Anisotropy in an ambiguous kinetic depth effect

Jeffrey B. Mulligan

Aerospace Human Factors Research Division, MS 262-2,
Ames Research Center, National Aeronautics and Space Administration, Moffett Field, California 94035

Received December 31, 1990; revised manuscript received October 28, 1991; accepted October 29, 1991

A set of animated stimuli (Lissajous figures), each element of which is physically consistent with two different three-dimensional shapes undergoing rigid rotations about orthogonal axes, is described. Human observers typically show a preference for one shape or the other; this preference may be biased by manipulating various parameters of the stimulus. Fairly good predictions of which shape will be seen are made by using an adaptation of Hildreth's smoothest-velocity-field computation. When a given stimulus is rotated 90° in the picture plane, the resolution of the ambiguity is often different, demonstrating anisotropy in the processing of the figures. The nature of this bias is such that for certain figures subjects see a three-dimensional object rotating about a vertical axis regardless of which two-dimensional orientation is used to present the stimulus. This bias is not predicted by the Hildreth model. One interpretation of the results is that the ambiguity in two-dimensional visual motion (i.e., the aperture problem) is not resolved before the interpretation of the three-dimensional structure.

1. INTRODUCTION

Human perception of three-dimensional shape from a sequence of two-dimensional images is an extraordinary feat that is often taken for granted. The visual system must compute both a depth and a third component of velocity for each point in the scene and do so in a way that produces overall consistency. An early observation of the recovery of three-dimensional structure from a sequence of two-dimensional silhouettes was reported by Miles[1]; Wallach and O'Connel[2] later dubbed the phenomenon the kinetic depth effect. For the case of rigid rotations of discrete points, Ullman[3] determined the conditions that must be satisfied for the solution to be computationally realizable. Ullman's more recent models[4] allow for some departure from rigidity, as might arise either from actual nonrigid motion or from noise in the input.

The complete problem becomes more difficult if one tries to extend this type of approach to images that are composed of line segments and smooth curves (not to mention gray-level images) instead of just isolated dots. This increased difficulty is due to the aperture problem: A motion detector viewing a moving line through a small aperture will be blind to motions of the line along its own length and will therefore be unable to report the actual two-dimensional image velocity within its field of view.

The aperture problem complicates the analysis of structure from motion because algorithms such as Ullman's require the two-dimensional image velocity at each point, while sensors with a small field of view can report only the local orientation and the orthogonal velocity. Many models of the structure-from-motion problem assume that estimates of the two-dimensional velocities are available as input.[3-7] One elegant approach has been proposed by Hildreth[8,9] for figures composed of closed curves. She suggested that a useful way to attack the problem would be to try to minimize the amount of variation in the hypothetical image velocities. Hildreth investigated a number of possible ways to define variation, but in most of her work she used the squared magnitude of the vector difference between the velocities of adjacent points on the curve. By summing the local variation over the entire figure, one may obtain a single number for a given hypothetical solution of the aperture problem. Calculus may be used to find the solution that minimizes this quantity; Hildreth called this solution the smoothest velocity field. In this paper I use the term roughness to denote the variational quantity that is minimized.

One objection to Hildreth's approach is that, although the solution of the aperture problem is subsequently used to analyze structure from motion, the aperture problem is approached without regard for possible three-dimensional constraints. Intuitively, we find it more information efficient to consider the ultimate three-dimensional interpretation when resolving the two-dimensional ambiguity, instead of being locked into a two-dimensional solution that could possibly be inconsistent with three-dimensional rigidity.

Hildreth has shown qualitatively that the algorithm makes mistakes in cases in which humans also experience illusory percepts, such as in the barber-pole illusion. The experiments described in this paper were performed in an attempt to assess the validity of Hildreth's algorithm as a description of human perception. The approach was to use a stimulus that had two distinct physically plausible three-dimensional interpretations; each interpretation corresponded to a different solution of the aperture problem. These ambiguous stimuli violate one of the conditions of Hildreth's algorithm, namely, that any intersections of the curve in the image should correspond to actual intersections of the three-dimensional generator. For this reason, initially no attempt was made to solve the minimization problem and find the smoothest velocity field. However, the roughness of a given hypothetical solution is still a well-defined mathematical quantity. Since it is known that when humans are presented with the ambiguous stimuli they usually see one of the two rigid interpretations, I simply computed the roughnesses for the two

0740-3232/92/040521-09$05.00

rigid interpretations; if the visual system solves the aperture problem by using a calculation such as roughness, then the subjects should perceive the interpretation with the lower roughness value.

2. METHODS

A. Lissajous Figures

The ambiguous stimuli were drawn from a class of curves known as Lissajous figures. Each figure has two distinct three-dimensional interpretations that correspond to quite different shapes. Each of the two interpretations is a three-dimensional curve that lies on the surface of a cylinder, but the orientation of the cylinder is different for each of the two interpretations.

The three-dimensional appearance of animated Lissajous figures was observed some time before the term kinetic depth effect was coined.[10] Philip and Fisichelli subsequently investigated the effects of various parameters on the spontaneous depth reversals in Lissajous figures.[11,12] These depth reversals are yet another ambiguity in the figures (similar to the depth reversals seen with the Necker cube), which is quite distinct from the ambiguity that is the topic of this paper; if one counts depth reversals as well as shape differences, there are a total of four possible interpretations. From these early reports it is impossible to determine whether the investigators were even aware that the two interpretations with different rotational axes were physically consistent with a single stimulus. For example, Fisichelli[12] describes changing the axis of rotation by interchanging the cables providing the deflection signals to the cathode-ray tube. The parameter values used in these studies, however, were ones that produce an extremely strong bias in favor of one of the two interpretations, so perhaps it was simply a matter of chance that spontaneous changes of rotational axis were not observed.

Before giving the explicit formulas describing Lissajous figures, it may be illustrative to consider the problem of depicting an unambiguous curve that lies on the surface of a cylinder. Imagine that we have a vertically oriented cylinder of unit radius and that we wish to paint on the surface of this cylinder a curve whose vertical position is defined as a function of the angular position $y = f(\theta)$. (I adopt a coordinate system in which x and y are the normal viewing screen coordinates, with z being a depth axis.) In this case I describe the curve parametrically by the following equations:

$$x(\theta) = \sin(\theta), \tag{1a}$$

$$y(\theta) = f(\theta), \tag{1b}$$

$$z(\theta) = \cos(\theta). \tag{1c}$$

The curve defined by these parametric equations will lie on the surface of a unit-radius cylinder regardless of the nature of the function $f(\theta)$. If we interchange the definitions of x and y, then we obtain a curve that lies on a horizontal cylinder. This observation is the key to understanding the ambiguity of Lissajous figures: If we let $f(\theta) = \sin(\theta)$, then the above equations are symmetric in x and y, and the resulting curve lies on both cylinders.

