Vision Res. Vol. 28, No. 6, pp. 747–753, 1988
Printed in Great Britain. All rights reserved

0042-6989/88 \$3.00 + 0.00
Copyright © 1988 Pergamon Press plc

THE APERTURE PROBLEM—II. SPATIAL INTEGRATION OF VELOCITY INFORMATION ALONG CONTOURS

KEN NAKAYAMA and GERALD H. SILVERMAN
Smith-Kettlewell Institute of Visual Sciences, 2232 Webster St, San Francisco, CA 94115, U.S.A.

(*Received* 12 *December* 1986; *in revised form* 24 *August* 1987)

Abstract—There exists a class of two-dimensional figures (including cumulative gaussian waveforms) whose contours have a limited range of orientations. These figures can appear as highly nonrigid if they undergo pure translation in the image plane. In the case of the cumulative gaussian waveform, it is the region around the inflection point that appears as nonrigid. Motivated by Hildreth's (1984) proposal, we made 5 predictions which were confirmed by the data: (0), rigidity of a figure can be dramatically increased if one attaches line terminators to the figure; (1), moving terminators "on" the figure increase rigidity far more than such terminators "off" the figure; (2) decreasing the velocity of the terminator decreases rigidity; (3) decreasing the distance between the terminator and the inflection point increases rigidity; (4) the effect of a moving terminator can be blocked by interposing a stationary terminator between it and a nonrigidly moving portion of the curve.

Motion Velocity Aperture problem Contour

INTRODUCTION

Numerous authors have noted that a simple local reading of velocity is insufficient to recover the true motion in an image (Wallach, 1935; Fennema and Thompson, 1979; Burt and Sperling, 1981; Horn and Schunk, 1981; Marr and Ullman, 1981; Adelson and Movshon, 1982). In theory, just two local readings are sufficient to reconstruct the true velocity field if the figure is undergoing pure translation in the image plane. This follows from the fact that any pair of locally defined constraint lines will intersect at the true velocity of motion in a velocity space. See Fig. 2 of our companion paper (Nakayama and Silverman, 1988).

For the case of general motion, consisting of translational, rotational and deformational motion, there is often no unique solution. This can be seen most clearly by referring to a figure adapted from Hildreth (1984) showing that the motion of a given point P between two successive frames is indeterminate (see Fig. 1).

The difficulties in recovering this velocity field is an example of the more general class of ill-posed problems outlined by Poggio *et al.* (1985). Such problems require specific assumptions about the environment and the corresponding optic array. Then and only then can one construct an algorithm to arrive at a candidate solution.

With respect to the reconstruction of velocity, we look for additional assumptions regarding the nature of the underlying field of velocity vectors. Horn and Schunk (1981) suggested that the velocity field might be assumed to vary gradually. Without further constraint, such a process would require the minimization of all spatial derivatives of motion (dV_x/dx, dV_x/dy, dV_y/dy, and dV_y/dx). Sharp discontinuities in the velocity field are common, however and this class of algorithms would smooth over areas of such discontinuity. Such smoothing is at odds with human perceptual observations where motion defined boundaries are seen as very sharp (Hildreth, 1984; Nakayama and Silverman, 1984).

Hildreth suggested an alternative approach which had greater ecological plausibility and received support from perceptual demonstrations. She suggested that local differences in velocity was indeed integrated but only along contours, proposing the following energy-like term which needed minimization.

$$\Theta(\mathbf{V}) = \int |d\mathbf{v}/ds|^2\, ds \qquad (1)$$

where **V** is a candidate set of velocity vectors along the contour and $\Theta(\mathbf{V})$ is a scalar associated with this set. Hildreth's algorithm computes the velocity field which minimizes

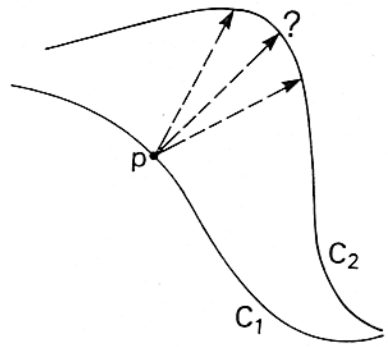

Fig. 1. Two successive frames of a stimulus undergoing general motion (a combination of translation, rotation, and deformation). Under these circumstances the velocity of point P cannot be specified from any succession of images (redrawn from Hildreth, 1984).

