Reprinted from Journal of the Optical Society of America A, Vol. *2*, p. 343, February 1985

Structure from stereo and motion

Whitman Richards

Natural Computation Group, E10-120, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received November 23, 1983; accepted September 5, 1984

Stereopsis and motion parallax are two methods for recovering three-dimensional (3D) shapes. Theoretical analyses of each method show that neither alone can recover rigid 3D shapes correctly unless other information, such as perspective and vertical disparity, is included. The solutions for recovering rigid structure from motion have a reflection ambiguity; the depth scale of the stereoscopic solution will not be known unless the fixation distance is specified in units of interpupil separation. (Hence the configuration will appear distorted.) However, the correct configuration and disposition of a rigid 3D shape can be recovered if stereopsis and motion are integrated, for then a unique solution follows from a set of linear equations. The correct interpretation requires only three points and two stereo views.

INTRODUCTION: THE PROBLEM

One of the essential tasks of vision is to determine the three-dimensional (3D) shape of objects in the world.[1] Once such information is available, a useful 3D model of an object can be constructed, which is suitable for recognition or manipulation, for example. Unfortunately, neither stereopsis nor motion parallax *alone* provides enough information to recover the correct 3D disposition or shape. Each method suffers serious defects unless other information is brought into play.

The critical defect with stereopsis is that the same rigid configuration of points seen at different distances will elicit different angular disparities on the two retinas. To recover the correct distance relations between the points using only horizontal disparity requires knowledge of the fixation distance.[2] Let an observer view an equilateral triangle lying in the horizontal plane at distance D_A as illustrated in Fig. 1(a). If the altitude of the triangle is z_A, then the angular disparity δx_A of the nearer point with respect to the farther two base points will be

$$\delta x_A = z_A(I/D_A{}^2), \tag{1}$$

where I is the interpupil separation between the two eyes (cameras) and small angle approximations are taken. Now, if the triangle is moved farther away to position D_B, then clearly the angular disparity δx_B of the near vertex will be reduced by the factor $D_A{}^2/D_B{}^2$. However, the angular width of the base will have decreased by only D_A/D_B. The triangle that previously appeared equilateral should thus appear squashed by the factor D_A/D_B as it is moved farther away. The triangle that *appears* equilateral based on (horizontal) disparity information alone must thus have a greater altitude, as shown in Fig. 1(b). In sum, the configuration or shape of a rigid set of points is not uniquely determined from stereopsis alone.

Recovering the 3D configuration from motion also presents problems unless information other than the (orthographic) motion of the points is provided. To illustrate the difficulty, let us assume that the motion-parallax solution [or equivalently the structure-from-motion (SFM) solution] requires at least three points and two views [for example, Hoffman and Flinchbaugh[3] and Bobick[4] show conditions and constraints under which the 3D configuration can be recovered from the two-dimensional (2D) projection of three points]; Ullman[5] used four points, and Prazdny[6] used five points. With one exception, all these solutions, including those velocity fields, are to a set of second-degree equations, which means that there is a duplicate solution that is a reflection about a plane. (More recently, Tsai and Huang[7] have obtained a linear solution for eight points.) For the given minimum number of points, therefore, each group containing this minimum has at least two solutions, one being a reflection of the other. Consider then the configuration of six points shown in Fig. 2(a). The triplets of points joined by solid lines are in a rigid relation, but the link between the two groups of triplets is not rigid (dashed line), as if the two parts are joined by a flexible rod. Because each of the two groups of triplets has a reflection ambiguity, other SFM interpretations of the entire configuration are possible, such as the one shown in Fig. 2(b). [Two other possible interpretations are the reflections of Figs. 2(a) and 2(b) about the horizontal line of four points.] A unique SFM solution thus requires removal of the reflection ambiguity.

By combining stereopsis with SFM we shall see that both ambiguities in the 3D interpretations can be eliminated. Stereopsis provides the sign needed to tell whether the ambiguous points seen with SFM are behind or in front of the others; SFM, on the other hand, correctly interprets the angular relations between the points, thus aiding stereopsis by eliminating the fixation-distance dependency.