This is illustrated graphically in Fig. 1. The top row shows the situation of a generic function $f(\theta)$ painted onto a vertical cylinder. The leftmost panel shows the cylin-

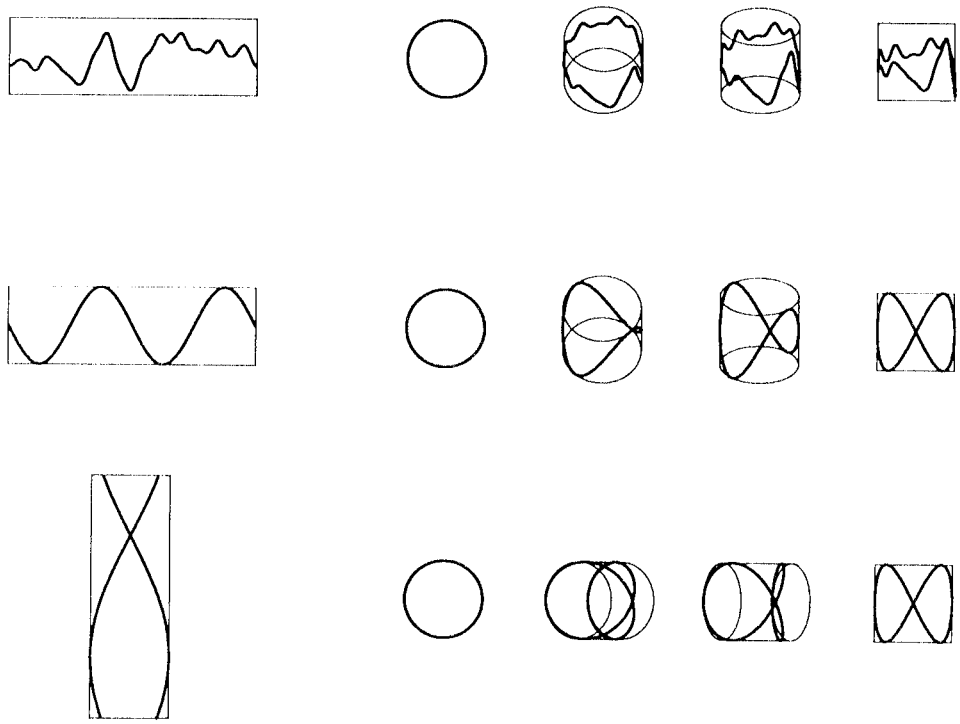

Fig. 1. Illustration of the three-dimensional ambiguity of Lissajous figures. The top row depicts a random waveform (shown at the left) that is rolled into a cylinder. Several views of the cylinder are shown. In the second row, the random waveform is replaced by two cycles of a sinusoid. When this curve is rolled into a cylinder and viewed from the side (far right), the parametric equations describing the projected curve are symmetric in x and y (with the exception of the frequency parameters), implying that the projected curve could equally well lie on a horizontally oriented cylinder. The curve that generates the same projection when rolled into a horizontal cylinder is shown in the bottom row. Although only a single phase is shown, this ambiguity remains when the cylinders are rotated.

der unrolled [i.e., a plot of $f(\theta)$]. The successive panels show orthographic projections of the cylinders from a number of viewpoints, ending with the side view. The second row shows the same process applied for the special case of $f = \sin(2\theta)$. In the third row, the function $f = \sin(\theta/2)$ is wrapped onto a horizontal cylinder. Note that the xy projections (the rightmost panels of the second and the third rows) are identical. Relatively unambiguous views of the two shapes are obtained in the intermediate rotations shown in Fig. 1. Note that there are no three-dimensional self-intersections in the saddle-shaped curve shown in the second row, while for the pretzel-shaped curve depicted in the third row the intersection in the final projection corresponds to a self-intersection of the three-dimensional curve.

The three-dimensional curves that project to Lissajous figures are described by the following sets of parametric equations:

Case 1 (curve lies on the vertical cylinder):

$$x_V(\theta) = A_x \sin(\omega_x \theta), \tag{2a}$$

$$y_V(\theta) = A_y \sin(\omega_y \theta + \phi_y), \tag{2b}$$

$$z_V(\theta) = A_x \cos(\omega_x \theta). \tag{2c}$$

Case 2 (curve lies on the horizontal cylinder):

$$x_H(\theta) = A_x \sin(\omega_x \theta + \phi_x), \tag{3a}$$

$$y_H(\theta) = A_y \sin(\omega_y \theta), \tag{3b}$$

$$z_H(\theta) = A_y \cos(\omega_y \theta). \tag{3c}$$

The frequency parameters ω_x and ω_y must be integers for the curve to close on itself as θ runs from 0 to 2π. Changing the phase (ϕ_x or ϕ_y) corresponds to rotating the cylinder about its axis.

We can see the equivalence of the projected curves in these two cases by making the following substitutions:

$$\theta = \theta' + \phi_x/\omega_x, \tag{4a}$$

$$\phi_y = -\phi_x \frac{\omega_y}{\omega_x}. \tag{4b}$$

By substituting these values into Eqs. (2a) and (2b), we easily see that

$$x_V(\theta') = x_H(\theta), \tag{5a}$$

$$y_V(\theta') = y_H(\theta). \tag{5b}$$

The z function is irrelevant since we assume orthographic projection onto the xy plane.

To animate the figures, let the phase [ϕ_x in Eq. (3a)] be a function of time:

$$\phi_x(t) = 2\pi\omega_t t, \tag{6}$$

where the parameter ω_t represents the angular velocity in revolutions per unit of time. Note that, from Eq. (4b), the angular velocities in the two interpretations differ by the ratio of the frequencies of the generating functions. Figure 2 depicts a few frames of a sequence, together with oblique views of the two possible shapes. An example showing the velocities associated with each of the two interpretations is shown in Fig. 3.

B. Psychophysical Procedures

The stimuli were presented on a cathode-ray tube (Tektronix Model 611). Signals for the X and the Y deflections were produced by digital–analog converters, or DAC's (ADAC Models 1023AD and 1023EX), under the control of a PDP11-23 computer. The x and the y gains of the display scope were carefully adjusted to provide the same spatial displacement for a given DAC increment, thereby correcting for any possible gain differences between the two DAC's.

