Pergamon

0042-6989(94)00136-7

*Vision Res.* Vol. 35, No. 4, pp. 569–578, 1995

0042-6989/95 \$9.50 + 0.00

# Restricted Ability to Recover Three-dimensional Global Motion from One-dimensional Local Signals: Theoretical Observations

NAVA RUBIN,*† SORIN SOLOMON,‡ SHAUL HOCHSTEIN*

*Received 30 August 1993; in revised form 28 February 1994; in final form 15 June 1994*

**Recovering 3D information from a 2D time-varying image is a vital task which human observers face daily. Numerous models exist which compute global 3D structure and motion on the basis of 2D local motion measurements of point-like elements. On the other hand, both experimental and computational research of early visual motion mechanisms emphasize the role of oriented (1D) detectors. Therefore, it is important to find out whether indeed 1D motion signals can serve as primary cues for 3D global motion computation. We have addressed this question by combining mathematical results and perceptual observations. We show that given the 2D-projected 1D instantaneous velocity field, it is mathematically impossible to discriminate rigid rotations from non-rigid transformations and/or to recover the rotation parameters. We relate this fact to existing results in cases where localized (point-like) cues are present, and to our own experiments on human performance in global motion perception when only 1D cues are given. Taken together, the data suggest a necessary role for localized information in early motion mechanisms and call for further physiological and psychophysical research in that direction.**

One-dimensional motion Two-dimensional motion Structure-from-motion

## 1. INTRODUCTION

When a large object passes through our visual field, the visual system faces a computational problem at more than one level. This is because, the receptive fields of primary visual cortical cells are too small to capture the motion of large objects. At one level, therefore, there is the problem of constructing *local motion detectors*: units that will reliably signal the motion measured in localized regions of the image. At another level, there is the problem of combining these local motion signals into the (correct) *global motion percept*.

Here, we analyze the computational constraints on models of integration of local motion signals for 3D global motion perception. We show that computational considerations have implications for the plausibility of different types of local motion detectors and of algorithms performing local-to-global motion computation (Marr, 1982).

### *Two approaches to global motion detection*

One family of local motion detectors that has been studied extensively, both experimentally in psychophysics and electrophysiology, and theoretically in modeling studies, is that of "oriented" units. These detectors are especially suited to signal the motion of straight (or approximately so) contours of a specific ("preferred") orientation within their receptive field. We therefore term them "(1D) detectors", and correspondingly use the term "1D velocity" (or 1D signals) to refer to the velocity component perpendicular to the contours' orientation (or to their neural representation). However, the motion of a straight contour is ambiguous because its motion along the direction of its elongation is unobservable (Wallach, 1935, 1976). Thus the motion information carried by 1D detectors is inherently ambiguous, a phenomenon commonly known as "the aperture problem".§

Previously it has been shown that in some cases, the aperture problem can be overcome by combining the

*Neurobiology Department, Institute of Life Sciences, and Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel.
†To whom all correspodence should be addressed at: Vision Sciences Laboratory, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, U.S.A. [*Email* nava@isr.harvard.edu].
‡Physics Department, Hebrew University, Jerusalem 91904, Israel.

§1D detectors respond also to the motion of non-elongated stimuli such as points or line endpoints (although usually to a lesser extent). However, these are spurious responses, in that they do not convey the veridical (and observable!) 2D stimulus velocity, but rather its component along the preferred direction of the 1D unit.

outputs of at least two oriented units. For the case of frontoparallel translatory motion, algorithms have been suggested which compute the global direction of motion exactly, such as the intersection of constraints (IOC) algorithm (Fennema & Thompson, 1979; Adelson & Movshon, 1982) or algorithms using a smoothing approach (Hildreth, 1984). For the case of frontoparallel rotation as well as translation, a procedure akin to the IOC can be used which produces an exact solution (Rubin, 1993). Moreover, it has been shown that resolving the aperture problem in the frontoparallel case can be realized by a simple neural network of only two layers of linear neurons (M. I. Sereno, 1989; M. F. Sereno, 1993).

Another possible approach to motion computation is to *avoid* the aperture problem in the first place, rather than *overcome* it in the local-to-global stage of the process. To do so, one must construct local motion detectors that would not suffer from the aperture problem, i.e. mechanisms which signal unambiguously the 2D velocity within their receptive fields, by detecting the motion of cues in the image such as points (localized regions of contrast gradients) and line endpoints. We shall term such mechanisms "2D motion detectors".

Until recently, this class of detectors received relatively little attention, both experimentally and theoretically. Electrophysiological research has concentrated on studying 1D, or oriented, motion sensitive cells (Orban, 1984), and the existence of local 2D motion detectors has not been directly demonstrated in electrophysiological experiments. In psychophysics too, research has been concentrated on orientationally tuned (1D) motion mechanisms (e.g. Anderson & Burr, 1985). Nevertheless, recently direct psychophysical evidence that 2D motion sensitive mechanisms exist in humans was provided (Anderson & Burr, 1991; Anderson, Burr & Morrone, 1991; Castet, Lorenceau, Shiffrar & Bonnet, 1993; Lorenceau, Shiffrar, Wells & Castet, 1993). A specific type of 2D units, namely line-terminator motion detectors, have been invoked to account for the so-called "barber-pole illusion" (Wallach, 1935; Shimojo, Silverman & Nakayama, 1989), and these authors have further hypothesized that the 2D motion-sensitive units should interact with mechanisms which signal occlusion and depth-relations between objects.

From a theoretical point of view, it has been demonstrated that the construction of 2D motion detectors necessarily involves non-linear stages of the "and" type, to avoid "false-positives"—spurious responses to stimuli that the unit is not designed to detect [for a comprehensive discussion of the issue see Zetzsche and Barth (1990)]. In contrast, the construction of oriented motion detectors generally involves non-linear operations of the rectifying or thresholding type (if any; see, e.g. Heeger, 1988 and references therein), and thus they lend themselves more easily to theoretical analysis. In addition, 1D detectors can rely on length summation and thus will tend to be more robust to noise in the image [note that in accord with the latter point, Lorenceau *et al.* (1993) and Castet *et al.* (1993) find that the 2D motion mechanisms are recruited only at higher contrasts].

