in Experiment 2. The hatched line represents the performance of the *def*-based model.

4.1. Method

4.1.1. Participants

The first author and four naïve observers participated in this experiment. The naïve observers had not participated in Experiment 1.

4.1.2. Stimuli

The signal and noise trials were generated as in Experiment 1, except that only planar surfaces were simulated ($\zeta = 0$). Each surface was defined by either 6, 12 or 48 dots.

4.1.3. Design

The number of dots in each stimulus display was the only within-subjects independent variable. Three values were used in three separate blocks. Otherwise, the design was like that described in Section 2.

4.2. Results

A d' measure was computed for each observer and each stimulus condition. Fig. 6 presents the results in terms of d' of the individual subjects as a function of the number of dots. All observers show a marked improvement in their performance as the number of dots defining the moving planes grew. A within-subjects ANOVA was conducted using the d's as the dependent variable. The effect of the number of dots was significant [$F(2, 8) = 15.13$, $P < 0.01$]. With 48 points, according to Marascuilo's one-signal test (Marascuilo, 1970), d' was significantly above zero ($P < 0.05$) for all observers; with 12 points, d' was significantly above zero for three observers; with 6 points, d' was not significantly above zero for any observer.

4.3. Discussion

The results of Experiment 2 indicate that dot numerosity had a large effect on performance in a SSFM discrimination task. None of the observers was able to perform above chance when the simulated surfaces were defined by only 6 dots. In contrast, d's were all significant when each surface was defined by 48 dots. This confirms the prediction of our model that performance should decrease as the number of dots drops. To compare the performance of human observers to that predicted by the model quantitatively, we performed a simulation and computed d' as before. We used 4, 5, 6, 12 and 48 dots per surface, with 200 stimuli in each condition. The results are shown in Fig. 6, together with those of the human observers. Again, we are interested in the trend, or the shape of the curve, not the absolute value of performance. The model's performance increases with dot number, roughly paralleling the psychophysical results. It can be seen, however, that

the model's performance rises faster than that of the human observers at small dot numbers (4–12). This may reflect internal noise, which affects human observers but not the model (in its present form), thus allowing it to perform above chance in small dot numbers. In addition, it may be that human observers gather information from local regions, as will be discussed in Experiment 5. In contrast, the model collected all the triplets in the display, even those that contain widely separated dots. Such a difference will have a greater effect in small dot numbers, where the dot density is low, drastically reducing the number of triplets in local sub-regions of the display.

Our results are consistent with those of Andersen (1996), who found that more than 20 dots were necessary for detecting sinusoidally corrugated surfaces defined by motion parallax information. They stand in contrast, however, with results such as those of Turner et al. (1995), who found that 5 points were sufficient to discriminated simulated rigid 3D motion from SFM displays depicting nonrigid noise. It is likely that dot number (or density) has a greater effect on tasks that rely on the formation of surfaces, like those that were used here and by Andersen (1996). The perceptual formation of surfaces has been previously shown to play an important role in SFM (Husain, Treue, & Andersen, 1989; Treue et al., 1991; Treue, Andersen, Ando, & Hildreth, 1995; Hildreth et al., 1995; Andersen & Atchley, 1997; Atchley et al., 1998).

5. Experiment 3

The purpose of Experiment 3 was to determine whether the signal-to-noise ratio similarly affects human performance and the *def*-based model. In this experiment, the stimuli were similar to those of Experiment 1, except that nonrigid noise was added to both signal and noise trials.

5.1. Method

5.1.1. Participants

The first author and four naïve observers participated in this experiment. The naïve observers had not participated in the previous experiments.

5.1.2. Stimuli

The displays were similar to those used in Experiment 1 and simulated the rotation of planar surfaces. Each simulated plane was defined by 18 dots. In the 'signal' trials, two planes were simulated as oscillating in 3D about a vertical axis in the image plane. The maximum angular displacement for the two planes was $\theta_1 = 25°$ and $\theta_2 = 35°$. In each trial, the initial 3D orientation of each of the simulated planes was randomly determined. Slant was chosen in the range (25°, 40°). Tilt was chosen in the range [60°, 90°] (first surface) and in the range [0°, 30°] (second surface), or in the range [300°, 270°] (first surface) and [0°, 330°] (second surface). In addition, a varying number of noise dots (0, 9, 18, 36) were added to each displays. These dots simulated nonrigid rotation about a vertical axis in the image plane by randomly assigning each of them a different angular displacement in the range (0, $(\theta_1 + \theta_2)/2$). At the beginning of the oscillation sequence, the noise dots lay on a plane, whose orientation was randomly chosen from the range 25–40° (slant) and 0–360° (tilt). Because of the different angular displacement (and thus also angular velocity) of each dot, they deviated from this simulated 3D plane once the motion sequence started, leading to nonrigid motion.