The DAC's incorporated a direct memory access controller (ADAC Model 1620DMA), which permitted lists of coordinate pairs to be transferred rapidly from memory. After each pair of coordinates was transferred, the interface generated a brief pulse that was used for the Z (brightness) input to the cathode-ray tube. This was a transistor–transistor logic pulse that had a duration of ap-

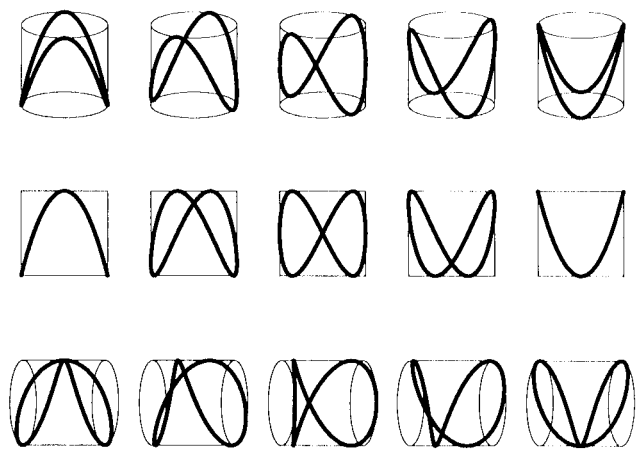

Fig. 2. Series of possible stimulus frames together with oblique views of the two possible generating shapes. The upper row shows slightly oblique views of the corresponding vertical cylinder inscribed with the saddle-shaped figure, while the lower row depicts views of a horizontal cylinder inscribed with the pretzel-shaped figure. Note that the cylinder in the upper row rotates through 90° from left to right, while the cylinder in the lower row rotates through a full 180°.

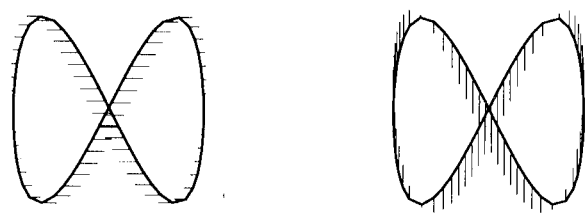

Fig. 3. Velocities associated with the two possible interpretations of an animated Lissajous figure. At the left are shown the velocities corresponding to perceived rotation about a vertical axis (merry-go-round motion). This corresponds to the saddle-shaped figure shown in the middle row of Fig. 1. Note that the velocities of the two limbs that intersect in the middle of the figure have opposite directions. At the right are shown the velocities corresponding to rotation about a horizontal axis (rolling-pin motion); this corresponds to the pretzel-shaped figure shown in the bottom row of Fig. 1. Note that where the curve intersects itself in the center the velocities match, since the intersection in the figure corresponds to an actual three-dimensional intersection in the projected figure.

proximately 1 μs. The front panel controls were adjusted to produce the maximum possible luminance. The time needed to write a single point with this apparatus was approximately 10 μs. The curves described by Eqs. (3a) and (3b) were produced by plotting 512 points at uniformly sampled values of the parameter θ. Unfortunately, this resulted in a nonuniform spacing of the points along the curve; although the point spacing was always less than the spot size (so that the curves appeared continuous), this did result in small intensity variations, inversely proportional to $\partial s/\partial\theta$, along the curve. The coordinate lists for the sequence of frames making up a single stimulus were computed (and were resident in memory) before the onset of the stimulus; the direct memory access transfers for the individual frames were initiated following interrupts from a real-time clock. The frame rate was 100 Hz. The computation of the coordinate lists for each stimulus was speeded by using tables of precomputed values of sine and cosine; before each trial, these tables were scaled by the aspect ratio factors to reduce the number of multiplications needed.

Each trial consisted of a 2-s presentation of a figure defined parametrically by Eqs. (3a) and (3b). For a given experimental condition, the frequency parameters ω_x and ω_y were fixed, but the aspect ratio A_x/A_y was varied from trial to trial. The product of A_x and A_y (and therefore the swept area of the stimulus) was held constant. The temporal frequency ω_t that determined the rotation frequencies was set so that the faster of the two rotations had a rate of 1 Hz.

After each trial the subjects were instructed to report whether the figure was perceived in rolling-pin motion or merry-go-round motion. Although the subjects almost universally reported that the stimuli appeared three dimensional, they were informed that, in the event that they did not see a three-dimensional figure, they could make the judgment on the basis of whether the two-dimensional motion was primarily up and down (rolling pin) or side to side (merry-go-round). The subjects were also instructed that, in the event that the percept changed during the course of the stimulus presentation, they should base their responses on the appearance at the end of the presentation. The subjects entered their responses by using the detached keyboard of the computer console.

It was noted in pilot experiments that elongation of the figure in one dimension tended to cause the rotation axis to be perceived in the same dimension as the elongation, i.e., large values of A_x/A_y produced a rolling-pin percept, while small values produced a merry-go-round percept. An up–down staircase was therefore used to control the selection of successive values of A_x/A_y, such that a rolling-pin response would decrease the value of A_x/A_y by a constant factor, while a merry-go-round response would increase it by the same factor. The factor used was 0.1 log unit, or approximately 1.26.

The subjects were tested under six conditions, consisting of two orientations of three pairs of values for ω_x and ω_y. These were $\omega_x = 2, \omega_y = 1$; $\omega_x = 3, \omega_y = 1$; and $\omega_x = 3, \omega_y = 2$. The remaining three conditions were obtained by simply exchanging ω_x and ω_y. Corresponding pairs of conditions were always run together. Two pairs of conditions were combined to make a block. Each of the three possible blocks was run twice, resulting in four replications of each condition. Within a block, each condition was assigned a single staircase; the trials were clustered into groups of four, consisting of one trial from each staircase. Within each cluster of trials the order of the conditions was controlled by a pseudorandom-number generator. Within each block, 50 judgments were made for each condition.

The subjects consisted of one experienced psychophysical observer (the author) and an undergraduate student who had some practice in making psychophysical judgments but who was naïve with respect to the purpose of the experiment. Additional subjects were tested in individual conditions but did not complete the experimental protocol; the (incomplete) results from these subjects were similar to those shown for the two subjects who completed the full regimen of observations. Later, an additional experienced subject (JAP) was tested with a different apparatus. This apparatus consisted of a raster graphics system with a frame rate of 60 Hz. Subject JAP completed three blocks in which all six conditions were interleaved. All the other parameters were identical to those described above.