STRUCTURE FROM STEREO PROPOSITION FOR TWO POINTS

Discrete Case

We will begin by considering the simple discrete case in which a stereo observer views a rigid configuration of points from one position (frame 1) and then moves to another position to obtain a second view (frame 2). Thus, although these discrete views do not make explicit the instantaneous velocities of the points, a measure of the relative velocities of the points can be obtained by keeping the temporal intervals between views

0740-3232/85/020343-07$02.00

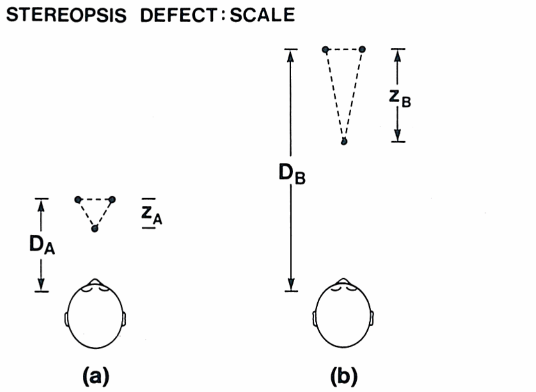

Fig. 1. Two kinds of failings in the recovery of 3D structure. For stereopsis, a given disparity will indicate a different distance, depending upon the observation distance, D. Thus (a) the near vertex of the equilateral triangle at distance D_A has the same disparity as (b) the near vertex of the isosceles triangle D_B.

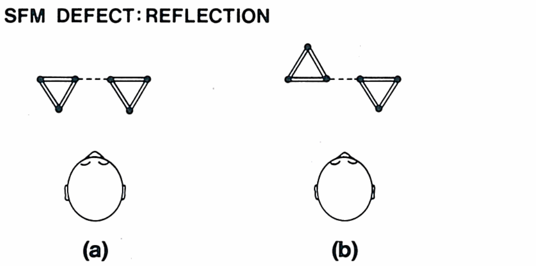

Fig. 2. When structure is recovered motion, there is a reflection ambiguity. This ambiguity becomes a problem as the structure becomes increasingly nonrigid, such as when there is a flexible link (dashed line) between two rigid components.

constant. (Below we treat the case for which the instantaneous velocities are available.) This is the approach used by Ullman[5] in his classical monocular SFM solution. The problem here, then, is to determine how many points P and how many stereo views V are needed to recover the correct configuration of points.

Figure 3 shows the viewing conditions and the coordinate system used. The bisector of the lines of sight is taken as the Z axis (note direction); the XZ (horizontal) plane is defined as including the two lines of sight. (The solution will assume that the horizontal axes of the two retinas or cameras lie in the XZ plane.) The Y axis is normal to the XZ plane at the fixation point O. The point P(x, y, z) and the origin O of the coordinate systems are assumed to be far away so that perspective information is nil; hence the projections are orthographic onto the *separate* frontal planes of the two eyes.

The basic problem is to recover the distance $OP(x, y, z)$ and the orientation σ, τ that the ray makes with the Z and Y axes. Because the views are orthographic and epipolar, τ appears in the image plane as does the elevation of P, namely, y_P. Because the azimuth of P, namely, x_P, also appears in the image, the problem reduces to recovering τ and the distance $OP_{xz} = (x_P{}^2 + z_P{}^2)^{1/2}$. Our two unknowns, σ_P and z_P, are thus entirely confined to the horizontal plane. Let us then consider only the top view of the situation, as shown in Fig. 4.

Here the projection of P(x, y, z) onto the XZ plane is denoted as P_1 for our first point, with the subscript 1 indicating our first view. The complementary angle $\theta_1 = (\pi/2) - \sigma_1$ has replaced σ. For any single view and point P, our unknown is either θ_{P1} or z_{P1}. Of course we know x_{P1}, which appears in the image, and because the viewing is stereoscopic, we also know the angular disparity of point P_1 with respect to O. Let this disparity be designated as δx_{P1}.

Unfortunately, knowledge of the angular disparity of P_1 is not sufficient to solve for its z coordinate, because by Eq. (1) we do not have knowledge of the interpupil separation or the fixation distance to O. This was the fatal defect of stereopsis alone. However, if we move our head (or cameras) slightly to one side, keeping the distance to O constant, then we have a second stereo view of P, namely, P_2 seen at azimuth x_{P2} with the observed disparity δx_{P2}. Although this lateral motion has introduced a new unknown, namely, z_{P2}, the ratio z_{P1}/z_{P2} will equal that of the observed disparities $\delta x_{P1}/\delta x_{P2}$, as can be seen readily from Eq. (1). Appendix A shows that this information is then in principle sufficient to recover the distance OP and its orientation to the viewer. Specifically, we can solve for the angle θ_2 in Fig. 2 as follows:

$$\theta_2 = \tan^{-1}\left(\frac{x_{P1}{}^2/x_{P2}{}^2 - 1}{1 - r_{P2}}\right)^{1/2}, \tag{2}$$

where $r_P = \delta x_{P1}/\delta_{P2}$. Because OP_2 is simply $x_{P2} \sec \theta_2$, we can calculate OP from y_P, which appears in the image plane. Hence we have the following SFM and stereo claim for two points:

Claim 1: Given two coplanar orthographic stereo views of two rigid points, their correct 3D disposition can be recovered uniquely, independent of fixation distance.