The two different types of local motion detectors lead, in turn, to two different approaches to global motion computation, differing in which type of local motion cues is assumed to play the main role in the computation of global motion. One possibility is, that the primary source of motion information arises from elongated parts in the image (and thus is carried by 1D detectors). Because of the aperture problem, the visual system would then need to follow the detection of 1D motion cues with a second stage, to overcome the ambiguity inherent in the 1D motion cues. We shall term this "the 1D approach". Since its first explicit formulation (Adelson & Movshon, 1982), the results of many studies were interpreted as supporting the 1D approach (e.g. Welch, 1989; Derrington & Suero, 1991). The relatively more extensive theoretical knowledge of 1D detectors (compared to 2D detectors) further contributed to the general acceptance of the 1D approach.

The alternative approach suggests that the visual system contains mechanisms which can reliably detect the 2D velocity of localized parts of the image, and that it relies on the outputs of these mechanisms in performing global motion computation. Thus, the disambiguation of 1D motion signals occurs due to the contribution of *another* type of units—2D detectors—and not via computations based on the outputs of 1D units alone. We term this "the 2D approach".

Recent psychophysical studies provide support for the 2D approach. It was shown that when presented with images that contain no 2D motion cues or impoverished 2D motion cues, human observers in many cases fail to detect the (mathematically) correct global direction of motion and along with it also fail to see the rigid interpretation of the time-varying image (Nakayama & Silverman, 1988; Ferrera & Wilson, 1990, 1991; Yo & Wilson, 1992; Shiffrar & Pavel, 1991; Mingolla, Todd & Norman, 1992; Rubin & Hochstein, 1993; see also Derrington & Badcock, 1992). Furthermore, if the image *does* contain 2D motion cues, they can have a dramatic effect on the perceived motion. When the 2D motion of the points or line endpoints is equal to the global direction of motion, the result is a coherent, rigid (and mathematically correct) global motion percept (Nakayama & Silverman, 1988; Rubin & Hochstein, 1993). When the 2D motion cues conflict with the global direction of motion, the effect is a departure from a coherent global percept (Lorenceau & Shiffrar, 1992).

### *3D global motion computation from 1D vs 2D signals*

In this study we present a computational analysis which further supports the 2D approach. Our analysis departs from the restricted case of frontoparallel motion and considers rotation of objects in 3D space.

A key point in hypothesizing the 1D approach was the assumption that the aperture problem can be overcome by integration of information from several 1D detectors. As mentioned above, this is indeed the case for frontoparallel motion. However, frontoparallel motion constitutes a vanishingly small set among all the possible motion types human observers are likely to encounter.

The question arises, therefore, whether it is theoretically possible to use 1D local motion signals to recover global object motion in the more general case of motion in 3D space.

The first thing to note in the case of 3D motion, is that now the measured 2D motion signals carry ambiguity too: this is because the component of motion along the line of sight is not measurable. In order to recover a unique interpretation for the underlying 3D motion, some extraneous assumptions, or constraints, must be added in the process of computation. The perceptual observation that humans often prefer the rigid interpretation among all possible ones (when it exists) (Wallach & O'Connell, 1953) led Ullman (1979), and subsequently others, to suggest that the visual system imposes a *rigidity constraint* in order to derive a unique structure and motion interpretation. Since then, a lot of theoretical work has been done, all aimed at the task of resolving the ambiguity of the 2D motion signals for the recovery of the underlying 3D motion which produced them. However, to decide between the 1D and 2D approaches, it is not sufficient to show that 3D global motion computation *can* be performed using 2D motion cues. Rather, it is necessary to demonstrate that a similar computation would *not* be possible (or would be much more difficult) based on 1D motion clues. Here, we provide such evidence.

In Section 2 we prove that the task of discriminating rigid from non-rigid 3D motion cannot be performed on the basis of any number of straight contours and their instantaneous velocities (or their views in two frames). This is because, the ambiguity inherent in the 1D motion signals is so great that for *any* (say, random) choice of contours and velocities, there exist many rigid-motion interpretations—in fact, an infinite number of them. We will subsequently refer to the nature of computations that could, in principle, result in recovering the 3D structure and motion from 1D motion information alone: these embody higher-order temporal calculations which couple the 1D motion information from prolonged measurements.

In Section 3, we discuss the relation of these results to human perception. In an accompanying paper (Rubin, Hochstein & Solomon, 1995) we report that human observers perform poorly in tasks which probe 3D motion perception when given an image which contains many 1D motion cues, but no or impoverished 2D motion cues. These experimental observations support our computational analysis and conclusions about the necessity of 2D motion detectors when they are taken together with other recent psychophysical data which show that the human ability to perform higher order temporal calculations for global motion perception is very limited.

## 2. FORMAL RESULTS

In this section, we present the formal results pertaining to the main question raised in the Introduction, namely, can the 1D approach be extended to the general case of motion in 3D? We present the results using the formalism of position measurements in discrete frames. From a strictly theoretical point of view, this formalism is more general since in fact it encompasses both the case of apparent motion and that of smooth motion, because the position measurements may be taken as dense as desired. However, all the results below could be re-derived in a formalism where the primary measurements are the velocity vectors of the instantaneous velocity field (Rubin, 1993).

We will be using throughout the term "a straight contour". By that we mean one which has no identifiable points (such as an endpoint, dashes etc.) on it, or that these identifiable points are not given to the system (e.g. because of lack of reliable 2D detectors). Thus, in what follows we will concentrate on what is and what is not mathematically possible to compute based solely on 1D motion information. We consider here pure rotations in 3D space. The treatment of translation can be done separately (e.g. Heeger & Jepson, 1992).

A word on conventions: we take the 2D-projection plane to be the $XY$ plane, and the viewing axis to be the $Z$ axis.

**Lemma:** Given two straight contours in the $XY$ plane, then for any choice of a 3D linear transformation (except for a set of zero measure), there exists a unique straight contour in the 3D space whose orthographic projections on the $XY$ plane before and after application of the linear transformation equal the two planar contours.

**Proof:** A straight contour in 3D space is fully characterized by its point of intersection with some plane and its direction in space. Let us choose a scheme whereby we specify the contour by its intersection with the $XZ$ plane, $(X_0,Z_0)$, and the two ratios $t_1 \equiv r_x/r_y$; $t_2 \equiv r_z/r_y$, where $(r_x,r_y,r_z)$ is a vector parallel to the contour.

The contour is therefore parametrized by the formula:

$$C=\left\{\mathbf{r}\middle|\mathbf{r}=\begin{pmatrix}X_0\\0\\Z_0\end{pmatrix}+\alpha\begin{pmatrix}t_1\\1\\t_2\end{pmatrix}\right\}$$

where $\alpha$ is a scalar that parametrizes position along the contour.