Typical data from a single run of a single condition are shown in Fig. 4. The percentage of rolling-pin responses is plotted against the log of A_x/A_y. The raw data from each block were fitted with a cumulative normal by using a weighted least-squares fitting procedure described in detail by Mulligan and MacLeod.[13] The inflection point of the curve is located at the aspect ratio for which we would expect to receive an equal number of rolling-pin and merry-go-round responses; I refer to this aspect ratio as the critical aspect ratio (CAR). For each of the six conditions, four replications provided independent estimates

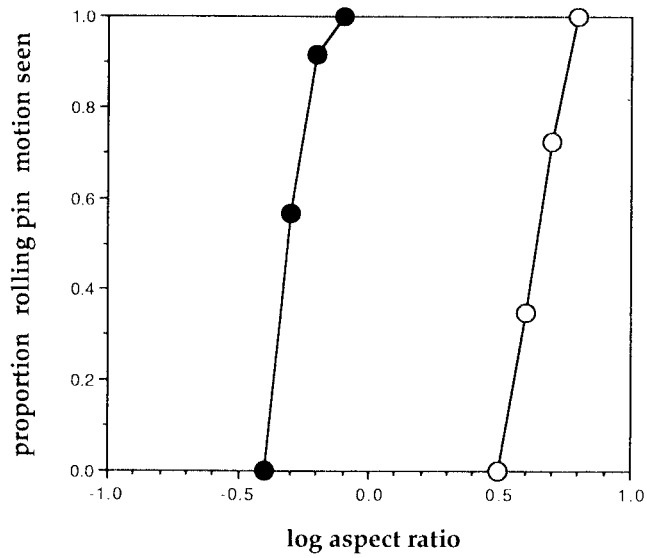

Fig. 4. Typical data from a single run of the experiment. The abscissa represents the log of the aspect ratio A_x/A_y, while the ordinate represents the proportion of responses indicating rolling-pin motion seen. The open circles are for the condition $\omega_x = 2, \omega_y = 1$, while the filled circles are for the dual condition $\omega_x = 1, \omega_y = 2$. Each curve represents 50 judgments collected with a single staircase. Raw data such as these were fitted with a cumulative normal to estimate the CAR at which the two percepts were equally likely. For these data, the fitting procedure produced estimates of the CAR of −0.30 and 0.64. The inequality of the absolute values of these numbers is evidence of anisotropy or bias.

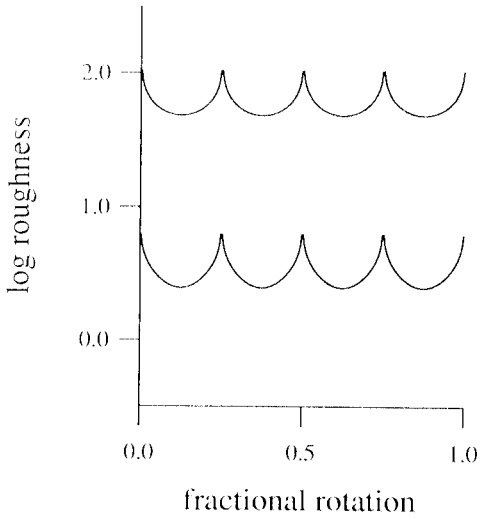

Fig. 5. Variation of the logarithms of the roughnesses $\mathbf{R}_x$ and $\mathbf{R}_y$ as a function of rotational phase, which is expressed as a fraction of 2π (one complete rotation). The plot shows the log of roughness computed at 256 uniformly spaced values of rotational phase ϕ_x. The values of the parameters were $A_x = A_y = 1, \omega_x = 1$, $\omega_y = 2, N = 1024$. The upper curve is the roughness computed for the rolling-pin interpretation, $\mathbf{R}_y$, while the lower curve is the roughness computed for the merry-go-round interpretation, $\mathbf{R}_x$. The separation of the curves indicates a strong preference of the model for the merry-go-round interpretation for this set of parameter values.

of the CAR for each subject. The fitting procedure also estimated the semi-interquartile difference, which is the change in the abscissa (log aspect ratio) required to change the response rate from 50% to 25% or 75%.

C. Smoothness Estimates

Predictions were made by using an adaptation of the computation proposed by Hildreth.[8,9] Hildreth defined a variational measure on possible two-dimensional velocity fields and solved for the velocity field that minimized this quantity, calling this the smoothest velocity field. Here I adopt a much simpler approach: Instead of finding the minimum, I simply compute the roughness of each of the two rigid solutions and assume that the visual system will prefer the interpretation having the lower value. No claim is made that either solution actually corresponds to a local minimum of the roughness function.

Following Hildreth, I used the following quantity to define roughness:

$$\mathbf{R}(\mathbf{v}) = \int \left|\frac{\partial v}{\partial s}\right|^2 \mathrm{d}s. \tag{7}$$

This integral was approximated as a discrete sum as

$$\mathbf{R}(\mathbf{v}) = \sum_{i=1}^{N} \frac{|\partial v/\partial\theta|^2}{|\partial s/\partial\theta|}\mathrm{d}\theta, \tag{8}$$

where θ is the parameter used to trace out the curves in Eqs. (2) and (3) and $\mathrm{d}\theta$ is equal to $2\pi/N$. This quantity was computed for each of the two possible interpretations: For rotations about a horizontal axis, all the velocities are in the vertical direction:

$$\mathbf{v}_y = A_y\omega_y \cos(\omega_y\theta), \tag{9}$$

$$\frac{\partial \mathbf{v}_y}{\partial\theta} = -A_y{\omega_y}^2 \sin(\omega_y\theta). \tag{10}$$

For vertical axis rotation, all the velocities are horizontal:

$$\mathbf{v}_x = A_x\omega_x \cos(\omega_x\theta + \phi_x), \tag{11}$$

$$\frac{\partial \mathbf{v}_x}{\partial\theta} = -A_x{\omega_x}^2 \sin(\omega_x\theta + \phi_x). \tag{12}$$

The remaining quantities needed to evaluate the sum in Eq. (8) are

$$\frac{\partial s}{\partial\theta} = \left[\left(\frac{\partial x}{\partial\theta}\right)^2 + \left(\frac{\partial y}{\partial\theta}\right)^2\right]^{1/2}, \tag{13}$$

$$\frac{\partial x}{\partial\theta} = A_x\omega_x \cos(\omega_x\theta + \phi_x), \tag{14}$$

$$\frac{\partial y}{\partial\theta} = A_y\omega_y \cos(\omega_y\theta). \tag{15}$$

Let us use the terms $\mathbf{R}_x$ and $\mathbf{R}_y$ to represent the roughnesses computed for motion about the vertical and the horizontal axes, respectively. These quantities depend on the amplitude factors A_x and A_y, the frequencies ω_x and ω_y, and the rotational phase ϕ_x. It turns out that the roughnesses vary as a function of rotational phase; this variation is shown in Fig. 5, where the log of the roughness is plotted as a function of phase. The phases at which the roughness measure attains a maximum correspond to the rotational positions where the front and the back sections of the generating curve project onto the same curve in the image, as occurs in the leftmost and the rightmost panels of Fig. 2.