Note that the above claim speaks only of the disposition of the two points (i.e., the angle θ_1). Although we have taken the azimuth x_{P1} and elevation y_P of P to be distances, in fact they are seen only as angles on the retina. Thus the correct configuration, or angular relations between a set of points, can be

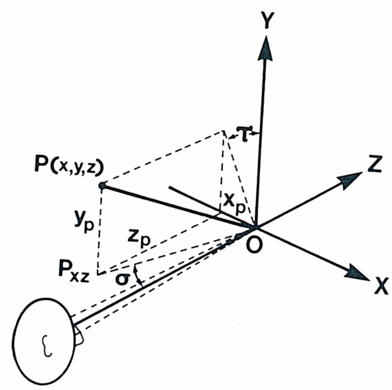

Fig. 3. Schematic showing the coordinate system and notation used.

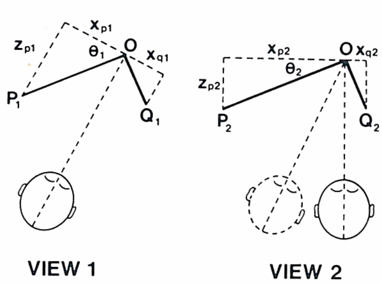

Fig. 4. Top view showing the projections of points onto the horizontal plane XZ. Note the angle has been replaced by its complement θ.

determined uniquely from two stereoscopic views, but not the actual absolute distances.

Continuous Case

Our visual system is remarkably sensitive to directional motion.[8] Rather than simply taking snapshots of a configuration of points as we move our heads, let us now assume that the instantaneous retinal velocity of any point is available, as well as its position. Under these conditions, Appendix B then shows that once again the angle θ may be recovered by using the following relation:

$$\theta = \tan^{-1}\left(\frac{\Delta\dot{x}/\dot{x}}{\Delta\delta x/\delta x}\right)^{1/2} \tag{3}$$

We thus make the following second claim:

Claim 2: Given one orthographic stereo view of two rigid points and their velocities, their correct 3D disposition can be recovered uniquely independent of fixation distance.

Thus we now have two methods of recovering the correct angular relations between a set of points.

THE INTERPRETATION RULE

The above two claims specify the minimal input required in order to obtain a unique solution for the 3D configuration of a rigid set of points, as seen in the 2D image. Should we then apply our solution for the 3D configuration of points to all pairs of points seen on our retinas? Clearly not, for some pairs will not be rigidly linked in three dimensions, and our interpretations will be incorrect. We thus need to be able to test from the image data whether or not a given pair of points is indeed rigidly linked. Specifically, we are required to identify false targets.

Appendixes A and B analyze the false-target possibility and show that either one more point or one more (stereo) view will allow the observer to eliminate point pairs that do not arise from rigid 3D configurations. Thus we may test and verify our rigidity hypothesis from the sense data. If the points pass the rigidity test, then we propose that the points be interpreted as arising from a rigid configuration.[5] We then have the following four interpretation rules:

Rule 1: (Discrete Case): If *three* coplanar stereo views of two points have a fixed separation according to the application of Eq. (2), then these points should be interpreted as being in a rigid configuration.

Rule 2: (Discrete Case): If *three* points and two coplanar stereo views have a fixed separation according to the application of Eq. (A2), then these points should be interpreted as being in a rigid configuration.

Rule 3: (Continuous Case): If two independent stereo views of two points plus their velocities suggest a fixed separation between these points according to Eq. (3), then these points should be interpreted as being in a rigid configuration.

Rule 4: If any of the above rules fail to apply (within certain as yet unspecified signal-to-noise considerations), then the points are not in a rigid configuration.