Now consider a linear transformation $\mathscr{A}$, which is specified by a $3\times 3$ matrix with elements $a_{ij}$, $i,j=1,\ldots,3$. As a result of applying $\mathscr{A}$ on $C$, it is transformed into a new straight contour, $C'$, given by:

$$C'=\left\{\mathbf{r}'\middle|\mathbf{r}'=\mathscr{A}\begin{pmatrix}X_0\\0\\Z_0\end{pmatrix}+\alpha\mathscr{A}\begin{pmatrix}t_1\\1\\t_2\end{pmatrix}\right\}$$

However, we can also parameterize $C'$ by four numbers $X_0'$, $Z_0'$, $t_1'$, $t_2'$, applying our scheme as before:

$$C'=\left\{\mathbf{r}'\middle|\mathbf{r}'=\begin{pmatrix}X_0'\\0\\Z_0'\end{pmatrix}+\alpha\begin{pmatrix}t_1'\\1\\t_2'\end{pmatrix}\right\}$$

In order for the two representations of $C'$ to be identical, two conditions must be met. (1) The vectors $(t_1',1,t_2')^{\mathrm{T}}$ and

$\mathscr{A}(t_1,1,t_2)^T$ should be equal up to some multiplicative factor $\beta$:

$$t_1'=\beta(a_{11}t_1+a_{12}+a_{13}t_2)$$

$$1=\beta(a_{21}t_1+a_{22}+a_{23}t_2)$$

$$t_2'=\beta(a_{31}t_1+a_{32}+a_{33}t_2).$$

Solving for $\beta$ in the second equation and substituting it into the other two, we get:

$$t_2'=\frac{(a_{11}t_1+a_{12}+a_{13}t_2)}{(a_{21}t_1+a_{22}+a_{23}t_2)} \quad (1)$$

$$t_2'=\frac{(a_{31}t_1+a_{32}+a_{33}t_2)}{(a_{21}t_1+a_{22}+a_{23}t_2)}. \quad (2)$$

(2) The second condition is that there exist some scalar $\alpha$ such that the vector $(X_0',0,Z_0')^T$ be equal to the vector $(\mathscr{A}(X_0,0,Z_0)^T+\alpha\mathscr{A}(t_1,1,t_2)^T)$:

$$X_0'=(a_{11}X_0+a_{13}Z_0)+\alpha(a_{11}t_1+a_{12}+a_{13}t_2)$$

$$0=(a_{21}X_0+a_{23}Z_0)+\alpha(a_{21}t_1+a_{22}+a_{23}t_2)$$

$$Z_0'=(a_{31}X_0+a_{33}Z_0)+\alpha(a_{31}t_1+a_{32}+a_{33}t_2).$$

Solving for $\alpha$ in the second equation and substituting it into the other two, and using the expressions for $t_1',t_2'$ from equations (1) and (2) above, we get:

$$X_0'=(a_{11}X_0+a_{13}Z_0)-t_1'(a_{21}X_0+a_{23}Z_0) \quad (3)$$

$$Z_0'=(a_{31}X_0+a_{33}Z_0)-t_2'(a_{21}X_0+a_{23}Z_0). \quad (4)$$

Proving the above lemma is now a straightforward matter of identifying the known and the unknown variables in equations (1)–(4): in the situation as described in the lemma, what we are given are the orthographically projected views of the 3D contour before and after applying the transformation, and the transformation $\mathscr{A}$ itself. Therefore the known variables are $X_0,t_1,X_0',t_1'$ and the $a_{ij}$s. What is unknown is the underlying depth structure of the contour, namely $Z_0$, $t_2$, $Z_0'$, $t_2'$. The problem (of specifying the unique 3D contour whose orthographic projections on the $XY$ plane before and after the application of $\mathscr{A}$ equal the two planar contours) reduces therefore to finding the unique solution to the four unknowns from equations (1)–(4). Doing this, we get:

$$Z_0=\frac{X_0'-a_{11}X_0+a_{21}t_1'X_0}{a_{13}-a_{23}t_1'} \quad (5)$$

$$t_2=\frac{a_{21}t_1t_1'+a_{22}t_1'-a_{11}t_1-a_{12}}{a_{13}-a_{23}t_1'} \quad (6)$$

$$t_2'=\frac{(a_{31}t_1+a_{32}+a_{33}t_2)}{(a_{21}+a_{22}+a_{23}t_2)} \quad (7)$$

$$Z_0'=(a_{31}X_0+a_{33}Z_0)-t_2'(a_{21}X_0+a_{23}Z_0). \quad (8)$$

There exists a set of measure zero of forbidden transformations. These are the matrices which would lead to vanishing denominators in the formulae of equations (5)–(8). Specifically, these matrices are of two families: one family is that in which $(a_{13}=a_{23}=0)$. These are rotations around the $Z$ axis. The second family of matrices is that in which $(a_{13}=a_{23}\ t_1')$ or $(a_{21}\ t_1+a_{22}+a_{23}\ t_2=0)$. These correspond to rotations in which, using Euler's notation $(tg\phi=-t_1)$ or $(tg\psi=t_1')$ (Goldstein, 1959, p. 109). In the case when the chosen axis of rotation lies in the plane of the screen, this means that axes which are perpendicular to one of the contours (in either frame) are not allowed.■

Note that the lemma implies that given a pair of contours, there exist an *infinite* number of interpretations of this pair as the 2D projections of some 3D contour before and after a linear transformation (e.g. a rigid rotation, a stretch along one of the axes, etc.). The uniqueness of the 3D contour is introduced only after a particular transformation is chosen. However, *a priori*, any linear transformation may be chosen (except those that belong to the set of measure zero).

Having proved the above lemma, we are now ready to derive from it a theorem that will bring out the extent to which 1D signals in two frames are insufficient for the recovery of any knowledge about the 3D motion.

**Theorem:** Given two sets $S_1$ and $S_2$ each containing $N$ straight contours in the $XY$ plane (and assuming the correspondence between the contours in one set to the other is known), then for any choice (up to a set of measure zero) of a rotation $\mathscr{R}$, there exist a set of $N$ contours in the 3D space whose orthographic projections before and after applying the rigid rotation $\mathscr{R}$ equal the sets of planar contours $S_1$ and $S_2$ respectively.