The measure of roughness defined in Eq. (8) is affected by the total arc length, which is a function of the rotational phase. If we double both of the amplitude factors A_x and A_y, the roughnesses $\mathbf{R}_x$ and $\mathbf{R}_y$ also double (as does the total arc length). Thus we see that, if we wish to have a roughness measure that depends only on shape and not on absolute size, we might obtain this by dividing by the total arc length L:

$$L = \int_0^{2\pi} \frac{\partial s}{\partial\theta}\mathrm{d}\theta. \tag{16}$$

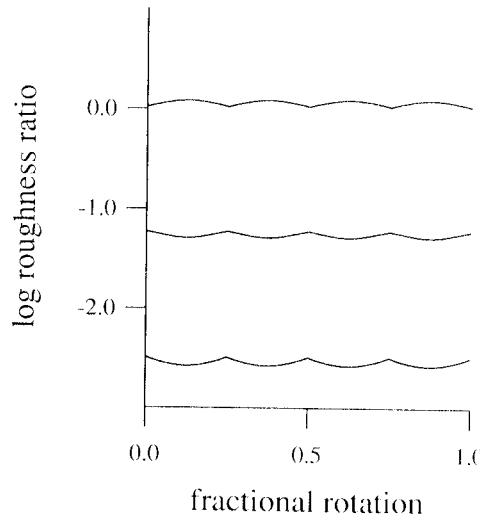

Fig. 6. Variation of log roughness ratio as a function of rotational phase. The values of the parameters were $\omega_x = 1, \omega_y = 2$, $N = 1024$. The three curves represent three different values of the aspect ratio A_x/A_y; from the upper curve to the lower curve the aspect ratios were 4.0, 1.0, and 0.25.

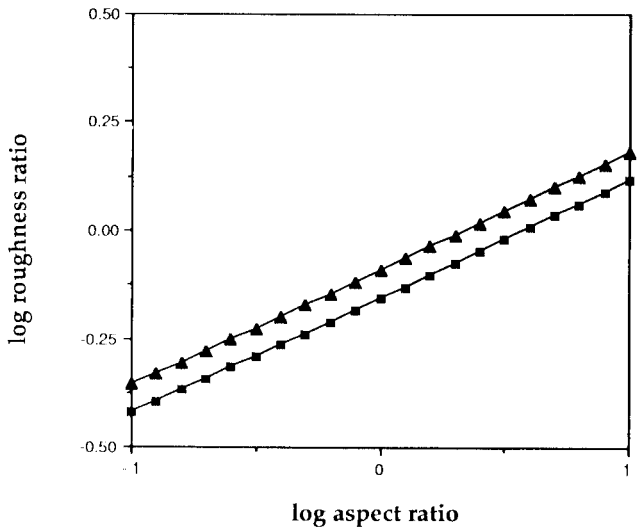

Fig. 7. Log roughness ratio integrated over rotational phase as a function of log aspect ratio. The log aspect ratio was sampled uniformly in 20 steps from −1 to 1. The values of the parameters used to generate the lower line were $\omega_x = 1, \omega_y = 2$, $N = 1024$, while the upper line represents $\omega_x = 2, \omega_y = 3$.

For the parameter values used to generate the data shown in Fig. 5, however, the arc-length variations as a function of phase are less than 5%, so the qualitative picture is not affected by this change.

Although the roughnesses vary as a function of phase, it can be seen from Fig. 5 that the ratio of the roughnesses is relatively constant, as indicated by the roughly constant vertical separation on the log ordinate. The log roughness ratio as a function of phase is plotted in Fig. 6 for different aspect ratios. Note that by taking the ratio of the roughnesses I have divided out the effect of total arc length.

A single number characterizing the relative smoothness of the two interpretations was obtained by integrating the log roughness ratio across the rotational phase. This was justified on the grounds that the ratio was relatively constant across phases and because, although subjects' percepts are bistable, the transitions do not seem to be phase locked with the rotation. Once this average ratio has been computed, we can estimate the predicted value of the CAR by solving for the aspect ratio that yields a mean log roughness ratio of 0. Because the visual system might integrate on some other transformed representation, we should be prepared to accept an error of the order of the vertical variation of the curves in Fig. 6.

Figure 7 plots the log roughness ratio as a function of the log aspect ratio for $\omega_x = 1, \omega_y = 2$ and $\omega_x = 2, \omega_y = 3$. The points represent the mean of 256 different phases sampled uniformly over the interval 0 to $\pi/4$. The sample phases were placed so as to straddle the phases at which the singularities occur, such as $\phi_x = 0$. It may be observed that the points fall close to a straight line, with a slope of 1/2. Linear regression was used to fit a line to the points; the log of the CAR was taken to be the X intercept from the regression equation. CAR's were obtained in this way for a number of pairs of frequencies (ω_x, ω_y); the log CAR is plotted against the log frequency ratio ω_x/ω_y in Fig. 8 (squares). Note that the points fall close to a line with a slope of −2. A weak explanation for this can be made from the fact that A_x appears in Eq. (12) with an exponent of 1, while ω_x appears with an exponent of 2. The graph in Fig. 8 is symmetric: The positions of the points in the upper-left-hand quadrant are simply the positions of the points in the lower-right-hand quadrant reflected through the origin. In a practical sense, this means that the shape of the figure at the CAR is unaffected if the entire figure is rotated 90°.

D. Deviations from the Smoothest Velocity Field

In Subsection 2.C predictions were made on the basis of which of the two rigid interpretations is smoother, i.e., which rigid interpretation has the lower value of the roughness measure introduced by Hildreth. In this subsection a slightly different approach is explored: I first solve for the smoothest velocity field (which does not correspond to the rigid motion of a three-dimensional figure) and then ask to which of the two possible rigid interpretations it is more similar.

The smoothest velocity field is obtained in the manner described by Hildreth. The expression for the roughness [Eq. (8)] is differentiated with respect to the (unknown) tangential component of the velocity at each sample point and then equated to zero. The result is a system of N linear equations in N unknowns, which is easily solved with standard linear algebra. Figure 9 shows the resulting velocity field for a representative figure.