differences in velocity along contours yet is consistent with contraints dictated by the aperture problem. In our opinion, the idea had two attractive features. First it copes more effectively with the problem of discontinuity by integrating along and not between contours. Second, it is in rough qualitative accord with a number of perceptual illusions where various figures such as ellipses undergo rotation in the image plane and are seen as nonrigid (Hildreth, 1984).

For the case of pure translational motion in the image plane, Hildreth's formulation is particularly simple since the minimization function can be reduced to zero. Thus Hildreth's approach subsumes previous treatments of the aperture problem (Fennema and Thompson, 1979; Adelson and Movshon, 1982) as a special case. Both predict that all translating figures should appear as rigid.

From our companion paper (Nakayama and Silverman, 1988), it should be evident that these ideas cannot be completely correct. We found that planar figures undergoing pure translational motion can be seen as highly nonrigid (Nakayama and Silverman, 1988). Figure 2 shows three different figures which are seen as nonrigid as they are translated vertically.

We argued that under some conditions not favorable for the computation of the true velocity field, the visual system has a tendency to default towards the local orthogonal velocity and this can lead us away from a rigid interpretation (Nakayama and Silverman, 1988). To incorporate this tendency into the present discussion, we suggest the addition of a second term to the Hildreth minimization scheme. It might be expected to have the following general form

$$\Theta(V) = \int \{|dV/ds|^2 + F[(\phi, |\mathbf{V}_L|)]\}\, ds \quad (2)$$

Where ϕ is the absolute angular deviation of the candidate velocity vector from the perpendicular local component, V_L is the magnitude of the local velocity vector, and $F(\phi, \mathbf{V}_L)$ is a monotonically increasing function of ϕ and $abs(\mathbf{V}_L)$ which summarizes the weak and graded tendency of the system to default towards the orthogonal motion. The influence of this hypothesized second term can be isolated conceptually by considering the special case of a straight line moving behind a circular aperture. Even though an infinite set of rigid interpretations are possible, this display is seen as a line moving orthogonal to its own orientation. An inspection of equation (2) provides a plausible interpretation. The $|dV/ds|^2$ term in equation (2) is zero for all rigid interpretations and thus makes no contribution to the solution. The orthogonal term $f(\phi, |\mathbf{V}_L|)$, however, varies for different directions of motion and correctly predicts the orthogonal motion that is seen.*

The candidate velocity field which minimizes the smoothness term is generally not the same field of velocities that minimizes the orthogonal tendency term. The scheme suggested by equation (2) indicates that the system reaches a compromise, minimizing the sum of these scalar energy terms. In the remainder of the paper, we report a number of experiments based on predictions derived from this modified form of Hildreth's thesis. Before outlining these predictions we describe our methods which are common to all experiments.

METHODS

A vertically oriented cumulative gaussian waveform [as depicted in Fig. 2(C)] was flashed on, moved for 100 msec at a velocity of 2.5 deg/sec, and then abruptly disappeared. This short duration was used to prevent any significant tracking by the oculomotor system. When presented alone, all subjects saw the

*Although this second term does seem to be necessary to account for some of these qualitative observations, we would not like to leave the impression that it is any more than a provisional way to reconcile perceived nonrigidity with Hildreth's smoothness notion. No formal mathematical work has been conducted, for example, to see whether the addition of such a term would result in a unique or computable solution.