The 'noise' trials were generated by simulating two groups, of 18 dots each, as follows. The initial positions of the dots in the two groups defined two planes, with slant and tilt chosen as for the 'signal' trials. The motion of one group of points simulated a rigid rotation, with an angular displacement of θ_1 or θ_2, depending on which plane they lay on. The remaining group simulated non-rigid motion by assigning each dot a random angular displacement in the range (0°, 1.5 θ_1) or (0°, 1.5 θ_2). As for the 'signal' trials, an additional group of 0, 9, 18 or 36 dots simulating nonrigid motion were added to the displays. They were generated as in the 'signal' trials.

5.1.3. Design

The number of noise dots was the only within-subjects independent variable. Four values were used in four separate blocks. Otherwise, the design was like that described in Section 2.

5.2. Results

A d' measure was computed for each observer based on the 80 trials in each of the four experimental condition. Fig. 7 presents the results of the individual subjects in terms of their d' as a function of the number of noise dots. All subjects show a deterioration in their performance as the number of noise dots is increased. A within-subjects ANOVA conducted using the d's as the dependent variable showed a significant effect of number of extra noise dots [$F(3,12) = 31.71$, $P < 0.001$]. Post-hoc comparisons (HSD, $P < 0.05$) showed d's for planar surfaces with no extra noise dots to be significantly higher than those for nine extra noise dots. d's for nine extra noise-dots were significantly higher than those for 36 extra noise-dots.

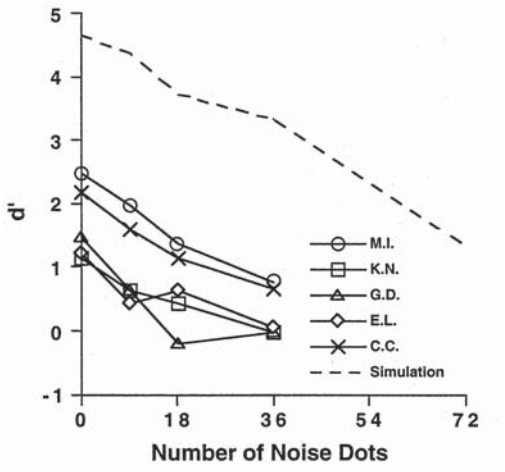

Fig. 7. d' scores are plotted as a function of the number of noise dots added to the displays of Experiment 3. The hatched line represents the performance of the *def*-based model.

5.3. Discussion

Noise had a strong detrimental effect on performance even though, at least for some observers, discrimination was still above chance when half or more of the dots in the signal trials were noise dots. This ability to perform successfully the segmentation task, even at these signal-to-noise ratios, replicates the results of Liter et al. (1994) with a different methodology. To compare the results to the performance of the model, we conducted simulations on displays like those used in Experiment 3, with the addition of a condition with 72 extra noise dots (200 trials in each of the five conditions). The simulation was performed by using the same procedure as that in Experiment 1. The d's obtained for the model are shown in Fig. 7. We find that the model, like the human observers, exhibits a smooth degradation with a decreased signal-to-noise ratio. As before, the important part of these quantitative results is not the absolute level of performance of the model, but rather its rate of degradation, which parallels that of the human observers.

6. Experiment 4

In Experiment 1, we found that the ability of observers to detect whether a SSFM display depicted one or two objects' rotations decreased as the simulated planar surfaces were perturbed into voluminous 'slabs'. The purpose of Experiment 4 was to extend these results. We simulated sinusoidally corrugated surfaces, varying the amplitude of corrugation. The *def*-based model predicts that performance should worsen as the amplitude increases. The reason is that, similarly to the case of the stimuli of Experiment 1, the deviation from planarity causes the triplets of dots in each surface to have a wide spectrum of *def* values, making it difficult to distinguish between distributions generated by two surfaces and those generated by a single surface embedded in noise.

6.1. Method

6.1.1. Participants

The first author and five naïve observers participated in this experiment. The naïve observers had not participated in the previous experiments.

6.1.2. Stimuli

The simulated surfaces were sinusoids of the form: $z = A \sin(fy)$, with the amplitude, A, taking the same values as the parameter ζ in Experiment 1: 0, 0.076, 0.152 times the radius of the stimulus window. The frequency, f, was 5π, and y ranged from -1 to $+1$, i.e. there were 5 cycles of corrugation within the 5.7° visual-angle window. Apart from these differences, the stimuli were like those of Experiment 1.

6.1.3. Design

The amplitude of the simulated sinusoidally corrugated surfaces was the only within-subjects independent variable. Three values were used in three separate blocks. Otherwise, the design was like that described in Section 2.