A measure of similarity between two velocity fields was formed by summing the squared vector differences between the rigid velocities and the corresponding velocity vectors from the smoothest field. This quantity was computed for each of the two rigid interpretations at a number of different aspect ratios for each set of parameter values. When the log ratio of these two quantities is plotted versus the log aspect ratio, a pattern of results similar to that seen in Fig. 7 is obtained. Linear regression was used to estimate the x intercept, i.e., the CAR for which the

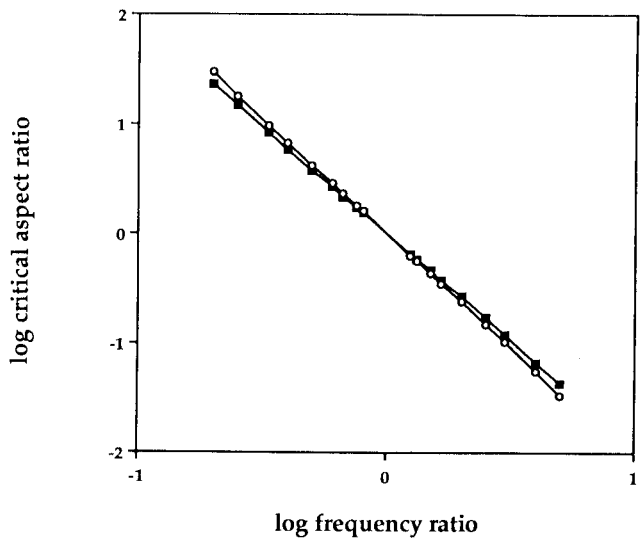

Fig. 8. Log CAR as a function of log frequency ratio ω_x/ω_y. The points sampled from the abscissa correspond to the following ordered pairs (ω_x, ω_y): $(1,5),(1,4),(1,3),(2,5),(1,2),(3,5),(2,3),(3,4),(4,5),(5,4),(4,3),(3,2),(5,3),(2,1),(5,2),(3,1),(4,1),(5,1)$. The squares indicate the predictions based on the roughness of the rigid interpretations; the circles indicate the predictions based on differences between the rigid interpretations and the smoothest velocity field.

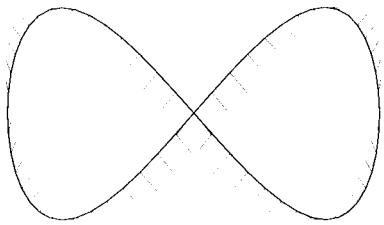

Fig. 9. Smoothest velocity field, computed by using Hildreth's algorithm.

smoothest velocity field was equally different (using the integrated squared vector difference metric) from each of the two rigid solutions. The predicted CAR's generated by this method are shown by the circles in Fig. 8, along with the predictions from the original method. Like the original predictions, these new predictions fall on a line with a slope close to -2, although this line is slightly steeper than that describing the original predictions.

3. RESULTS

The experimental results are shown in Fig. 10. The mean log CAR over the four replications is plotted as a function of the log frequency ratio for subjects LR (triangles), JBM (circles), and JAP (open squares). The small squares show the corresponding predictions from Fig. 8. The numerical data used to generate Fig. 10 are given in Table 1, along with the standard errors.

Several features of Fig. 10 are notable. First, although the data do deviate from the predictions, the overall slope of the data points is close to -2, as in the prediction, giving qualitative support to our modified Hildreth model. Second, the deviations from the predictions are exclusively upward from the predictions, indicating a bias in favor of the merry-go-round percept. The bias is evinced not simply by deviations from the predictions but from the fact that pairs of points that correspond to two orientations of a given shape are not located symmetrically with respect to the origin. Any set of data possessing central symmetry would be evidence for isotropy, regardless of how unlike the prediction it might be. The prediction does show this symmetry, since there is no anisotropy built into the model.

Although the deviations from the predictions shown in Fig. 10 look rather modest, the ordinate is a logarithmic scale, and small deviations therefore correspond to profound differences in shape. To assess the significance of the anisotropy, however, it is necessary to compare these deviations with the range of aspect ratios for which the percept is bistable. This is indicated by the transition zone of the psychometric function shown in Fig. 4, which is typical for all the subjects. In most cases the size of the anisotropy effect (the sum of the logs of the CAR's for corresponding conditions) is larger than the width of this transition zone, indicating that there are stimuli that are consistently perceived in merry-go-round motion regardless of the orientation in which they are presented.

4. DISCUSSION

A. Vertical–Horizontal Illusion

The vertical–horizontal illusion (VHI) refers to the fact that a vertical line will appear longer than a horizontal line of the same length. The details of this much-studied illusion are summarized well by Robinson.[14] Observers similarly overestimate the vertical component of motion in obliquely moving targets.[15] This suggests a simple explanation of the anisotropy observed in the data, namely, that the visual input is subjected to an affine distortion before the motion is analyzed. A deformation of the image consistent with the VHI would produce an anisotropy of the correct sign for the subjects whose data are shown in Fig. 10; if the visual input were stretched in the vertical dimension by a factor α and compressed in the horizontal dimension by the same factor, then the lines in Figs. 8 and 10 would simply be shifted horizontally by an amount $2\log(\alpha)$. We can estimate the amount of deformation needed to account for the subjects' data by calculating the horizontal shift necessary to superimpose the regression lines for the predicted and the observed results in Fig. 10. When this is done, an aspect ratio factor of 1.20 (0.079 log unit) is obtained for subject JBM, and a factor of 1.21 (0.0825 log unit) is obtained for subject LR.

B. Anisotropy in Two-Dimensional Apparent Motion Correspondence

Gengerelli[16] demonstrated an anisotropy of two-dimensional apparent motion correspondence with the stimulus illustrated in Fig. 11(a), sometimes referred to as a bistable quartet. The stimulus consists of two pairs of luminous dots that are flashed in alternation. The two pairs are located at opposite corners of a rectangle. When the pairs are alternated in time, several percepts are possible: The two dots that are visible at any given