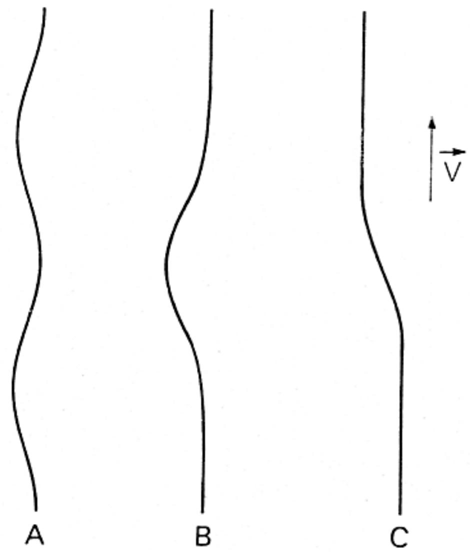

Fig. 2. Examples of three planar curves which can appear as highly nonrigid when they undergo vertical translational motion. (A) sinewave, (B) gaussian waveform, (C) cumulative gaussian or error function (erf).

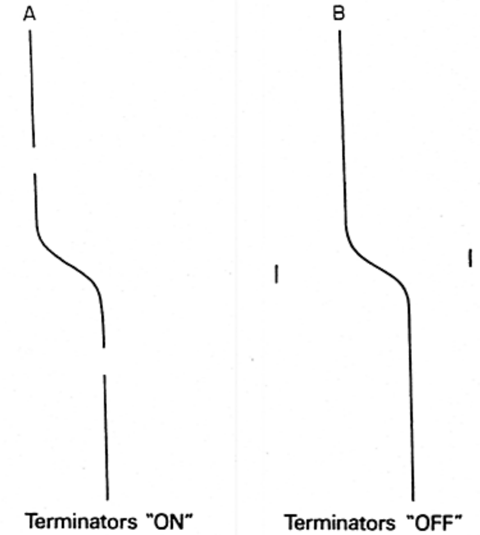

Fig. 3. Two experimental arrangements of the terminators. In (A), the terminators are on the line. In (B), they are off the line.

figure as highly nonrigid. For the experiments reported here, the standard deviation of the cumulative gaussian was 0.2 deg of visual angle. Thus it was about 0.8 deg in vertical extent. Amplitude or offset of the waveform was 1.0 deg of visual angle. In all other respects the display was the same as that described in our companion paper (Nakayama and Silverman, 1988).

To test some of the ideas proposed, we added moving "terminators" to the image. Such terminators, defined as the end of a line segment, have essentially all orientations and contain sufficient local information to solve the aperture problem.† So in distinction to moving local contours, the motion of terminators is not ambiguous. Adding a moving terminator is analogous to injecting a fixed and highly defined velocity at selected regions of the image plane. As such, we used the terminator as a probe to see how it influenced the perception of motion at other regions of the image.

Two different layouts of terminators were employed. We could put a pair of moving terminators "on" the line which appeared as gaps [see Fig. 3(A)] or we could place the pair of terminators "off" the line [see Fig 3(B)] where they consisted of short line segments having the same length as the gaps. The critical distance between a given terminator and the inflection point could also be varied. In the case of the gaps these distances were in the vertical direction and in the horizontal direction for the line segments. These gaps or line segments could be made to move independent of the movement of the cumulative gaussian waveform. Thus they could remain stationary, move faster, or slower than the moving waveform.

We generated this display on a Hewlitt Packard 1332A monitor oscilloscope. To generate the cumulative gaussian waveform, we used a custom one-dimensional video buffer which enabled us to present a continuous waveform with no visible digital quantization artefacts. Custom electronic hardware was also constructed to generate the motion of the gaps and the line segments. Frame rate was 100 Hz.

To measure the amount of perceived rigidity, we used a rating scale system. Observers were presented with an interleaved series of conditions and were asked to give a number which characterized their perception of nonrigidity. We always presented the highly nonrigid waveform without terminators as a standard to anchor the scale near one end and relied on the observers sense of rigidity to anchor it at the other. To deal with the issue of experimenter bias, we replicated all of the major results with naive observers who had no knowledge of the nature or purpose of the experiment. The differences in perceived rigidity under the different experimental conditions were very

†The two-dimensional Fourier spectra of a punctate line terminator contains widely dispersed energy in the Fourier plane. Thus different mechanisms sensitive to different orientations will be stimulated and provide sufficient local information to solve the aperature problem (see Adelson and Movshon, 1982).