**Proof:** Choose the linear transformation mentioned in the above lemma to be the orthogonal transformation $\mathscr{R}$. Now consider the procedure outlined in the proof of the lemma for the determination of the unique 3D contour on the basis of its two projections and the transformation. Since this procedure can be applied to each contour separately and no coupling between different contours is introduced, it can be applied to any number $N$ of pairs of planar contours, resulting in a unique set of $N$ 3D contours which correspond to the required 3D structure. The set of rotation transformations for which such underlying 3D structure does not exist is obtained by the unification of all forbidden transformations for each pair of contours, as described in the proof to the lemma above.■

The fact that the number of contours $N$ may be as large as desired may seem very surprising at first glance. After all, intuitively one might expect that increasing the number of contours would eventually provide enough information to discriminate a set of arbitrary contours from a set consistent with rigidity, and possibly also to determine the rotation parameters in the latter case. However, it can be seen where this expectation fails by using a simple counting argument: according to equations (1)–(4) above, each additional contour contributes 4 equations and 4 unknowns. Therefore, adding more and more contours cannot help in determining the elements of the matrix $\mathscr{A}$. Consequently, it is impossible to check whether or not $\mathscr{A}$ is orthogonal, and thus discriminate rigid from non-rigid cases. Even after imposing rigidity constraints, which reduce the number of unknowns in the transformation $\mathscr{A}$ from 9 to 3, the rotation parameters remain (in the generic case) completely unfixed regardless of the number of contours.

This state of affairs is to be contrasted with the situation when information from three frames is available. Here, each additional contour contributes 8 equations (4 for each frame-to-frame transition), but only 6 unknowns—2 for each frame. If we impose rigidity constraints, the number of unknowns contributed by the transformations is 6: 3 for each frame-to-frame transition. Therefore, for four or more contours (in each frame) the number of equations exceeds the number of unknowns, thus allowing the rejection of cases inconsistent with rigidity. As for the recovery of depth information from three frames in the cases which are consistent with rigidity, these computations involve in general the solution of transcendental equations, and therefore need to be performed numerically and the solution cannot be given in closed form. A detailed account of the algorithm that leads to the numerical solution is beyond the scope of this paper (but see Spetsakis & Aloimonos, 1990). The reason for not considering in detail the case of three-frame analysis in the present study will become evident in the next section: although having a mathematical interest of their own, the three-frame computations appear to be irrelevant to the computations done by the visual system and to perceptual phenomena. To see that, we have to integrate the theoretical results obtained thus far with other theoretical and perceptual observations about 2D (point) motion, which strongly suggest that two-frame calculations are the main source of global motion information to the visual system. This will be done in Section 3.

At this point, it is constructive to bring a concrete visual demonstration of the statement expressed in the theorem above. For this purpose, we use a pair of images which should be fused stereoscopically in two different ways. Each half of the picture in Fig. 1(a) contains a (slightly*) different set of four straight contours. The theorem states that for (almost) any choice of rotation axis and rotation angle, there is an underlying 3D structure whose projections before and after the rotation equal the two images. If the chosen axis lies in the plane of the figure, we can use stereoscopic vision (instead of motion) to actually see this underlying structure: by orienting the figure such that the chosen axis-of-rotation is parallel to the (body-centered) vertical and fusing the two images, what we get (assuming the stereoscopic matching procedure is done correctly by the visual system in this case) is a 3D percept that is identical to the one that would result from applying the theorem.

Fusing the two images in Fig. 1(a) therefore produces a percept of the underlying 3D structure when choosing the vertical axis to be the axis of rotation. In Fig. 2(a), the same two halves are drawn one above the other, and therefore rotating the page by 90 deg and fusing them produces a percept of the underlying 3D structure when the chosen axis is the (page-centered) horizontal one.

The percept of 3D structure produced by these figures may not seem overwhelmingly strong. Indeed, the reader may even find it non-trivial to achieve fusion and see any depth at all without allowing for prolonged viewing periods. This is related to the fact that the results derived above have implications for stereo perception as well as for motion perception. To limit ourselves to discussing motion phenomena, however, we will take a more pragmatic approach and refer the reader to Figs 1(b, c) and 2(b, c). To facilitate fusion, dots were superimposed on the lines in Figs 1(b) and 2(b) in locations consistent with the chosen axis. In Figs 1(c) and 2(c), the dots alone are drawn. After viewing the structure in these figures, most readers will find it easier to go back to Figs 1(a), 2(a) and see the 3D structures.

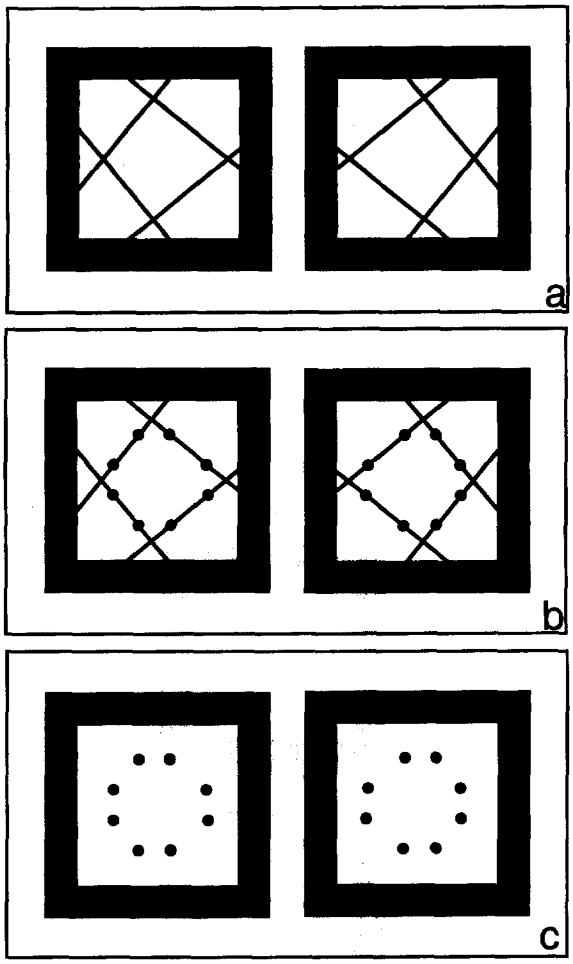


FIGURE 1. Each pair of images should be fused stereoscopically to obtain a 3D percept. In order to facilitate fusion of (a), it is recommended that the viewing order be (c); (b); (a). (a) When the two sets of contours are fused, the emerging 3D percept demonstrates the underlying 3D structure that would be obtained by applying the proof of the theorem (see text) if the chosen axis of rotation is the (page-centered) vertical. (b) Dots are superimposed on the contours in locations consistent with stereoscopic fusion when the page is viewed upright. (c) Only the dots are drawn, to decrease images disparity and facilitate stereoscopic fusion.