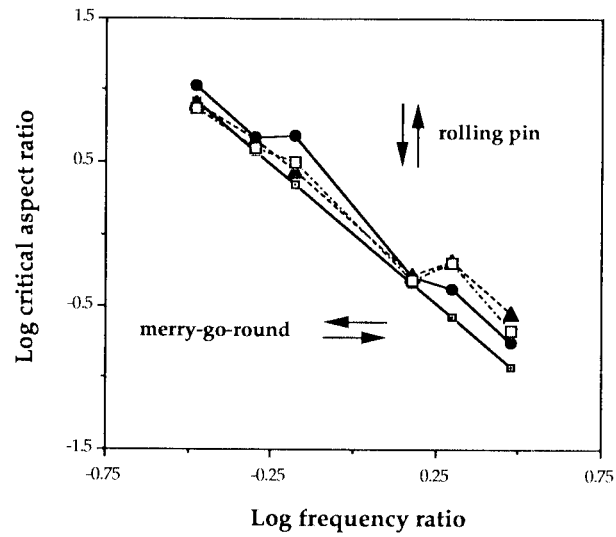

Fig. 10. Log CAR versus log frequency ratio for three subjects plotted together with model predictions. The model predictions are indicated by the small squares that lie on the straight line with a slope approximately equal to -2, the triangles indicate the data for subject LR, the circles indicate the data for subject JBM, and the large squares indicate the data for subject JAP. Frequency/aspect ratio combinations in the upper-right-hand quadrant are seen primarily in rolling-pin motion, while those in the lower-left-hand quadrant are seen primarily in merry-go-round motion. The negatively sloped lines indicate the boundary in the parameter space between these two regimes. The fact that pairs of data points representing rotated stimuli are not located symmetrically with respect to the origin indicates the anisotropy, which for these data favors the merry-go-round interpretation.

Table 1. Raw Data Used To Generate the Graph in Fig. 9[a]

Subject	ω_x	ω_y	log CAR	SEM
JBM	1	2	0.667	0.048
	2	1	0.385	0.084
	1	3	1.030	0.020
	3	1	0.752	0.066
	2	3	0.680	0.052
	3	2	0.294	0.032
LR	1	2	0.657	0.040
	2	1	0.185	0.022
	1	3	0.913	0.020
	3	1	0.547	0.064
	2	3	0.436	0.020
	3	2	0.283	0.062
JAP	1	2	0.594	0.011
	2	1	0.201	0.003
	1	3	0.869	0.007
	3	1	0.673	0.004
	2	3	0.502	0.005
	3	2	0.318	0.007

[a]For subjects JBM and LR, the mean log of the CAR was computed over four replications of each of the six frequency pairs, while three replications were used for subject JAP. The fifth column shows the standard error for each mean.

time may be seen to oscillate in either a horizontal or a vertical direction. It would also be physically consistent for the dots to be seen in circulating motion around the perimeter of the figure, but this is rarely observed.

In this stimulus the aspect ratio of the figure affects the perceived direction of motion. When the horizontal separation is small relative to the vertical separation, it is more likely that horizontal motion will be seen. It is possible to measure a psychometric function relating the aspect ratio to the proportion of the time that horizontal motion is seen. The inflection point of this psychometric function corresponds to a CAR for this task, i.e., the aspect ratio for which horizontal and vertical motions are equally likely to be perceived. With an aspect ratio of unity, the subjects showed a preference for vertical correspondence when the figure was fixated centrally. Note that this is the opposite of what would be predicted if motion correspondence were determined simply by proximity after a deformation consistent with the VHH.

Gengerelli[16] found that this bias disappeared when the display was fixated eccentrically, so that the entire display fell within a single cortical hemifield, and concluded that the bias resulted from a preference to make correspondence within a cortical hemifield. Ramachandran *et al.*[17] performed a similar study on commissurotomy, or split-brain, patients and found that, although the bias was exaggerated, the stimulus was still quite ambiguous, suggesting that the perception of apparent motion across the vertical midline was easily mediated by subcortical structures.

These results pose a puzzle with respect to the results of the present experiment: If there is, for whatever reason, a preference for vertical correspondences in ambiguous motion displays, then we would expect to see a preference for rolling-pin rotation in the ambiguous kinetic depth effect figures used in the present study. This is the opposite of the bias observed in the present study.

If, as is commonly supposed, the computation of two-dimensional optic flow is a necessary precursor to the computation of structure from motion, then any biases inherent in the two-dimensional process should be reflected in the responses of the three-dimensional system. If there are no biases in the two-dimensional motion process, then there cannot be any biases in the three-dimensional process if the ambiguity must be resolved at the two-dimensional stage.

An alternative possibility is that two-dimensional ambiguities are not resolved before the computation of three-dimensional structure. An architecture that would permit this is a distributed representation in which all the possible two-dimensional velocities are represented; the perceived direction would usually correspond to the most-active unit, but less-active units could still pass their signals to higher levels. Even if a bias existed at an early stage, it would be possible for a different (stronger) bias at a later stage to dominate the resolution of the ambiguity. Distributed models for solving the aperture problem have been proposed by Heeger,[18] Sereno,[19] and Perrone.[20] Simoncelli *et al.*[21] have proposed a form of this architecture that is based on probabilities and that includes stages for three-dimensional representation.

C. Ecological Considerations

I have considered several possibilities for the site at which the bias is introduced; I have not, however, said anything about why the bias might be present or whether it has any functional significance. I have mentioned the possibility that the bias is a direct result of an early warping of the

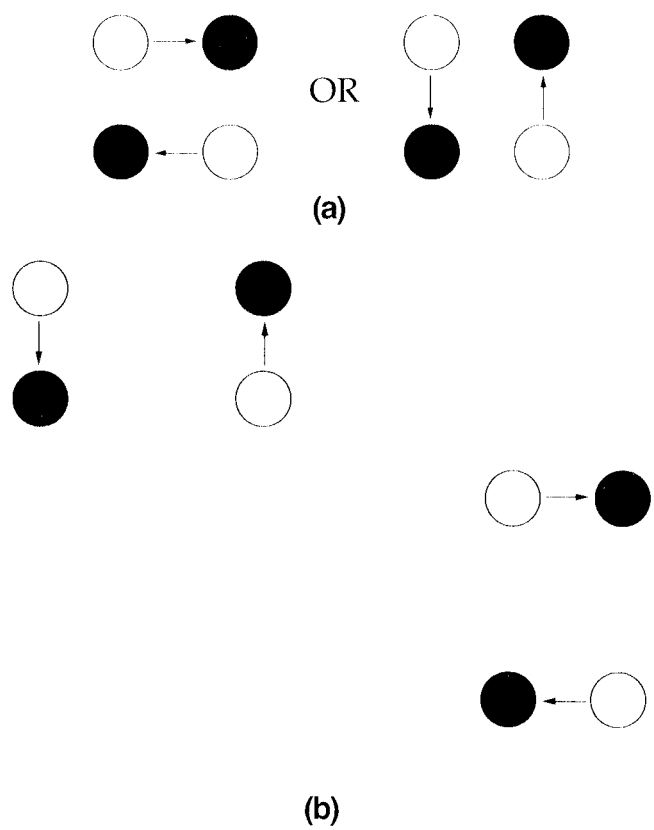

Fig. 11. (a) Stimulus configuration used by Gengerelli[16] and Ramachandran *et al.*[17] The open circles represent dots present at time t_1, which are replaced by dots at the positions shown by the filled circles at time t_2. When this sequence is presented cyclically, the percept is usually of a pair of dots in oscillatory motion, either side to side or up–down. (b) The perceived direction of motion can be biased by changing the aspect ratio of the figure.

visual field by the VHH and has no intrinsic significance. If, on the other hand, the bias is restricted to the interpretation of three-dimensional objects and scenes, then we are faced with the intriguing possibility that the bias arises because of some aspect of the three-dimensional world. According to signal detection theory, an observer's criterion is influenced by the *a priori* probability of a given stimulus[22]; in the present case, it would be sensible to expect a bias similar to that observed if merry-go-round motion were in fact more prevalent in the environment.