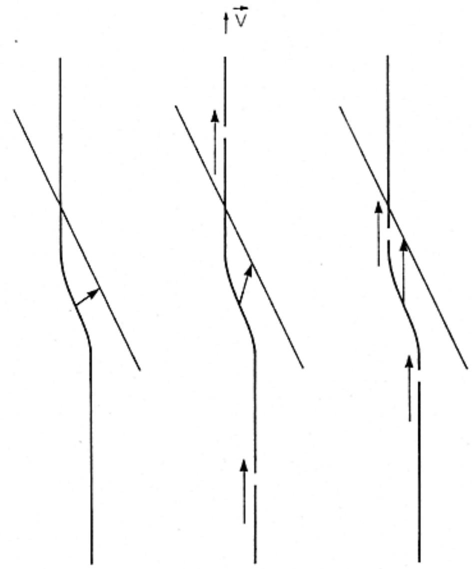

Fig. 4. In each diagram the moving waveform is depicted along with the constraint line corresponding to local velocity at inflection point. V refers to the velocity of the figure. (A) No terminator exists and we suggest that the orthogonal tendency at the inflection point keeps its corresponding velocity vector to a nearly perpendicular orientation along its constraint line. As a consequence we predict that the figure would appear as highly nonrigid. (B) A moving terminator somewhat distant from the inflection point is added. This creates a finite $|dV/ds|^2$ term which drives the vector at the inflection point further out on its own constraint line. This figure should look more rigid than in (A). (C) Same as in (B) except the terminator is even closer. As a consequence, the figure should appear as even more rigid.

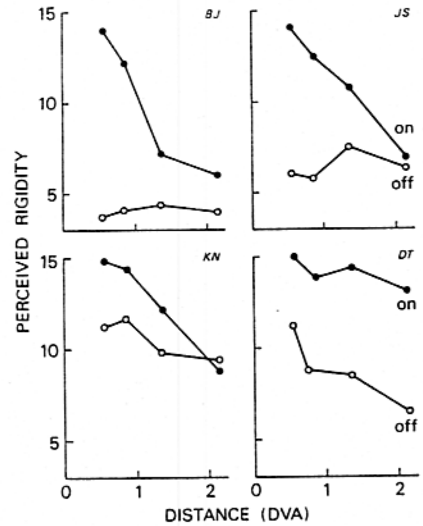

Fig. 5. Perceived rigidity ratings plotted as a function of distance for terminators which are on the line (solid dots) and for terminators which are off the line (open circles). Note that terminator and waveform velocity were equal and set to 5 deg/sec. Data taken for 4 observers. A rating of 5 corresponds to the perceived rigidity of the moving figure without terminators and a rating of 15 corresponds to the case where the stimulus looks unequivocally rigid.

large and observers required essentially no practice to make ratings. No systematic differences in the results were seen between the naive observers (B.J, D.T.) and ourselves (K.N., J.S.).

THEORETICAL PREDICTIONS AND RESULTS

Prediction No. 1: Rigidity should increase as one decreases the distance from the terminators and the inflection point

First let us consider the moving waveform alone without terminators [see Fig. 4(A)]. In this situation we suggest that the orthogonal term overshadows the $|dV/ds|^2$ term. As such the candidate velocity vector at the inflection point is on its constraint line but is misaligned with the true movement of the figure. Thus, nonrigid motion is predicted. Imagine, however, that we switch "on" a pair of moving terminators at some distance from the inflection point [as seen in Fig. 4(B)]. The presence of these terminators drives up the $|dV/ds|^2$ term which counteracts the orthogonal tendency. The system will reach equilibrium by aligning the local vector more closely with the true motion of the waveform, resulting in an increased perception of rigidity. This will happen to an even greater extent for terminators which are even closer to the inflection point [see Fig. 4(C)]. The alignment will be better and the figure can be expected to appear even more rigid. In sum, there should be a distance effect such that the closer terminators should increase rigidity more than distant terminators.