*Note that in principle, the theorem holds for two sets of $N$ contours that differ as much as desired: greater contour-disparity, in this case, merely implies greater depth-gradient. However, to get proper fusion, we had to demonstrate the theorem by using sets of lines that differ only slightly.

Note how different the two 3D structures in Figs 1(a) and 2(a) are: in Fig. 1(a), each pair of neighboring lines intersects in space, to form a wire-frame of four connected lines that resembles the skeleton of a kite, bent around its symmetry axis. In Fig. 2(a), the structure is one of four lines in 3D space that do not touch each other at all. Instead, their directions seem to "chase" each other in a whirlpool-like fashion. However, note that the 3D structures of Fig. 1(a, b) were obtained from *identical* two image-halves that were only oriented differently for stereoscopic fusion. This completes our visual demonstration of the statement made in the theorem, that two sets of straight contours have many possible interpretations as the 2D projections of 3D structures before and after a rigid rotation, and that these interpretations may differ from one another significantly.

So far, we have referred to the possible *rigid* interpretations of the given time-varying image. However, the lemma derived above allows us also to deduce the existence of different classes (each of an infinite number of members) of *non-rigid* interpretations consistent with the time-varying image: for one, since the transformation $\mathscr{A}$ in the lemma was not constrained to be orthogonal but only linear, we may choose $\mathscr{A}$ such that it introduces stretching (or shrinking) of differing degrees along the major axes. Second, since as noted before, the procedure of determining the 3D structure on the basis of the 2D-projected image and the chosen transformation can be carried out for each contour independently, we may choose an altogether different transformation for each of the contours—a different axis of rotation, a different rotation angle, different scaling factors for the major axes, or any combination of these variables.

To summarize, we have shown that contrary to the case of frontoparallel motion, where local 1D motion information alone can be used to compute the correct global motion, in the generic case when the object viewed moves in 3D space, the instantaneous 1D velocity field is not sufficient to recover the object's global motion. Roughly speaking, this arises from the fact that now the ambiguity inherent in contour motion allows for an infinite number of (very different) depth interpretations for the same time-varying image, even after the introduction of specific constraints such as requirement for a rigid interpretation. In particular, it was shown that for (almost) any arbitrary choice of a 3D rotation, there exists a 3D structure which would produce the 2D-projected images given in two frames (or, alternatively, the image and its instantaneous velocity field). Therefore, when given such a time varying image, the visual system cannot use the 1D velocity measurements to compute the axis of rotation. Furthermore, the mere discrimination between rigid and non-rigid motion is not possible in this case, thus eliminating the use of a strategy whereby first consistency with rigidity is checked and then the structure and motion are computed under a rigidity assumption.

## 3. RELATION TO PERCEPTION

It has long been known that human observers readily detect the 3D structure and motion of objects from their time-varying 2D-projected image, even in the absence of other depth cues (Wallach & O'Connell, 1953 and references therein; see also Sperling, Landy, Dosher & Perkins, 1989 and references therein). In order to assess the relative contribution of 1D vs 2D motion cues to the

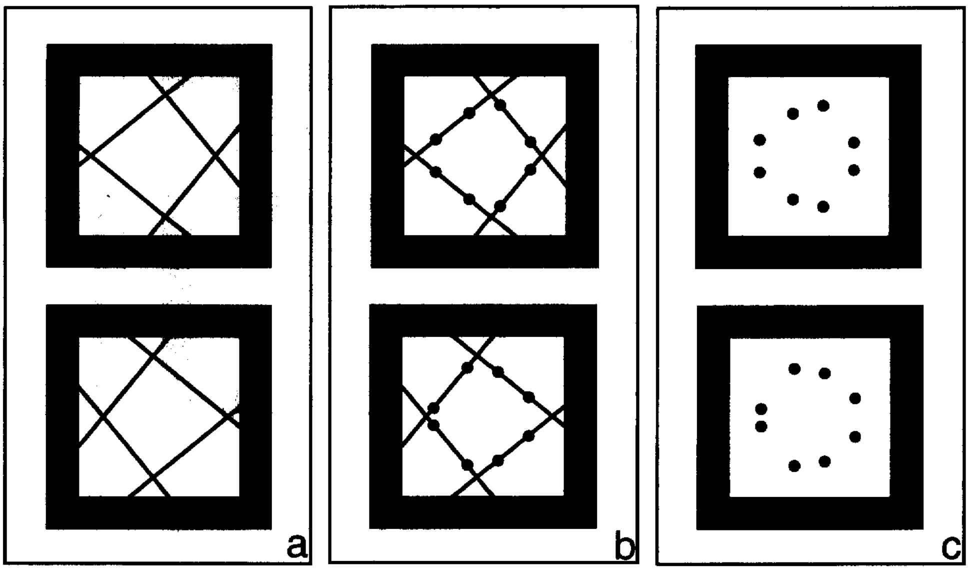


FIGURE 2. The pairs of images from Fig. 1(a) are drawn here one above the other (instead of alongside). After rotating the page by 90 deg, each pair of images should be fused stereoscopically to obtain a 3D percept. Again, recommended order of viewing is (c); (b); (a). (a) When the two sets of contours are fused, the emerging 3D percept demonstrates the underlying 3D structure that would be obtained by applying the proof of the theorem (see text) if the chosen axis of rotation is the (page-centered) horizontal. (b) Dots are superimposed on the contours in locations consistent with stereoscopic fusion when the page is rotated by 90 deg. (c) Only the dots are drawn, to decrease images disparity and facilitate stereoscopic fusion.

computations leading to this stable 3D percept, we have presented subjects with time-varying images of 3D wire-frames made of several connected bars that rotated about a fixed axis. However, we have eliminated or substantially weakened the 2D motion cues in these images, using various techniques (Rubin *et al.*, 1995). Under these conditions, subjects' performance in global motion detection tasks deteriorated dramatically. Poor performance was found not only in tasks of discriminating rigid from non-rigid motion, but also in tasks as simple as discriminating rotation about a vertical axis from rotation about a horizontal axis (both in the plane of the screen). These results hold even for prolonged observations, comprising as many as 80 consecutive frames and sweeping up to 32 deg in rotation. Observers report that these images do not elicit a notion of movement in depth at all.

While the results of Section 2 would fully explain these observations for apparent motion sequences made of two frames, the relation of the theoretical results to the more general conditions of prolonged viewing need further elaboration. From a purely mathematical point of view, the visual system could use the 1D motion information acquired from prolonged viewing (three frames or more) to perform the required tasks.

Nevertheless, taken together with recent results from studies on the perception of point motion (see below), the theoretical results presented in Section 2 are at the heart of understanding the relative role of 1D vs 2D motion cues for the formation of the global motion percept.