Does such an ecological imbalance exist? If so, it is likely to arise from observer self-motion, as opposed to the motion of other objects. Note that there is usually a rotational component to the relative motion between objects and a moving observer, the axis of which depends on the relation between the object and the direction of motion. For example, when an ambulant observer moves through the forest, the tree trunks at eye level have a small component of merry-go-round motion in addition to a large translational component. Similarly, when the observer surmounts a fallen log, the log has a component of rolling-pin motion. Perhaps the bias developed in creatures that were living under open skies and that never had any fallen logs overhead, i.e., that portion of the superior visual field that would have produced a component of rolling-pin motion was devoid of pattern. This argument seems somewhat contrived, however, and requires many assumptions about both the nature of the environment and the behavior of the observer with regard to locomotion and eye movements. If one assumes that an ambulant observer tends to look in the direction of motion, then one would expect that the types of motion encountered would be different for the different regions of the visual field; in particular, this might produce a bias toward rolling-pin motion in the inferior visual field. An investigation of the dependence of the observed bias on position in the visual field would be an interesting topic of future research.

5. CONCLUSIONS

In spite of the fact that Hildreth's theory does not predict the anisotropy seen in this study, the fact that a modified version of the theory predicts the correct dependence on frequency ratio (i.e., the correct slope in Fig. 10) is strong support for the theory. The theory can easily be made to predict the biases if an affine transformation consistent with the VHH is assumed to precede the motion analysis. Reported anisotropies in two-dimensional motion correspondence, however, are inconsistent with this view.[16,17] One possibility is that the phenomena studied by Ramachandran *et al.*[17] involve completely different mechanisms subject to their own distinct biases. An alternative explanation is that the observed bias is introduced at a level involving three-dimensional representation; an implication of this hypothesis is that the two-dimensional aperture problem is not resolved independently of three-dimensional interpretation.

ACKNOWLEDGMENTS

The bulk of the experimental data were collected while the author was a graduate student in the Department of Psychology at the University of California, San Diego, and supported by National Institutes of Health grant EY-01711 (awarded to Donald I. A. MacLeod, whose personal support is also gratefully acknowledged) and an IBM graduate fellowship. Computational modeling and preparation of the manuscript were performed at the NASA Ames Research Center, where the work was supported by NASA Research Technology Operating Plan (RTOP) 506-47. The initial implementation of the Hildreth algorithm was based on a mathematica program provided by Maggie Shiffrar and Al Ahumada. Thanks to John Perrone for donating some subject hours.

REFERENCES

1. W. R. Miles, "Movement interpretation of the silhouette of a revolving fan," Am. J. Psychol. **43,** 392–405 (1931).
2. H. Wallach and D. O'Connel, "The kinetic depth effect," J. Exp. Psychol. **45,** 205–217 (1953).
3. S. Ullman, *The Interpretation of Visual Motion* (MIT Press, Cambridge, Mass., 1979).
4. S. Ullman, "Maximizing rigidity: the incremental recovery of 3-D structure from rigid and rubbery motions," Perception **13,** 255–274 (1984).
5. H. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," Proc. R. Soc. London Series B **208,** 385–397 (1980).
6. D. D. Hoffman, "Inferring local surface orientation from motion fields," J. Opt. Soc. Am. **72,** 888–892 (1982).
7. D. J. Heeger and A. Jepson, "Visual perception of three-dimensional motion," Neural Computat. **2,** 127–135 (1990).
8. E. Hildreth, "The computation of the velocity field," Proc. R. Soc. London Ser. B **221,** 189–220 (1984).
9. E. Hildreth, "Computations underlying the measurement of visual motion," Artif. Intell. **23,** 309–355 (1984).
10. C. O. Weber, "Apparent movement in Lissajou figures," Am. J. Psychol. **42,** 647–649 (1930).
11. B. R. Philip and V. R. Fisichelli, "Effect of speed of rotation and complexity of pattern on the reversals of apparent movement in Lissajou figures," Am. J. Psychol. **58,** 530–539 (1945).
12. V. R. Fisichelli, "Effect of rotational axis and dimensional variations on the reversals of apparent movement in Lissajous figures," Am. J. Psychol. **59,** 669–675 (1946).
13. J. B. Mulligan and D. I. A. MacLeod, "Reciprocity between luminance and dot density in the perception of brightness," Vision Res. **28,** 503–519 (1988).
14. J. O. Robinson, *The Psychology of Visual Illusion* (Hutchinsons, London, 1972).
15. R. B. Post and M. Chaderjian, "Perceived path of oblique motion: horizontal–vertical and stimulus-orientation effects," Perception **16,** 23–28 (1987).
16. J. A. Gengerelli, "Apparent movement in relation to homonymous and heteronymous stimulation of the cerebral hemispheres," J. Exp. Psychol. **38,** 592–599 (1948).
17. V. S. Ramachandran, A. Cronin Golomb, and J. J. Myers, "Perception of apparent motion by commissurotomy patients," Nature (London) **320,** 358–359 (1986).
18. D. J. Heeger, "Model for the extraction of image flow," J. Opt. Soc. Am. A **4,** 1455–1471 (1987).
19. M. E. Sereno, "Implementing stages of motion analysis in neural networks," in *Proceedings of the 9th Annual Conference of the Cognitive Sciences Society* (Erlbaum, Hillsdale, N.J., 1987), pp. 405–416.
20. J. A. Perrone, "Simple technique for optical flow estimation," J. Opt. Soc. Am. A **7,** 264–278 (1990).
21. E. P. Simoncelli, D. J. Heeger, and E. H. Adelson, "Perception of 3D motion in the presence of uncertainty," Invest. Ophthalmol. Vis. Sci. Suppl. **31,** 173 (1990).
22. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).