To test this idea, we varied the distance of a pair of terminators which moved at the same velocity as the waveform itself and obtained rigidity judgements for different distances. The results of 4 observers is seen as the solid circles in Fig. 5. Note that in support of the hypothesis, there is a decreasing perception of figural rigidity as distance is increased.

Prediction No. 2: Rigidity should increase only if a terminator is on the contour

Hildreth's model only admits for integration along the contour. No provision is made for integration between contours. As a consequence, we should expect that figural rigidity should increase only for the case of the gaps which are on the line [see Fig. 3(A)] and not for the case of the line segments which are off the moving waveform [see Fig. 3(B)].

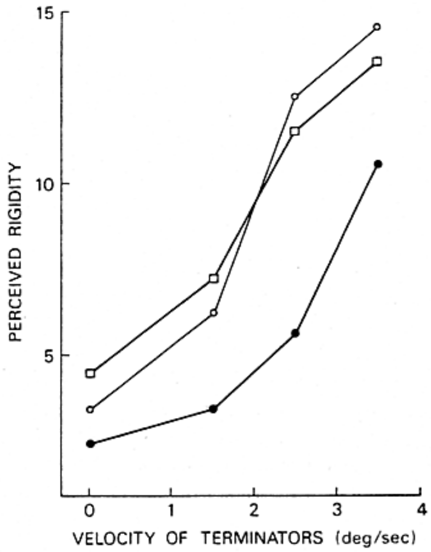

Fig. 6. Rigidity ratings for 3 observers as a function of velocity of the terminators. Terminators were "on" the line [as in Fig. 3(A)] and were 1.4 deg of visual angle from the inflection point. Cumulative gaussian waveform was on for 100 msec and had a velocity of 2.5 deg/sec. Solid circles (B.J.), open squares (K.N.), open circles (J.S.).

To test this idea we measured the perceived rigidity as a function of the distance of the terminator, both for the case where it is on the line and the case where it was off the line. Distance is defined as the distance between the nearest part of the terminator and the inflection point.

Our results provided a striking though partial confirmation of the prediction. The open circles in Fig. 5 show the data from instances where the terminators were off the line. It is clear that the effect on figural rigidity is far less than when the terminator is on the line and this can be seen for all four observers. That some effect is evident for terminators off the line, however, indicates that some form of integration must occur across contours although in weakened form.

Prediction No. 3: Figural rigidity should decrease as one decreases the velocity of the moving terminators

The reasoning here is very similar to prediction No. 1. By decreasing the velocity of the terminators, one reduces the size of the $|dV/ds|^2$ term and thus permits the orthogonal term to become more dominant.

Figure 6 shows that there is a strong dependence on velocity. Thus decreasing velocity has a large effect on the perception of rigidity, as predicted.

Prediction No. 4: Rigidity induced by a remote set of terminators can be "blocked" by a non-moving terminator interposed between the moving terminator and the inflection point

In order for velocity information to be passed from one portion of a contour to another, intermediate points along the contour must assume intermediate velocity values because the driving force behind the Hildreth algorithm is the reduction of velocity differences. It follows that if one were to prevent this from occurring, velocity information from a remote region would not propagate towards the center of a figure. Thus, the effect of a moving terminator in making the inflection point look more rigid could be blocked if we were to inject a zero velocity point between the terminator and the inflection point. A stationary terminator provides such a fixed velocity of zero. Thus we would predict that if we were to make an inflection point look more rigid by adding a remote moving terminator, this could be blocked if a stationary terminator were interposed between the moving terminator and the inflection point.

To see if this was the case we asked a number of observers ($n > 10$) to see whether the target became more rigid when a moving terminator was added to the figure. As expected, all reported an increase in figural rigidity. Then to test the prediction, we placed a stationary pair of terminators halfway between the moving terminators and the inflection point. All observers reported that the figure again looked highly nonrigid. This is consistent with our expectation that the stationary terminator should block the propagation of velocity information from the moving terminator.