6.2. Results

A d' measure was computed for each observer based on the 80 trials in each stimulus condition. The results are shown in Fig. 8. As predicted by the model, all observers showed a deterioration in their performance as the amplitude of corrugation was increased. A within-subjects ANOVA was conducted using the ds as the dependent variable. The effect of the amplitude of the sinusoidally corrugated surfaces was significant [$F(2,10) = 16.59$, $P < 0.001$]. Post-hoc comparisons (HSD, $P < 0.05$) showed ds for planar surfaces to be significantly higher than those for the smallest amplitude.

6.3. Discussion

Experiment 4 extends the results of Experiment 1 to SSFM displays depicting surfaces perturbed away from planarity by sinusoidal corrugation. The amplitude of the sinusoidal corrugations had a similar effect to that of the thickness perturbation of Experiment 1. According to our *def*-based model, in both cases, the manipulation of the independent variable affected the difficulty of determining whether the *def* distribution had one or two peaks, because the perturbations broadened the

peaks and flattened both distributions. Consequently, it became harder to distinguish whether the stimulus was produced by two rigid rotations (two peaks) or by one rigid rotation embedded in noise (one peak). The results of model simulations based on 200 randomly generated displays for each of the three conditions of Experiment 4 are shown in Fig. 8. As in the previous cases, the absolute level of performance of the model is higher than that of the human observers, but the rate of degradation closely parallels the psychophysical curves.

There is, however, a difference between the situations in Experiments 1 and 4: as we noted in the discussion of Experiment 1, the results there made sense from an ecological point of view, since most of the visual information arrives from the surfaces of objects and not from their voluminous 'inside'. But in Experiment 4 the stimuli were not voluminous—the planar surfaces were perturbed by 'bending' them. In other words, they *could* arise from the surface of a (non-planar) object. How do we explain, therefore, that the model (as well as the human observers) performed so poorly in the high amplitude case? Surely, we do not mean to suggest that the visual system is capable of performing SSFM only for planar surfaces. The answer is, of course, that this will depend on the extent of deviation from planarity. Applying ecological considerations again, highly rough surfaces such as those used in Experiment 4 are not very common in nature. However, it is reasonable to expect good SSFM performance for smoothly, moderately curved surfaces, like the faces of everyday objects. We need, therefore, to address the question of how to extend the model to such non-planar surfaces. This will be done in the next section.

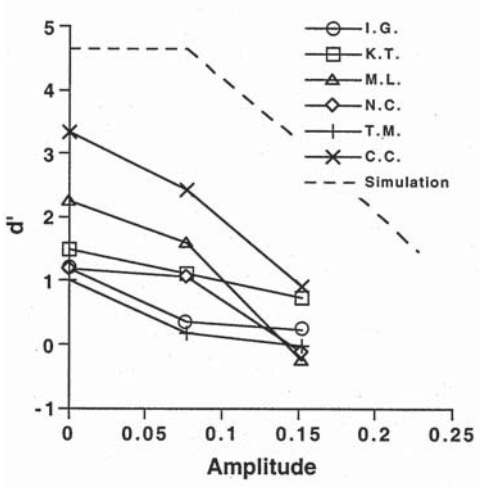

Fig. 8. *d′* scores as a function of corrugation amplitude in Experiment 4. Corrugation amplitude took on the same values as the parameter ζ in Experiment 1. The hatched line represents the performance of the *def*-based model.

7. Experiment 5

The *def*-based model described in Section 1 was based on a *global* computation: the distributions of *def* values were computed by collecting information from every triplet of dots in the display. Consequently, triplets made of dots relatively far apart from each other contributed to the result as much as triplets made of closely neighboring dots. This led to a situation where any deviation from planarity caused deterioration in the performance of the model. The reason is that the model relied on all triplets of a surface having the same (or very similar) *def* value, and that value is directly related to the slant of the plane defined by the triplet (Eq. (6)). But when a surface is curved, triplets of dots on it can have a wide range of slants, leading to a *def* distribution with no clear peaks. The limitation in performance on curved surfaces made the global-computation model, albeit useful to illustrate the principles of a *def*-based approach, unrealistic. Clearly, human SSFM is not limited to planar surfaces. For example, the global-computation model would fail on the stimulus introduced by Ullman (1979) of two rotating concentric cylinders, while human observers interpret this stimulus easily. Therefore, we should expect from a realistic model of SSFM to perform well also for curved surfaces.