SENSITIVITY TO IMPERFECT DATA

The above analysis shows that, in principle, stereo and motion can be combined to recover the correct 3D configuration of

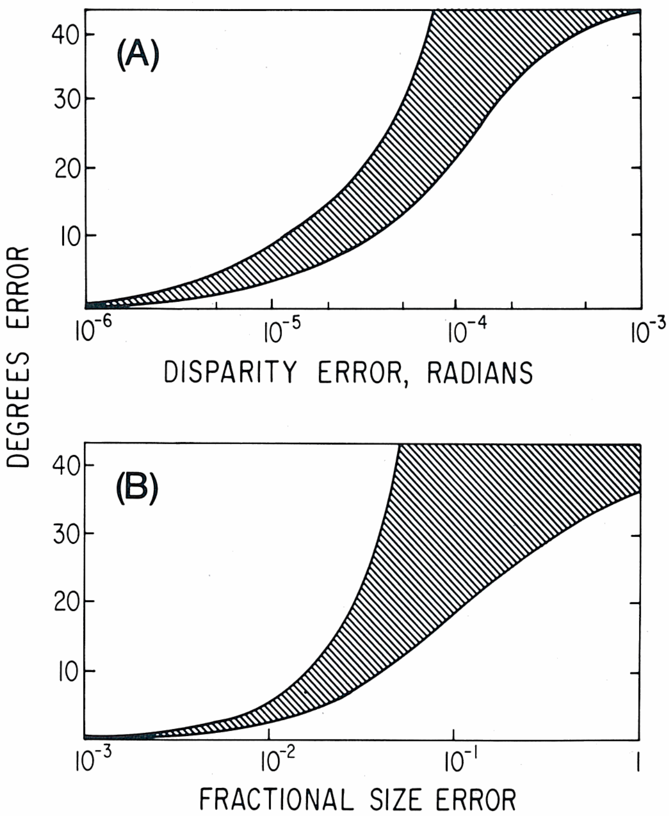

Fig. 5. Effect of noisy image data on the recovery of a 3D configuration at 10 m. The ordinate shows the error in estimating a configuration of two points separated at 1.4 m and oriented at 45° to the observer in the horizontal plane. (A) Configuration errors introduced by incorrect disparity measurements. (B) Configuration errors introduced by incorrect angular-size measurements. The estimation errors can be reduced substantially if greater lateral movements are used or if the configuration is closer to the observer.

points from 2D-image data. However, what kind of precision is required of the data in order that the 3D configuration can be reconstructed with reasonable accuracy? Will only a slight amount of noise in the data cause gross changes in the interpreted 3D shape, or do the estimation errors increase monotonically as the data become increasingly unreliable?

Figure 5 shows the result of one simulation using two stereo views of two points. (For this analysis, it is not necessary to add a third view or point.) The test case was two points separated by 1.4 m at 10 m, with one point lying at a 45° orientation to the other in the horizontal plane. The interpupil distance was taken as 6.4 cm, and the two stereo views corresponded to a 2° movement about the fixation point (i.e., a lateral motion of about 0.3 m). Figure 5(A) shows the result of increasing the disparity-measurement error; Fig. 5(B) shows the result of increasing the error of the angular-size measurements. The crosshatched regions indicate the range of the misestimation error, which depends upon whether positive- or negative-measurement errors are introduced. Note that all estimation errors rise monotonically as the measurement errors increase. Hence the interpretation process will be well behaved under measurement error.

Returning to Fig. 5(A), we see that for the particular conditions chosen, a disparity error of 10^{-5} rad will introduce a 7° error in one's estimate of the configuration or orientation of the second to the first point. (The error in estimating the separation between points is trivial—less than 10% for most of the data illustrated.) A disparity error of 10^{-5} rad corresponds to about 2 sec of arc, which is the limit of stereo acuity. An order-of-magnitude improvement in performance occurs, however, if the fixation distance and configuration is reduced to 1 m with the points separated by 14 cm. Alternatively, the lateral movement could be increased. Thus the exquisite human stereo sensitivity allows stereo and motion to be combined successfully, especially during locomotion. Figure 5(B) shows the effect of introducing errors in the measurement of angular size. Again, because human sensitivity to size[9] is of the order of 3%, the recovery of the 3D structure from the 2D-image data seems feasible. Combining both the stereo and size errors does not seriously change these results and may in fact improve performance when the signs of the two errors are chosen appropriately.

PSYCHOPHYSICAL PREDICTIONS

The above analysis suggests three possible schemes for recovering the correct 3D configuration of points using stereopsis together with motion. To date, no psychophysics is available to favor one scheme over another. However, we can present some past results showing that stereopsis and motion are indeed intimately coupled modules in the human visual system.