DISCUSSION

Given that all of our observations were made *after* considering the implications of Hildreth's theory, we were surprised by the general agreement between theory and data. Of particular interest are predictions Nos 2 and 4 which we would not have ordinarily made without having Hildreth's notions in mind. Since other interpretations may offer at least a partial explanation of the data, we consider some of them in turn.

First we note that the "on" vs "off" the contour condition is subtly confounded with a difference related to shearing and compressive components of motion (see Nakayama *et al.*, 1985 for a discussion of compression and shearing motion). In the case where the terminators are "on" the contour and thus aligned along the vertical, the rigid percept insures that there are no vertical compressive components or horizontal shearing components. The case of the terminators off the line would insure that there is no vertical shearing components or horizontal compressive shearing components. Recent neurophysiological evidence indicates the existence of specialized velocity detectors for curl and divergence (Saito *et al.*, 1986). Thus there is a possibility that a smoothness algorithm, based on the selective detection and integration of divergence $(\mathrm{d}V_x/\mathrm{d}x + \mathrm{d}V_y/\mathrm{d}y)$ or curl $(\mathrm{d}V_x/\mathrm{d}y - \mathrm{d}V_y/\mathrm{d}x)$ could be of importance.

Second we note a phenomenon of motion capture originally discussed by Ramachandran and colleagues. They found that if a set of fine random dots was replaced by another set of random dots, "illusory" coherent motion of the dots could be seen if the replacement was accompanied by the motion of a low spatial frequency sinusoid which was superimposed on the display (Ramachandran and Inada, 1985). Aside from the very different nature of the inducing and test motions, one other difference between motion capture and the present phenomenon is evident. With motion capture, the inducing motion is essentially distributed over the same retinal area as the induced dots whereas in our case they are in spatially distinct retinal regions.

Third, one can ask whether the results might be better understood in terms of spatial frequency filtering at the earliest stages. For example, suppose the receptive fields for motion were sufficiently large to encompass the terminators and the region of the inflection point for shorter and not so for longer terminator distances. Could this explain the distance effect and possibly eliminate the need to postulate a second or higher stage of motion processing? This alternative explanation is flawed, however, because it fails to account for the large difference between the "on" vs "off" the line condition. In fact given the configuration of our experiment and what is known about the spatial frequency tuning of receptive fields, it is possible to have just the opposite expectation—that rigidity should be greatest when the terminators are off to the side.

Consider the fact that the contrast detecting aspects of the receptive fields for motion are anisotropic as revealed both by electrophysiological (Hubel and Wiesel, 1962) and psychophysical observations (Nakayama *et al.*, 1985). Thus the long axis of the receptive field is oriented orthogonal to the preferred direction of motion. This means that in the present experimental configuration velocity should be better integrated for the case where the terminators are off to the side as this corresponds to the long axis of a hypothetical receptive field. As such, a conception based on the early spatial frequency filtering properties of motion units predicts a result which was exactly opposite to that obtained.

Given that the Hildreth model handles these results and that other analogies or alternatives are less convincing, what can be said for its neurophysiological plausibility? As we consider known mechanisms we see very little that is available. First there is no direct physiological evidence that velocity signals or any other signals are selectively propagated along contours. There is indirect psychophysical evidence for excitation traveling over areas as is the case of color filling-in (Yarbus, 1967) but this is a spread over a bounded area and not along a contour. Furthermore color spreading also has no plausible physiological substrate. Second, we come to the concept of velocity in a surrounding neighborhood influencing the responsiveness of a local area. Such interactions exist but they are usually inhibitory not excitatory, emphasizing differences between surround and center (Frost and Nakayama, 1983; Allman *et al.*, 1985). In our case we are looking for an excitatory linkage between surround velocities and those in the center, perhaps analogous to the linking of co-linear line fragments seen by Nelson and Frost (1985) in cat cortex or van der Heydt *et al.* (1984) in monkey area 18. Such interactions have been hypothesized by Grossberg and Mingolla (1985).

So what is missing in the present formulation is a plausible linkage to known physiological mechanisms. It remains to be seen whether it will be replaced by a theory based on current physiological notions or whether the identification of new physiological processes will be required.