The natural way to extend the model so that it works on curved surfaces is to *restrict the computation of def to local neighborhoods*. Mathematically, any surface can be approximated as piecewise planar (e.g. a ball can be approximated by a collection of flat pentagons and hexagons sewn together appropriately; in fact, this is how soccer balls are made traditionally). Conversly, a small area just around any point on a surface can be approximated as a plane. Therefore, by computing *def* distributions only from localized regions, we are practically back in the situation of planar surfaces, for which we already have a solution. However, for this approach to work, we need to choose the right scale of local neighborhoods, or size of 'integration window' around each point. On the one hand, this window should be small enough so that the planar approximation is still appropriate. On the other hand, it cannot be too small—since we need to have enough dots in it to obtain the distribution of *def*. The results of Experiment 4 can be understood now in this context: as the amplitude of corrugation grew, the area that can be approximated as a plane around each point shrank, until it was too small to obtain a reliable *def* distribution. What determines when the area is 'too small'? The number of dots in a region is determined by the density of dots, and therefore, a higher density will allow smaller regions to be used. But when considering the visual system, it is likely that constraints of biological implementation confine the size of the integration win-

dow to be within some range, setting a limit on how small it can get. The purpose of Experiment 5 was to test the limits of human performance as a surface becomes increasingly curved. As in Experiment 4, we simulated sinusoidal surfaces, but this time, we fixed the amplitude of corrugation and varied its frequency. A low frequency means a small amount of corrugation, and therefore we expect performance to be good, and to deteriorate as the frequency is increased.

In addition, we modified the *def*-based model described before so that it computed the distributions of triplets only within restricted regions. From here on, we shall call this modified version the *local model*, and refer to the original model as the *global model*. We ran the local model on the stimuli of Experiment 5, to verify that indeed it can handle curved surfaces, and compared its performance to the psychophysical results.

7.1. Method

7.1.1. Participants

The first author and four naïve observers participated in this experiment. The naïve observers had not participated in the previous experiments.

7.1.2. Stimuli

The stimuli were similar to those used in Experiment 4, except for the parameters used. The simulated surfaces were sinusoids of the form: $z = A \sin(fy)$. The amplitude, A, was fixed at 160 pixels, or 0.38 times the radius of the stimulus window (more than twice than that of the largest amplitude used in Experiment 4). The frequency f was 0, 0.75π, 1.5π, 6π; and y ranged from -1 to $+1$. Each surface was defined by 250 dots. The maximum magnitudes of rotation for the two simulated surfaces were $\theta_1 = 55°$ and $\theta_2 = 40°$. The nonrigid motion was produced by randomly choosing a different magnitude of rotation in the range 25–55° for each of the 250 dots defining one of the two simulated surfaces.

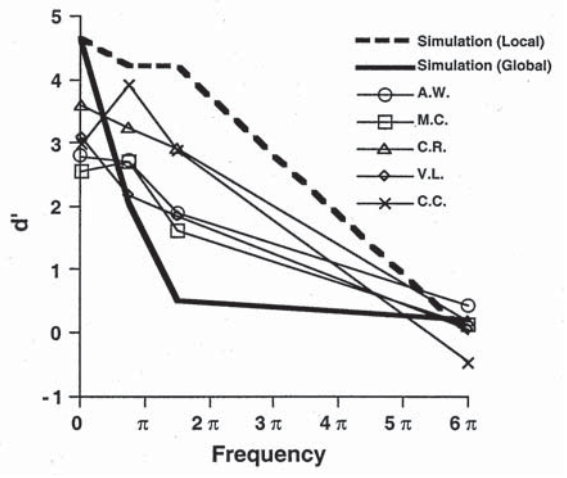

Fig. 9. d' scores as a function of corrugation frequency in Experiment 5. The thick line represents the performance of the global model; the thick hatched line represents the performance of the local model.

7.1.3. Design

The frequency of corrugation of the simulated sinusoidal surfaces was the only within-subjects independent variable. Four values were used in four separate blocks. Otherwise, the design was like that described in Section 2.

7.2. Local model

For the simulations of the local model, the display was decomposed into local regions, and *def* distributions were computed independently for each region. The centers of the local regions were located at the vertices of a 7×7 regular lattice. Their size was determined empirically in pilot trials so as to optimize performance in each condition. Only triplets made of dots that all fell within a region contributed to the computation of the distributions. The distributions of all groups were then normalized to units of Z-scores, and the distributions were subsequently pooled together. If there were less than 16 points within a local region, it did not contribute to the pooled distribution of *def* values.

7.3. Results

A d' measure was computed for each observer and stimulus condition. d' was based on 80 trials, half of which were signal trials. The results are shown in Fig. 9. We find that observers are able to perform the task well at the lowest frequency of corrugation (0.75π); they exhibit a slight deterioration at the intermediate frequency (1.5π), but perform very poorly at the highest frequency (6π). A within-subjects ANOVA conducted using the d's as the dependent variable showed a significant effect of the frequency of the sinusoidal surfaces [$F(3,12) = 45.01$, $P < 0.001$]. When the frequency was equal to 6π, detection differed significantly from the average of the other conditions [$F(1,4) = 70.11$ $P < 0.01$]. In this condition, d' was not significantly above zero ($P < 0.05$) according to Marascuilo's one-signal test for any of the observers. d' was significantly above zero for all observers in all other conditions.