To explain this statement, we need first to go back to the theoretical knowledge about the computation of 3D structure and motion based on 2D motion signals. To recover the full Euclidean structure of an object and its corresponding motion, a minimum of three orthographically-projected views (frames) of at least four object points is needed (Ullman, 1979). It may therefore seem that the situation is similar to the case of contour motion described above. However, this initial impression is misleading: this is because, although the *full* description of structure and motion cannot be recovered, a substantial amount of information *can* be inferred from only 2 frames when 2D signals are given.

We will summarize below what is the information that can be recovered from 2 frames (or, equivalently, the instantaneous velocity field) of 2D cues. For the derivation of these mathematical results see Bennet, Hoffman, Nicola and Prakash (1989) and Koenderink and van Doorn (1991). First, a discrimination can be made as to whether the motion is consistent with a rigid rotation or not. This means that the vast majority of pairs of images (generated say, by some random procedure) will be rejected by such a discrimination process. Recall that such a "correct rejection" process on the basis of only two frames cannot be carried out on the basis of 1D motion information alone.

But one can extract yet more information from the two frames when 2D motion measurements are given: once consistency with rigidity has been established, it is possible to determine the overall shape of the object up to a stretch and a shear by some factor along the line of site. With respect to the motion of the object, considerable information can be recovered also: denote the axis of rotation by a vector $(\omega_x,\omega_y,\omega_z)$, where the direction of the vector parallels the axis of rotation, and its magnitude equals the rotation angle. On the basis of point-location in two frames (or the instantaneous 2D velocity field) one can determine the value of $\omega_z$, and the value of $\omega_x/\omega_y$. This means that although the axis of rotation cannot be fully recovered, the direction of its projection on the plane of the screen is determined already from two frames, as is the magnitude of rotation around the line of sight. (Choosing a particular scale for the $Z$ direction, in return, will determine the values of $\omega_x,\omega_y$ and thus the axis of rotation and the angle of rotation.)

How do these theoretical results stand in relation to human perception of depth from motion? This question was studied experimentally by a number of authors, and their results lead to the conclusion that the visual system indeed takes advantage of the possibility to extract 3D information from two-frame point-motion sequences. Braunstein, Hoffman and Pollick (1990) showed, that human observers can discriminate two-frame apparent motion sequences of 4–6 points that are consistent with rigid 3D motion from sequences that are not. Lappin, Doner and Kottas (1980), Braunstein *et al.* (1990), Todd and Bressan (1990), and others, note that such two-frame apparent motion sequences lead to a strong percept of structure and motion in depth. It is particularly important to note that, although the Z scale cannot be determined (mathematically) from only two views (see above), this ambiguity is *not* experienced perceptually. Rather, viewers generally report perceiving a scene with well-defined depth!

To reconcile these seemingly contradictory facts, we propose below three hypotheses which together describe the computational approach taken by the visual system for global motion computation. The first hypothesis states that:

● **Hypothesis 1:** *The visual system constructs a* completed 3D percept *already from first-order temporal computations. In this process of global motion computation, 2D motion cues are used, since (mathematically) it cannot be performed on the basis of 1D cues alone.*

The term "completed" is used in Hypothesis 1 because, as noted above, not *all* of the 3D information can be recovered from two-frame sequences. Thus, the missing information has to be "filled-in", or in fact *guessed* by the visual system. We suggest that:

● **Hypothesis 2:** *In generating the* completed 3D percept, *the visual system determines the scale of the object in the Z direction according to one or more*

*more of a set of heuristic rules. Two examples of such rules* are:*

*— the Z scale is chosen to be of the same order of magnitude as the scale in the X,Y directions;*

*— if available, perspective cues are used to deduce the structure of the object and thus determine the Z scale.*

So far we have talked about the use of information obtained from only two frames. What about the subsequent frames—what is their role in determining the final percept?

When viewing long sequences of apparent motion (or prolonged smooth motion) and perceiving an underlying 3D structure and motion, our notion usually, is that the large number of frames plays a significant role in establishing our stable 3D percept. However, we must ask *what exactly* is the contribution of these additional frames? What we claim, is that all existing psychophysical data can be accounted for—and in fact in some cases can be explained only (see below)—by assuming that the contribution of subsequent frames is in fact fairly limited:

● **Hypothesis 3:** *The use of motion information arriving after the* completed 3D percept *is obtained (i.e. information from third frame and beyond) is limited to the* re-validation *of this 3D percept. If the information in subsequent frames is consistent with it, this percept will be further strengthened. If not, it will lead to a breakdown of this percept and the process will begin de novo from Hypothesis 1.*

We propose that Hypotheses 1–3 encompass all the stages that lead to a 3D percept of structure and motion from a time varying 2D image. Note, that this is a description of the *computational strategy* taken by the visual system. It is far from being a full description of all the processes involved: we do not refer to the specific nature of the neural computations undertaking these stages. But the computational approach determines the goals—and limitations—of what the algorithm and implementation must achieve. In particular, Hypothesis 3 implies that the visual system is unable to perform the higher-order temporal calculations needed for completing the missing structure and motion information on the basis of third-frame measurements. This hypothesized limitation should be contrasted with the models suggested thus far which assume that such high-order temporal calculations are in fact performed by the visual system (Ullman, 1983; Landy, 1987; Rieger & Lawton, 1985; Heeger & Jepson, 1992; for a more complete exposition of the existing models and the presentation of an approach congenial to ours see Todd and Bressan (1990) and Koenderink and van Doorn, 1991).

Before moving on to review the evidence, an important point should be clarified. The heart of our claim is that the visual system relies heavily on first-order temporal calculations in depth from motion perception. This is *not* the same as saying that it can only carry out computations on images presented successively (in time): given, for example, an apparent motion sequence whereby each image is very close to its preceding one, the visual system may still perform the first-order temporal computations of the kind discussed in Hypotheses 1 and 2 on two images that are, say, five frames apart, thus increasing the apparent motion signal and getting a better signal-to-noise ratio. Similarly, when viewing smooth motion, the instantaneous velocity field that will serve as signal for the necessary first-order temporal calculations will probably be an approximation obtained from some finite temporal extent rather than the actual infinitesimal velocity field.

Two questions must be answered when examining the validity of a particular theoretical approach. Firstly, can it account for the existing experimental evidence? Secondly, can it account for experimental evidence which was not understood before? For Hypotheses 1–3, the answer to both questions is positive.