CONCLUDING REMARKS

We have shown that a variety of targets can appear as highly nonrigid, even if they are undergoing simple translational motion in the image plane. These examples provide additional counterexamples to the supposition that the visual system automatically defaults to a possible rigid interpretation. It also provides a set of examples which could not be accounted for by previous models (Fennema and Thompson, 1979; Adelson and Movshon, 1983; Hildreth, 1984).

To handle these new results, we postulate the existence of an orthogonal tendency [embodied in equation (2)] in which the absence of competing signals will bias the interpretation of the velocity field in favor of motion orthogonal to the local contour. In order to overcome the erroneous interpretation that could result from such a tendency, two conditions must be met. First, which we described in our companion paper (Nakayama and Silverman, 1988), is the need for the local velocity signals to come from a sufficiently large range of orientations to counteract the inherent noise associated with the reading of velocity signals. Otherwise the intersection of constraint lines will be too imprecise. Second, and this has been described in the present paper, the differently oriented velocity signals required to solve the aperture problem have to be in sufficiently close proximity and also to be on the same contour for a solution to be reached.

Acknowledgements—Supported in part by grants 5P30 EY-01186 and 1RO1 EY-05408 from the National Institutes of Health and The Smith-Kettlewell Eye Research Foundation. A preliminary abstract of this paper was presented as: Silverman G. H. and Nakayama K. (1984) Propagation of velocity information along moving contours. *J. opt. Soc. Am.* **73,** 56. Author's address: 2232 Webster Street, San Francisco, CA 94115, U.S.A. We thank Steven Chung for the design and construction of specialized electronic hardware.

REFERENCES

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

Allman J., Miezin F. and McGuiness E. (1985) Direction and velocity-specific responses from beyond the classical receptive field in the middle temporal area (MT). *Perception* **14,** 105–126.

Burt P. and Sperling G. (1981) Time, distance, and feature trade-offs in visual apparent motion. *Psychol. Rev.* **88,** 171–195.

Fennema C. L. and Thompson W. B. (1979) Velocity determination in scenes containing several moving images. *Comput. Graph. Image Proc.* **9,** 301–315.

Frost B. J. and Nakayama K. (1983) Single unit neurons code opposing motion independent of direction. *Science, N.Y.* **220,** 744–745.

Grossberg S. and Mingolla E. (1985) Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* **92,** 173–211.

Hildreth E. C. (1984) The measurement of visual motion. *ACM Distinguished Dissertation* Series) MIT Press, Cambridge, Mass.

Horn B. K. and Schunk B. G. (1981) Determining optical flow. *Artif. Intell.* **17,** 185–203.

Hubel D. H. and Wiesel T. N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol., Lond.* **160,** 106–154.

Marr D. and Ullman S. (1981) Directional selectivity and its use in early visual processing. *Proc. R. Soc. B* **211,** 151–180.

Nakayama K. and Loomis J. M. (1974) Optical velocity patterns, velocity sensitive neurons and space perception: a hypothesis. *Perception* **3,** 63–80.

Nakayama K. and Silverman G. (1984) Temporal and spatial aspects of the upper displacement limit in random dots. *Vision Res.* **24,** 293–299.

Nakayama K., Silverman G., MacLeod D. I. A. and Mulligan J. (1985) Sensitivity to shearing and compressive motion in random dots. *Perception* **14,** 225–238.

Nakayama K. and Silverman G. H. (1988) The aperture problem—I. Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Res.* **28,** 739–746.

Nelson J. I. and Frost B. J. Contour illusions in the visual cortex.

Poggio T., Torre V. and Koch C. (1985) Computational vision and regularization theory. *Nature, Lond.* **317,** 314–319.

Ramachandran V. S. and Inada V. (1985) Spatial phase and frequency in motion of random-dot patterns. *Spatial Vision* **1,** 57–67.

Saito H., Yukie M., Tanaka K., Hikosaka K. Fukada Y. and Iwai E. (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J. Neurosci* **6,** 145–167.

Yarbus A. L. (1967) *Eye Movements and Vision.* Plenum Press, New York.