7.4. Discussion

The results of Experiment 5 indicate that perceptual performance was strongly affected by the frequency of the simulated sinusoidal surfaces: as frequency increased, performance decreased. To compare these results with the model predictions, we performed

simulations of both the local model and the global model. Each model was run on 200 randomly generated displays (100 signal displays) for each of the four conditions of Experiment 5. For the global model, only 100 dots were used in the computation, so as to approximately equate to the number of triplets used in the simulation of the local model. A d' measure was computed for each model and each condition, and the results are shown in Fig. 9, together with those of the human observers. The two models perform similarly for the planar surfaces (frequency 0), but the global model shows a sharp decline in performance already at the lowest frequency of corrugation (0.75π), and falls apart completely at the intermediate frequency (1.5π). In contrast, the local model maintains its good level of performance at those corrugation frequencies and shows a significant decline only at the highest frequency (6π). This behavior parallels that found for the human observers. Thus, the strategy of restricting the computation of *def* to local neighborhoods overcomes the limitations of the global model and offers a good model for human performance.

To further test the local model, we ran it on the two-cylinder stimulus introduced by Ullman (1979). We tested whether the model would be able to discriminate one- from two-cylinder displays. Because each transparent cylinder produces two velocity fields that move in opposite directions (the 'front' and 'back' faces), we could perform a simple first segmentation step by dividing the vectors of the instantaneous velocity field into two sets according to their direction of motion (it is well known that human observers can easily perform such segmentation; see, for example, Qian, Andersen, & Adelson, 1994). For each of the resulting sets, we then computed the *def* distribution within each local regions and estimated the number of peaks as described before. In a simulation based on 200 trials (100 of each condition), the local model discriminated the two cylinder displays from a single cylinder embedded in nonrigid noise with $d' = 3.28$. Thus, the local *def*-based model can account also for the perception of these more naturalistic stimuli.

8. General discussion

We studied human performance in tasks that involve segmentation and the recovery of multiple-object structure from motion (SSFM) and presented a model for how SSFM may be accomplished by the visual system. The purpose of the psychophysical experiments was twofold: one, to characterize SSFM in greater detail than was done so far, and two, to test specific predictions of the model.

In five experiments, we compared perceptual performance with the outcomes of a model based on the distribution of the *def* component of the velocity field. Experiments 1, 4 and 5 examined the effect of the geometry of the moving objects observers' ability to discriminate two rigidly rotating objects from one rigid object embedded in noise; Experiments 2 and 3 quantified the effect of lowering the signal in the SSFM displays on performance. In Experiment 1, planar surfaces were turned into voluminous 'slabs' by perturbing the depth value of the dots defining them; discrimination performance decreased with increased magnitudes of perturbation. In Experiments 4 and 5 we manipulated the curvature of surfaces by gradually adding sinusoidal corrugation to planes. Performance was impaired when the amplitude (Experiment 4) or frequency (Experiment 5) of corrugation increased. Experiment 2 examined the effect of the number (or density) of the dots defining the surfaces. Discrimination performance increased with dot numerosity. Experiment 3 quantified the effect of adding noise to the displays, revealing that the visual system has a high tolerance for noise in discriminating one versus two rigid objects' rotations.

The model that we proposed is based on computing the distribution of values of a property of the velocity field called *def*, which is a function of the spatial derivatives of the projected (2D) velocity field, and in the case of discretized fields defined by dots (such as used in SFM displays), it is defined for triplets of dots. Intuitively, *def* measured the local amount of shear (or deformation) in the velocity field [together with two other measures, of the in/outflow ('*div*') and pure rotation ('*curl*'), they characterize the instantaneous velocity field up to a global rotation]. In the general case, *def* changes across the field; but in the special case of a rigidly rotating plane the value of *def* is constant, because it is related to the local slant of the surface, and a the slant of a planar surface is constant in space. Based on that, we proposed a model for how the visual system may perform SSFM. The model is based on computing the distribution of *def* values in the display and finding the number of peaks in the distribution. If two (or more) peaks can be identified in the distribution, then two overlapping surfaces will be perceived. If different peaks cannot be discerned, perceptual segmentation of the velocity field will not occur.

In the first version, which we called 'the global model', the *def* distribution was collected from all possible triplets in the entire display. This model (which is the simplest version of our proposed approach, for experimental and illustrative purposes), performs well for displays depicting two transparent, rigidly rotating planes, but breaks down for voluminous objects or curved surfaces. Simulations of the global model yielded a good agreement with human performance in Experiments 1–4, in terms of the amount of deterioration as the planes were perturbed in the ways described above. The modified version, termed 'the local model',

collected *def* distributions from spatially restricted neighborhoods of each dot. This model works well also for moderately curved surfaces, since they can be approximated as piecewise planar (but it still performs poorly for voluminous objects). Simulations of the local model yielded good agreement with the results of Experiment 5, showing a good performance for sinusoidally corrugated surfaces at moderate frequencies, and sharp deterioration as the frequency of corrugation increased.