Research of Regan and Beverley

It has long been known that changing an object's size can produce a compelling impression that the object is moving in depth.[10] The physiological basis for this phenomenon, often described as looming, which can be seen monocularly, is different from motion in depth created binocularly by changing disparity.[11–13] Over the past 10 years, Regan and his colleagues have amassed considerable evidence for the presence of separate and quasi-independent channels that each respond selectively either to changing-size stimulation or to changing-disparity stimulation.[12,14–18] Regan's data thus support the plausibility of the human visual system's ability to compute Eq. (3), for example, which requires measurements of changing size or velocity (Δx) and of changing disparity ($\Delta\delta x$).

Regan and Beverley[19] also show that the changing-size and changing-disparity channels feed into a common motion-in-depth stage. This conclusion is reinforced by more-recent data of Richards and Liberman,[20] who explore the nature of the interaction. These independent results thus support our computational prediction that both motion and disparity information should come together early in the processing in order that the correct 3D configuration of objects can be determined. According to Eq. (3), one possible form of this interaction would be a division, or, more simply, a subtraction if a logarithmic transformation of the signals were made en route to the common stage.

In their 1979 paper, Regan and Beverley[19] show that the one advantage of comparing changing size with changing disparity is that absolute size of the moving object can be recovered up to a constant scale factor, namely, the separation between the eyes. Alternatively, one might view the yardstick for absolute size as simply the interpupil separation.

In this paper, another role is suggested for a stage that combines size change with changing disparity, namely, the ability to recover the correct configuration of objects in space. To do this, however, requires that the changing size and changing disparities be measured relative to their current magnitudes rather than that the actual increments themselves be used as proposed by Regan and Beverley. Thus we use $\Delta x/x$ and $\Delta\delta x/x$ rather than Δx and $\Delta\delta x$.

Demonstration

Perhaps a most convincing argument for the plausibility of combining stereo and SFM is a simple demonstration. Examine a tree from your window, or perhaps even your fingertips arranged in a pentagon and held vertically at arm's length. If you view this tree (or the fingers) with one eye and rock your head sideways just a bit, then indeed a 3D shape emerges from the motion parallax. Similarly, with binocular viewing and no head motion a 3D shape is also apparent. But are these impressions correct? As soon as one combines binocular viewing with the lateral head motion, then the *correct* 3D configuration becomes clear and vivid.[21] Something is clearly gained by combining the two modules.

SUMMARY

Combining stereo disparity with SFM is one way that the correct 3D configurations and relations between objects can be recovered from 2D images. Neither stereopsis nor motion parallax nor SFM can do this alone. That the human visual system indeed combines these two computational schemes into one appears plausible. Not only do our impressions of the 3D world improve by the combination, but psychophysical evidence suggests that the required neural mechanisms are present. One immediately is led to inquire whether other modules in combination, such as stereo and shape from shading,[22] or motion and shape from shading, would offer similar advantages.

APPENDIX A: STRUCTURE FROM STEREO PROPOSITION FOR TWO POINTS

Proposition 1

Given two coplanar orthographic stereo views of two rigid points, their 3D disposition may be recovered uniquely independent of fixation distance.

Proof

Let the two lines of sight from each stereo view lie in the XZ plane and intersect at O, as shown in Fig. 3. Any point $P(x, y, z)$ can then be specified by its distance from O and two angles σ, τ. Because the views are orthographic, τ appears in the image plane, as does the elevation of P, namely, y_p and its azimuth x_p. The problem then reduces to recovering σ or P_{xz}, the projection of $P(x, y, z)$ onto the XZ plane.

As can be seen from the above remarks, the projection of $P(x, y, z)$ onto the XZ plane is shown in Fig. 4. For notational convenience P_1 has replaced P_{xy} and $\theta = \pi/2 - \sigma$. Our unknowns are thus θ_i and z_{pi}, because x_{pi} appears in the image plane.

From the fact that the length OP_i is constant over all views, we obtain

$$\overline{OP}_1{}^2 = \overline{OP}_2{}^2 = x_{p1}{}^2 + z_{p1}{}^2 = x_{p2}{}^2 + z_{p2}{}^2 \qquad \text{(A1)}$$

with unknowns z_{p1}, z_{p2}.

From the fact that each view is stereoscopic, we obtain the distance-disparity relation

$$\frac{\delta x_{p1}}{\delta x_{p2}} = \frac{z_{p1}}{z_{p2}} = r_p, \tag{A2}$$

where δx_{pi} is the measured disparity, thereby making r_p a known constant. This relation follows from the fact that the horizontal disparity of P relative to O is given by

$$\delta x_{pi} = z_{pi}(I/D^2) \tag{A3}$$

where I is the interpupil distance and D is the line-of-sight distance to O, and given that the distance OP is much smaller than D. Taking the ratio of Eq. (3) for $i = 1, 2$ eliminates the (I/D^2) dependency.