To answer positively the first question, one must show that there are no existing data that *contradict* the theory. In our case, this would imply direct evidence for the ability of human observers to perform tasks that could not be performed on the basis of first order temporal calculations and possibly a following stage of re-validation. To this end, we refer the reader to Todd and Bressan (1990). They scan the literature and show that the structure and motion information contained in two-frame sequences is sufficient for the performance of all the psychophysical structure-from-motion tasks. When use of the information from subsequent frames (third and beyond) for re-validation or rejection of the two-frame percept is taken into account, the approach we suggest can account also for results obtained more recently (Norman & Todd, 1993).

The results presented in Braunstein *et al.* (1990), in particular, provide direct support to the notion of *re-validation* proposed in Hypothesis 3: they analyze the observed increase in performance of a rigid/non-rigid discrimination task as a function of number of frames, and show that this observed increase is due exclusively to a decrease in the "false-alarms", where a false-alarm is defined as a case where the stimulus is inconsistent with rigidity but the subject mistook it for a rigid one. In other words, the subjects were able to use the subsequent frames to reject cases where the motion information in these frames was inconsistent with their (erroneous) rigid percept based on two frames, but were not able to use this subsequently obtained information to increase their "hit" rate (which was below 80% correct). Weinshall (1992) reports that in a task of determining surface curvature from motion, performance does not improve when the number of frames is increased from two to four. Todd and Bressan (1990) also discuss in detail some experimental results which at first glance seem to suggest otherwise (Hildreth, Grzywacz, Adelson & Inada, 1990), and show that these results can be accounted for by the confounding effects of the duration of the stimulus or the extent of angular rotations between two frames.

Moving to answer the second question, we first describe the (so far unaccounted for) phenomenon: this is a class of objects whereby motion sequences which *are* mathematically consistent with a rigid rotation of the

*For further discussion of various heuristics used by observers see Braunstein (1976).

object, are perceived by human observers as highly non-rigid and distorting.

Adelson (1985) writes that if a set of identical 3D objects are stretched along the $Z$ (viewing) axis to varying degrees, then while rotating, "some members of this family look rigid, but others look quite non-rigid". He then goes on to state that "the effect is not simply due to static depth cues, since it works with figures that have weak static depth (e.g., pseudo-random wire-frame figures, or rigid constellations of random dots)". We have replicated these results, and observed that the notion of non-rigidity comes into play in the cases when, as a result of stretching (or shrinking) the previously-isotropic object along the $Z$ axis, its extension along the $Z$ axis was markedly different from its $X, Y$ scale (by a factor of 3 or more). In these cases, even prolonged viewing of many complete 360 deg rotation cycles, and the top-down knowledge that these are in fact rigid objects, did not help: an overwhelming sense of distortion reigns.

The explanation we propose for these observations goes as follows. The heuristic rules that serve to add the missing information and to obtain a completed-3D-percept from a pair of frames, fail in these cases, and lead to *erroneous* percepts—of objects whose $Z$ scale is similar to the observed $X, Y$ scales. The next frames, therefore, supply information that is in conflict with these erroneous percepts, and subsequently lead to the breakdown of the rigid interpretation and the perceptual experience of a distorting image. Indeed, it is our (as yet informal) observation, that in the demonstrations noted above, the perceptual notion of non-rigidity and distortions becomes *more*, not less, pronounced with longer motion sequences, that sweep a larger overall rotation angle. Lastly, as suggested by Hypothesis 3, the discrepancy between the "completed-3D-percept" and the information obtained from subsequent frames leads to a breakdown of the rigid interpretation, and not to repeated readjustment of the perceived 3D structure, till arriving to the correct one (as would be the case in existing models of structure-from-motion).

To summarize, Hypotheses 1–3 suggest that in computing 3D global structure and motion, the visual system relies heavily on computations performed on local motion signals obtained from two-frame motion sequences. In doing so, the visual system by necessity uses 2D local motion signals, since this strategy could not be used with 1D motion information alone. This implies, that local detectors that reliably signal 2D motion in the image must exist in the early stages of visual processing.

## 4. DISCUSSION

We have reported here some mathematical results which relate to the recovery of global object motion from local 1D motion measurements. We showed that 1D signals alone cannot subserve 3D global motion computation if only first-order temporal calculations are available (based on two-frame motion or instantaneous velocity field). This means that it is not possible to recover the underlying 3D structure and motion by constructing a 3D generalization of what has been termed (in the case of frontoparallel motion) a "velocity space construction". Rather, some form of localization information is needed. If the system can perform higher than first-order temporal calculations, then 1D motion signals alone can in principle be used to recover 3D structure and motion information.

We compared the results to previous theoretical results which use 2D signals. In this case, a lot of information can be recovered already from two-frame motion sequences. Based on these results and on related perceptual observations, we suggested a computational approach which outlines the strategy used by the visual system in recovering 3D structure and motion. This proposed strategy relies heavily on the results of computations obtained from two-frame motion sequences, and therefore 2D local motion signals must be used in this process of global motion computation.

Our results provide support for the 2D approach, since they suggest that the contribution of 2D detectors is critical for the computation of global motion. We should note, however, that the nature of the 2D motion cues used in the global motion computation need not necessarily be confined *only* to the detection of point and endpoint motion. Sections of high curvature in a contour may also be used, at some level, as sources of unambiguous motion information (see e.g. Philip & Fisichelli, 1945; Mulligan, 1992; Braunstein & Andersen, 1984). In fact, it has been suggested that endpoint detecting mechanisms may be closely related to curvature detecting mechanisms (Dobbins, Zucker & Cynader, 1987, 1989).

Finally, an important point should be clarified. Our claim that reliable 2D signals must exist in the visual system should not be taken to imply that the 1D signals play no role in global motion perception. Rather, what we claim is that 1D signals *alone* cannot account for the observed characteristics of human global motion perception. The contribution of identifiable points, i.e. of 2D cues, is essential for the overcoming of a deep ambiguity inherent in contour motion. The "aperture problem", which is a term used in the context of the ambiguity inherent in the *rigid* motion of straight contours, is in fact only one manifestation of the more general "correspondence problem" inherent in contour motion: since, when viewing non-rigid (such as elastic) moving objects, knowledge of the motion of one point on the curve does not lead to knowledge of the motion of other points along it (see Wallach, 1976, Chap. IX, Part 2 for a more complete discussion of this issue). However, once this ambiguity has been resolved by the detection of identifiable points along (or near) smoothly curved contours, the 1D signals arising from the contour motion—which are abundant in images in general and are more robust to image noise—*do* participate in the determination of the global motion percept. The question of how the visual system combines the information from 2D motion cues with 1D motion signals to produce a coherent global motion percept is an important one for further research.