Our findings complement previous reports which showed that the *def* component of the first-order velocity field is a good predictor of human performance in single-object SFM. This has been shown for the discrimination of rigid from nonrigid motion (Domini et al., 1997), the discrimination of constant from variable angular velocities in rotating ellipsoids and planes (Domini et al., 1998b), the perceived orientation of the axis of rotation and the accuracy in discriminating fixed-axis from nonfixed-axis rotations (Caudek & Domini, 1998), the perception of depth-order relations (Domini & Braunstein, 1998; Domini et al., 1998a), and the perception of surface slant (Loomis & Eby, 1988; Freeman, Harris, & Meese, 1996; Domini & Caudek, 1999). Recently, Andersen and collaborators have established the importance of *def* also in surface detection tasks. Andersen and Atchley (1997), for example, asked observers to discriminate velocity fields generated by points positioned on a corrugated 3D surfaces from velocity fields generated by points randomly positioned within a 3D volume. They manipulated frequency, amplitude, density and surface complexity, and found that human performance could be predicted by analyzing the difference between the expected values of the *def* distributions produced by signal and noise trials. Here, we extended their approach to SSFM by inspecting another property of the distribution of *def*: whether or not it is bimodal. The results of the simulations reveal that the delectability of bi-modality is indeed a good predictor of human SSFM. Moreover, by applying the *def*-approach to orthographic projections we extended the results previously obtained with perspective translations (see Andersen & Atchley, 1997).

The *def*-based approach can account for the ability of human observers to perform well in SSFM tasks in a variety of stimulus conditions. But, more importantly, it also predicts the *limits* of human performance. This latter fact is what distinguishes the *def*-based model from other models suggested previously. The predominant approach in SFM has been to recover the Euclidean or affine 3D structure of the moving objects, under a rigidity assumption that allows the space of solutions to be constrained (e.g. Ullman, 1979; Adiv, 1985; Hildreth et al., 1990; Koenderink & van Doorn, 1991). This approach is mathematically rigorous and produces the veridical solution (when it exists). This can be very useful for purposes of machine or computer vision, but as a model for human perception, this is precisely the weakness of the approach: it will produce veridical solutions also in cases where human observers fail to do so. In particular, a model based on a rigidity assumption will correctly recover the structure and motion parameters of all the stimuli used in Experiments 1, 4 and 5, because its validity is not restricted to near-planar surfaces but extends also to any curved surface and voluminous object. Thus, such a model would maintain its good performance even for the highest amounts of perturbation away for planarity in Experiments 1, 4 and 5, in stark contrast to the poor performance exhibited by our observers. In order to 'fit' the model to human performance we would have to introduce arbitrary, ad hoc limitation on it; in its essence, it will not show sensitivity to the variables we manipulated in the experiments. In contrast, our hypothesis that the human visual system relies on computing the distribution of the deformation component of the velocity field was able to account for the successful human performance as well as its failures. The limitations that we observed can thus be understood as natural consequences of the *def*-based approach, rather than arbitrary failures of the visual system.

Why would the visual system adopt a strategy that may lead to erroneous performance, rather than take a rigidity-assumption approach that guarantees a veridical solution? There may be several reasons for that. First, as already mentioned, in the generic case, the light reflected from objects comes from their surface, not the entire volume, since most objects are opaque. As we have seen (Experiment 5), the *def*-based approach is capable of doing a good job in SSFM of curved surfaces, so long as their curvature is not too extreme. In other words, the perturbed surfaces used in our experiments represent unusual cases crafted in lab conditions, and therefore, failing in those cases may be a small price to pay in order to obtain a more robust performance in the generic case. One significant advantage of the *def*-based approach is that it can produce acceptable results about structure and motion also for velocity fields which were not created by rigidly moving objects (e.g. the face of water). The analysis described in Eqs. 1–6 remains valid also in this case, allowing the slant of the surface to be estimated locally at various points. In contrast, all a rigidity-based algorithm could tell us in this case is that the velocity field did not originate from the motion of a rigid object. Another important advantage is that the computation of *def* requires only measurement of the spatial derivative of the velocity field. There are numerous examples of computations of spatial derivatives in the visual system (e.g. edges are obtained from the derivative of luminance), and it is generally believed that spatial derivatives are easy for neural systems to compute (cf. Marr,

1982; Koch & Segev, 1998). The *def*-based approach is therefore biologically plausible—an important attribute for any model that aims to capture human perception.