We now have two equations (A1) and (A2) in two unknowns, z_{p1}, z_{p2}, which can be solved for θ_2:

$$\theta_2 = \tan^{-1}\left(\frac{x_{p1}^2/x_{p2}^2 - 1}{1 - r_{p^2}}\right)^{1/2} \tag{A4}$$

The length OP_2 is then simply $x_{p2} \sec \theta_2$, from which OP can be calculated because y_{p2} appears in the image plane.

Uniqueness

The square root in the solution (A4) for the angle θ_2 allows only positive values for θ_2. Yet the correct value for θ_2 may be either positive or negative, depending on whether point P_2 lies in front of or behind the frontal plane containing the fixation point O. The solution (A4) for θ_2 is thus not unique unless the sign of z_{p2} is known. However, the sign of z_{p2} is known. The sign of z_{p2} is the same as that for the disparity of P_2, namely, δx_{p2}. Hence the position of P_1 and thus also $P(x, y, z)$ can be determined uniquely.

Degeneracies

Under some conditions, Eq. (4) cannot be solved for θ_2. The only case is when the denominator $(1 - r_p{}^2)$ is zero. This corresponds to $\delta x_{p1} = \delta x_{p2}$, or when P_1 and P_2 both lie in the same frontal plane. [This can be shown to be the only singular condition by evaluating the Jacobian of Eqs. (A1) and (A2).[23] The value of this determinant will be zero only when $r_p = z_{p2}/z_{p1}$. But because $r_p = z_{p1}/z_{p2}$, this singularity corresponds to $z_{p1} = z_{p2}$, as before.]

False Targets

Is it possible that another pair of points not in a rigid configuration will also satisfy Eq. (A4)? If so, then a valid interpretation of this equation is not possible, because the observer would have no way of determining whether the solution came from a rigid configuration.

Let us assume that points O and Q also satisfy Eq. (A4) and thus appear rigid, although they are not. Let the competing rigid solution be O, P. Then, as seen in the image plane, P and Q must be coincident:

$$x_{pi} = x_{qi}; \quad y_{pi} = y_{qi}. \tag{A5}$$

The only ambiguity is in the Z values of P and Q. For two views, we may relate these Z values by the parameter a_i as follows:

$$z_{q1} = a_1 z_{p1},$$
$$z_{q2} = a_2 z_{p2}. \tag{A6}$$

However, because the disparity ratios for the two views of P and Q are known, they must also be identical for P and Q to appear the same. Hence from Eq. (A2) we have

$$\frac{z_{q1}}{z_{q2}} = r_q = r_p = \frac{z_{p1}}{z_{p2}}. \tag{A7}$$

Thus, combining Eq. (A7) with Eqs. (A6), we have

$$\frac{z_{q1}}{z_{q2}} = \frac{a_1 z_{p1}}{a_2 z_{p2}} = \frac{z_{p1}}{z_{p2}}, \tag{A8}$$

requiring that $a_1 = a_2$. Thus the only false-target condition is when

$$z_{q1} = a \cdot z_{p1},$$
$$z_{q2} = a \cdot z_{p2}. \tag{A9}$$

To explore this single false-target possibility, we will determine the values of a that lead to false targets. Recall that x_{pi} must equal x_{qi}. Hence we may combine Eq. (A1) renotated to point Q with expressions (A9) to obtain

$$\overline{OQ}_1{}^2 = x_{q1}{}^2 + z_{q1}{}^2 = x_{p1}{}^2 + a^2 \cdot z_{p1}{}^2,$$
$$\overline{OQ}_2{}^2 = x_{q2}{}^2 + z_{q2}{}^2 = x_{p2}{}^2 + a^2 \cdot z_{p2}{}^2. \tag{A10}$$

The difference in length $\overline{OQ}_1{}^2 - \overline{OQ}_2{}^2$ is thus

$$\overline{OQ}_1{}^2 - \overline{OQ}_2{}^2 = (x_{p1}{}^2 - x_{p2}{}^2) - a^2 \cdot (z_{p2}{}^2 - z_{p1}{}^2). \tag{A11}$$

But because OP is rigid (of fixed length), we may eliminate the z_{pi} term using Eq. (A1) to obtain the conditions on Q_1 and Q_2 required to produce a false target, namely,

$$\overline{OQ}_1{}^2 - \overline{OQ}_2{}^2 = (x_{p1}{}^2 - x_{p2}{}^2)(1 - a^2). \tag{A12}$$

From Eq. (A12) we see immediately that there is no rigid false target OQ_i because then the left-hand side of Eqs. (A9) would be zero, forcing $a = 1$, which from Eqs. (A9) makes point Q identical to P. How then can nonrigid false targets be excluded?