## REFERENCES

Adelson, E. H. (1985). Rigid objects that appear highly non-rigid. *Investigative Ophthalmology and Visual Science (Suppl.)*, *26*, 56.

Adelson, E. H. & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*, 523–525.

Anderson, S. J. & Burr, D. C. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research*, *25*, 1147–1154.

Anderson, S. J. & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *Journal of the Optical Society of America A*, *8*, 1330–1339.

Anderson, S. J., Burr, D. C. & Morrone, M. C. (1991). Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. *Journal of the Optical Society of America A*, *8*, 1340–1351.

Bennet, 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*, 1052–1069.

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

Braunstein, M. L. & Andersen, G. J. (1984). A counterexample to the rigidity assumption in the visual perception of structure from motion. *Perception*, *13*, 213–217.

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

Castet, E., Lorenceau, J., Shiffrar, M. & Bonnet, C. (1993). Perceived speed of moving lines depends on orientation, length, speed and luminance. *Vision Research*, *33*, 1921–1936.

Derrington, A. M. & Badcock, D. R. (1992). Two-stage analysis of the motion of 2D patterns, what is the first stage? *Vision Research*, *32*, 691–698.

Derrington, A. M. & Suero, M. (1991). Motion of complex patterns is computed from the perceived motions of their components. *Vision Research*, *31*, 139–149.

Dobbins, A., Zucker, S. W. & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, *329*, 438–441.

Dobbins, A., Zucker, S. W. & Cynader, M. S. (1989). Endstopping and curvature. *Vision Research*, *29*, 1371–1387.

Fennema, C. L. & Thompson, W. B. (1979). Velocity determination in scenes containing several multiple moving objects. *Computer Vision, Graphics and Image Processing*, *9*, 301–315.

Ferrera, V. P. & Wilson, H. R. (1990). Perceived direction of moving two-dimensional patterns. *Vision Research*, *30*, 273–287.

Ferrera, V. P. & Wilson, H. R. (1991). Perceived speed of moving two-dimensional patterns. *Vision Research*, *31*, 877–894.

Goldstein, H. (1959). *Classical mechanics*. New York: Addison-Wesley.

Heeger, D. J. (1988). Optical flow using spatiotemporal filters. *International Journal of Computer Vision*, *1*, 279–302.

Heeger, D. J. & Jepson, A. D. (1992). Subspace methods for recovering rigid motion. I: Algorithm and implementation. *International Journal of Computer Vision*, *7*, 95–117.

Hildreth, E. C. (1984). *The measurement of visual motion*. Cambridge, Mass.: 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. & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, *8*, 377–385.

Landy, M. S. (1987). Parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A*, *4*, 864–877.

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.

Lorenceau, J. & Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research*, *32*, 263–273.

Lorenceau, J., Shiffrar, M., Wells, N. & Castet, E. (1993). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, *33*, 1207–1217.

Marr, D. (1982). *Vision*. New York: Freeman.

Mingolla, E., Todd, J. T. & Norman, J. F. (1992). The perception of globally coherent motion. *Vision Research*, *32*, 1015–1031.

Mulligan, J. B. (1992). Anisotropy in an ambiguous kinetic depth effect. *Journal of the Optical Society of America A*, *9*, 521–529.

Nakayama, K. & Silverman, G. H. (1988). The aperture problem—II. Spatial integration of velocity information along contours. *Vision Research*, *28*, 747–753.

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

Orban, G. A. (1984). *Neuronal operations in the visual cortex*. Berlin: Springer.

Philip, B. R. & Fisichelli, V. R. (1945). Effect of speed of rotation and complexity of pattern on the reversals of apparent movement in Lissajous figures. *American Journal of Psychology*, *58*, 530–539.

Rieger, J. H. & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A*, *2*, 354–359.

Rubin, N. (1993). The use of one-dimensional and two-dimensional motion information for the computation of global motion. Ph.D. thesis, Hebrew University, Israel.

Rubin, N. & Hochstein, S. (1993). Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional images. *Vision Research*, *33*, 1385–1396.

Rubin, N., Hochstein, S. & Solomon, S. (1995). Restricted ability to recover 3D global motion from 1D motion signals: Psychophysical observations. *Vision Research*, *35*, 463–476.

Sereno, M. I. (1989). Learning the solution to the aperture problem for pattern motion with a Hebb rule. *Advances in Neural Information Processing Systems*, *2*.

Sereno, M. E. (1993). *Neural computation of pattern motion: Modelling stages of motion analysis in the primate visual cortex*. Cambridge, Mass.: MIT Press.

Shiffrar, M. & Pavel, M. (1991). Percepts of rigid motion within and across apertures. *Journal of Experimental Psychology, Human Perception and Performance*, *17*, 749–761.

Shimojo, S., Silverman, K. & Nakayama, K. (1989). Occlusion and the solution to the aperture problem. *Vision Research*, *29*, 619–626.

Sperling, G., Landy, M. S., Dosher, B. A. & Perkins, M. E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology, Human Perception and Performance*, *15*, 826–840.

Spetsakis, M. E. & Aloimonos, J. (1990). Structure from motion using line correspondence. *International Journal of Computer Vision*, *4*, 171–184.

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

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Mass.: MIT Press.

Ullman, S. (1983). Maximizing rigidity: The incremental recovery of 3D structure from rigid and non-rigid motion. *Perception*, *13*, 255–274.

Wallach, H. (1935). Uber visuell wahrgenommene bewegungsrichtung. *Psychologische Forschung*, *20*, 325–380.

Wallach, H. (1976). On perceived identity. In Wallach, H. (Ed.), *On perception*. New York: Quadrangle.

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

Weinshall, D. (1992). Shortcuts in shape classification from two images. *Computer Vision, Graphics and Image Processing*, *56*, 57–68.

Welch, L. (1989). The perception of moving plaids reveals two motion-processing stages. *Nature*, *337*, 734–736.

Yo, C. & Wilson, H. R. (1992). Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research*, *32*, 135–147.

Zetzsche, C. & Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research*, *30*, 1111–1117.


*Acknowledgements*—Support for this study was provided by the U.S.-Israel Binational Science Foundation and by the Hebrew University Center for Neural Computation. We thank Margalit Harmats, Shimon Ullman, Hendrik van Veen, Daphna Weinshall and Peter Werkhoven for helpful discussions. Special thanks to Robert Shapley for many helpful discussions and for invaluable help in reviewing the manuscript.