In conclusion, we showed that human observers are able to correctly segment velocity fields produces by two overlapping transparent objects in a wide variety of conditions, and presented a model that is based on the computation of a simple, local quantity—the deformation (*def*) component of the velocity field. The model can account for the good segmentation and structure from motion (SSFM) performance exhibited under certain conditions, as well as predict the conditions where SSFM will be impaired and shed light on the reasons for those limitations in performance.

Acknowledgements

Corrado Caudek was supported by the NYU Sloan Program for Theoretical Neuroscience. The authors wish to thank Shimon Ullman and Fulvio Domini for helpful discussions.

References

Adiv, G. (1985). Determining 3D motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 7*(4), 384–401.

Andersen, G. J. (1996). Detection of smooth three-dimensional surfaces from optic flow. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 945–957.

Andersen, G. J., & Atchley, P. (1997). Smoothness of the velocity field and three-dimensional surface detection from optic flow. *Perception & Psychophysics, 59*(3), 358–369.

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. (1986). The computation of structure from fixed axis motion: nonrigid structures. *Biological Cybernetics, 51*, 293–300.

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, A, 6*(7), 1052–1069.

Braunstein, M. L. (1962). The perception of depth through motion. *Psychological Bulletin, 59*, 422–433.

Braunstein, M. L. (1976). *Depth perception through motion*. New York: Academic Press.

Braunstein, M. L. (1989). Structure from motion. In J. J. Elkind, & S. K. Card, *Human performance models for computer-aided engineering* (pp. 89–105). Washington, DC: National Academy Press.

Braunstein, M. L. (1994). Decoding principles, heuristics and inference in visual perception. In G. Jansson, S. S. Bergstroem, & W. Epstein, *Perceiving events and objects: resources for ecological psychology* (pp. 436–446). Hillsdale, NJ: Erlbaum.

Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1990). Discriminating rigid from nonrigid motion: minimum points and views. *Perception & Psychophysics, 47*(3), 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 strucutre. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 335–343.

Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 609–621.

Caudek, C., & Proffitt, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception and Performance, 19*(1), 32–47.

Dijkstra, T. M. H., Snoeren, P. R., & Gielen, C. C. A. M. (1994). Extraction of three-dimensional shape from optic flow: a geometric approach. *Journal of the Optical Society of America, A, 11*, 2184–2196.

Domini, F., & Braunstein, M. L. (1998). Recovery of 3D structure from motion is neither Euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1273–1295.

Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance, 25*(2), 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 & Performance, 23*(4), 1111–1129.

Domini, F., Caudek, C., & Richman, S. (1998a). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics, 60*(7), 1164–1174.

Domini, F., Caudek, C., Turner, J., & Favretto, A. (1998b). Discriminating constant from variable angular velocities in structure from motion. *Perception & Psychophysics, 60*(5), 747–760.

Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion tasks. *Vision Research, 35*(20), 2927–2941.

Freeman, T. C., Harris, M. G., & Meese, T. S. (1996). On the relationship between deformation and perceived surface slant. *Vision Research, 36*(2), 317–322.

Gibson, J. J. (1950). *The perception of the visual world*. Cambridge, MA: Houghton Mifflin Co.–The Riverside Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gregory, R. L. (1998). *Eye and brain: the psychology of seeing* (5th ed.). Princeton, NJ: Princeton University Press.

Griffiths, A. F., & Zaidi, Q. (1998). Rigid objects that appear to bend. *Perception, 27*(7), 799–802.

Grzywacz, N. M., & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: position-based versus velocity-based formulations. *Journal of the Optical Society of America, A, 4*(3), 503–518.

Hildreth, E. C., Ando, H., Andersen, R. A., & Treue, S. (1995). Recovering three-dimensional structure from motion with surface reconstruction. *Vision Research, 35*(1), 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*(1), 19–36.

Hoffman, D., & Bennett, B. (1986). The computation of structure from fixed axis motion: rigid structures. *Biological Cybernetics, 54*, 1–13.

Hogervorst, M. A., Kappers, A. M., & Koenderink, J. J. (1996). Structure from motion: a tolerance analysis. *Perception & Psychophysics, 58*(3), 449–459.

Hogervorst, M. A., Kappers, A. M., & Koenderink, J. J. (1997). Monocular discrimination of rigidly and nonrigidly moving objects. *Perception & Psychophysics, 59*(8), 1266–1279.

Husain, M., Treue, S., & Andersen, R. A. (1989). Surface interpolation in three-dimensional structure from motion perception. *Neural Computation, 1*, 324–333.

Kaiser, M. K. (1990). Angular velocity discrimination. *Perception & Psychophysics, 47*(2), 149–156.