If the distance between a pair of points is nonrigid, then the value of a will be different from 1. Furthermore, because the distance between O and Q will change from one view to the next, so must the value of a (otherwise OQ is a rigid configuration). Thus the simplest strategy to eliminate false targets is to add an extra (third) view and determine whether the distance OP indeed remains constant. If it does, then a must have been constant. The probability of this occurrence by chance for arbitrarily chosen values of a is zero, except if the configuration is rigid.

Alternatively, a third (rigid) point R may also be included in the configuration. In this case, the angle POR must be consistent with the lengths OP, OR, and PR, again overconstraining the solution.

This result now leads to the following two interpretation rules:

Rule 1

If *three* coplanar stereo views of two points have a fixed separation according to the application of Eq. (A4), then these points should be interpreted as being in a rigid configuration.

Rule 2

If *three* points and two coplanar stereo views have a fixed separation according to the application of Eq. (A4), then these points should be interpreted as being in a rigid configuration.

APPENDIX B: STRUCTURE FROM STEREO PROPOSITION FOR TWO POINTS PLUS VELOCITIES

Proposition 2

Given one orthographic stereo view of two rigid points and their velocities, their 3D disposition may be recovered uniquely independent of fixation distance.

Proof

Once again, the relations between the viewer and point $P(x, y, z)$ are as shown before in Fig. 3. Because the projections x_p and y_p are known, the problem reduces to recovering σ or P_{xz}, the projection of $P(x, y, z)$ onto the XZ plane.

From above, the projection of $P(x, y, z)$ onto the XZ plane is shown in Fig. 2 as before, with the substitution $\theta = (\pi/2) - \sigma$. Because more details about the geometry of OP are required, this portion of Fig. 4 is further expanded to become Fig. 6. The notations here have also been simplified by dropping the subscript p. The problem now is to show how θ can be measured from the projection of P onto the X axis.

As the observer rotates about the fixation point O by an angle ϕ, the XZ axes will rotate by the same angle because they are defined with respect to the observer's position. Let R be the projection of P onto the X axis, lying at distance x_1 from O. Then for any fixed angle of observer rotation ϕ, R moves to R', causing x_1 to increase to x_2 and z_1 to decrease to z_2. Note that both R and R' will lie on the same arc because OP is fixed and z_i is perpendicular to x_i by definition. Thus, at any instant, the motion of R will be tangent to the circle ORP. As $\phi \rightarrow 0$, this tangent then describes the direction of change R in the XZ plane. As is shown in Fig. 6, the tangent vector will have a length Δx in the X axis and Δz in the Z axis. From the geometry,

$$\tan\theta = \frac{\Delta x_1}{\Delta z_1} = \frac{z_1}{x_1}. \tag{B1a,b}$$

Recalling now from Eq. (A3) of Appendix A the relation between disparity δx_{p1} and distance z_{p1},

$$z_1 = \delta x_1(D^2/I), \tag{B2}$$

where D is the fixation distance and I the interpupil separation. Noting that the same relation (B2) holds between Δz_1 and $\Delta\delta x_1$, we can eliminate the (D^2/I) term by division to obtain

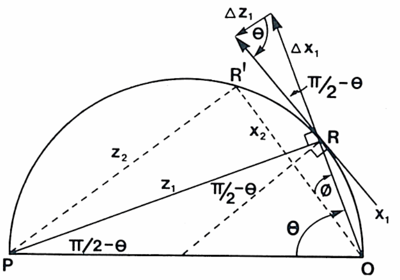

Fig. 6. An expanded view of a portion of Fig. 4.

$$\frac{\Delta\delta x_1}{\delta x_1} = \frac{\Delta z_1}{z_1}. \tag{B3}$$

We now have three equations, Eqs. (B1a,b) and (B3), in the three unknowns Δz_1, z_1, and θ. Solving for θ, we find that

$$\theta = \tan^{-1}\left(\frac{\Delta x/x}{\Delta\delta x/\delta x}\right)^{1/2}, \tag{B4}$$

where the expression in parentheses is simply the ratio of the increment of the projection of OP onto the X axis to its relative disparity increment. Or, in terms of velocities, it is the ratio of the x component of the velocity of P to the rate of disparity change, both normalized by their distances from O.