Kaiser, M. K., & Calderone, J. B. (1991). Factors influencing perceived angular velocity. *Perception & Psychophysics, 50*(5), 428–434.

Koch, C., & Segev, I. (1998). *Methods in neuronal modeling*: *from ions to networks* (2nd ed.). Cambridge, MA: MIT Press.

Koenderink, J. J. (1986). Optic flow. *Vision Research*, *26*(1), 161–179.

Koenderink, J. J., Kappers, A. M., Todd, J. T., Norman, J. F., & Phillips, F. (1996). Surface range and attitude probing in stereoscopically presented dynamic scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 869–878.

Koenderink, J. J., & van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America*, *A*, *3*(2), 242–249.

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America*, *A*, *8*(2), 377–385.

Landy, M. S., Dosher, B. A., Sperling, G., & Perkins, M. E. (1991). The kinetic depth effect and optic flow—II. First- and second-order motion. *Vision Research*, *31*(5), 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*(4457), 717–719.

Lappin, J. S., & Fuqua, M. A. (1983). Accurate visual measurement of three-dimensional moving patterns. *Science*, *221*(4609), 480–482.

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception*, *22*(12), 1441–1465.

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1994). Detection of one versus two objects in structure from motion. *Journal of the Optical Society of America*, *A*, *11*(12), 3162–3166.

Longuet-Higgins, H. C., & Pradzny, K. (1984). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, *208*, 385–397.

Loomis, J. M. & Eby, D. E. (1988). Perceiving structure from motion: failure of shape constancy. Paper presented at the Second International Conference on Computer Vision, Tampa, FL.

Marascuilo, L. A. (1970). Extensions of the significance test for one-parameter signal detection hypotheses. *Psychometrika*, *35*, 237–243.

Marr, D. (1982). *Vision*: *a computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman.

Miles, R. (1931). Movement interpretation of the silhouette of a revolving fan. *American Journal of Psychology*, *43*, 392–405.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics*, *53*(3), 279–291.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America*, *A*, *Optics*, *Image Science and Vision*, *16*(3), 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*(5), 666–679.

Pollick, F. E., Nishida, S., Koike, Y., & Kawato, M. (1994). Perceived motion in structure from motion: pointing responses to the axis of rotation. *Perception and Psychophysics*, *56*, 91–109.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *Journal of Neuroscience*, *14*(12), 7357–7366.

Rubin, N., Hochstein, S., & Solomon, S. (1995a). Restricted ability to recover three-dimensional global motion from one- dimensional motion signals: psychophysical observations. *Vision Research*, *35*(4), 463–476.

Rubin, N., Solomon, S., & Hochstein, S. (1995b). Restricted ability to recover three-dimensional global motion from one- dimensional local signals: theoretical observations. *Vision Research*, *35*(4), 569–578.

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature*, *384*(6608), 460–463.

Sparrow, J. E., & Stine, W. W. (1998). The perceived rigidity of rotating eight-vertex geometric forms: extracting nonrigid structure from rigid motion. *Vision Research*, *38*(4), 541–556.

Thompson, W. B., Kersten, D., & Knecht, W. R. (1992). Structure-from-motion based on information at surface boundaries. *Biological Cybernetics*, *66*(4), 327–333.

Todd, J. T. (1982). Visual information about rigid and nonrigid motion: a geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 238–252.

Todd, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, *36*(2), 97–103.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, *48*(5), 419–430.

Todd, J. T., & Perotti, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception & Psychophysics*, *61*(8), 1577–1589.

Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of three-dimensional space in the perceptual analysis of motion and stereo. *Perception*, *24*(1), 75–86.

Todorovic, D. (1993). Analysis of two- and three-dimensional rigid and nonrigid motions in the stereokinetic effect. *Journal of the Optical Society of America*, *A*, *10*, 804–826.

Tomasi, C., & Kanade, T. (1992). Shape and motion from image streams under orthography—a factorization method. *International Journal on Computer Vision*, *9*(2), 137–154.

Treue, S., Andersen, R. A., Ando, H., & Hildreth, E. C. (1995). Structure-from-motion: perceptual evidence for surface interpolation. *Vision Research*, *35*(1), 139–148.

Treue, S., Husain, M., & Andersen, R. A. (1991). Human perception of structure from motion. *Vision Research*, *31*(1), 59–75.

Turner, J., Braunstein, M. L., & Andersen, G. J. (1995). Detection of surfaces in structure from motion. *Journal of Experimental Psychology: Human Perception & Performance*, *21*(4), 809–821.

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*(3), 255–274.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*, 205–217.

Wallach, H., O'Connell, D. N., & Neisser, U. (1953). The memory effect of visual perception of three-dimensional form. *Journal of Experimental Psychology*, *45*, 360–368.