To recover the length OP_1, we note that $\cos\theta = x_1/OP_1$. Hence

$$OP_1 = x_1 \sec\theta = x_1(1 + \tan^2\theta)^{1/2}. \tag{B5}$$

Substituting Eq. (B4) into Eq. (B5), we find that

$$OP_1 = x\left(1 + \frac{\Delta x/x}{\Delta\delta x/\delta x}\right)^{1/2}. \tag{B6}$$

Thus the disposition and the length between two points O and P are recoverable from one dynamic stereo view that generates relative motion of disparity and angular extent.

Uniqueness

As before in Appendix A, although there is a square-root solution for θ, Eqs. (B4) and (B6) will yield unique solutions because the sign of z is the same as that for Δx and is known. Hence the position of P_{xz}, and hence $P(x, y, z)$, can be determined uniquely.

Degeneracies

Equation (B5) cannot be solved when x or $\Delta\delta x$ is zero, corresponding to $\theta = \pi/2$. Referring to Fig. 6, we see that this condition is equivalent to point P lying in the sagittal YZ plane. (Note that this degeneracy would not occur if perspective, rather than orthographic projection, were assumed.) As long as the observer's motion is such that the configuration OP will undergo some rotation, this degenerate condition will not occur in practice.

False Targets

Here we wish to determine the conditions in which a point other than P will also satisfy Eqs. (B4) and (B6). Let us assume that there is such a point Q, with position coordinates (x, y, z_q) and velocities $(\Delta x/\Delta t, \Delta y/\Delta t, \Delta z_q/\Delta t)$. Because the $x, y, \Delta x, \Delta y$ values appear in the image, the z_q and Δz_q are the only unknowns. These unknowns for point Q can be related to the corresponding values z_p and Δz_p for point P as follows:

$$z_q = a_1 z_p,$$
$$\Delta z_q = a_2 \Delta z_p. \tag{B7}$$

However, Eq. (B3) gives us the relation between the known disparity ratios for points P and Q:

$$\frac{\Delta\delta x_p}{\delta x_p} = \frac{\Delta z_p}{z_p}, \tag{B8a}$$

$$\frac{\Delta\delta x_q}{\delta x_q} = \frac{\Delta z_q}{z_q} = \frac{a_2\Delta z_p}{a_1 z_p}. \tag{B8b}$$

But the disparities $\Delta\delta x_{p,q}$ and $\delta x_{p,q}$ are observables and hence must be the same. Equating Eqs. (B8a) and (B8b), we see that $a_2 = a_1$. Thus the only false-target condition is when

$$\Delta z_q = a\Delta z_p,$$
$$z_a = a z_p. \tag{B9}$$

To explore this single false-target possibility, we will determine the values of a that lead to false targets.

Referring to equation (B1), the angular values θ_p and θ_q for P and Q satisfy

$$\tan\theta_p = \frac{\Delta x_q}{\Delta z_p} = \frac{z_p}{x},$$
$$\tan\theta_q = \frac{\Delta x_q}{\Delta z_q} = \frac{z_q}{x}. \tag{B10}$$

Thus

$$x\Delta x = z_p\Delta z_p,$$
$$x\Delta x = z_q\Delta z_q = a^2 z_p\Delta z_p, \tag{B11}$$

where Eqs. (B9) have been used to express the z values for Q in terms of those for P. But Eqs. (B11) force $a^2 = 1$ for all Q's. Hence from Eqs. (B9) we see that Q is identical to P and there are no false targets. (This result may have been anticipated, because the solution for the configuration of OP was based on instantaneous values of the position and velocity of P.) This result now leads to the following interpretation rules:

Rule 1
If two independent stereo views of two points plus their velocities suggest a fixed separation between these points according to Eq. (B6), then these points should be interpreted as being in a rigid configuration.

Rule 2
If Rule 1 fails to apply (within certain yet-to-be-specified signal-to-noise considerations), then the two points are not in a rigid configuration.

Thus, because Proposition B is based on an instantaneous analysis of the sensory data, it provides the basis for a potentially more powerful scheme for interpreting the structure of both rigid and nonrigid configurations.

ACKNOWLEDGMENT

This report describes research done in the Natural Computation Group at the Department of Psychology, utilizing the facilities of the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for this work is provided by the National Science Foundation and the U.S. Air Force Office of Scientific Research (AFSOR) under a combined grant for studies in Natural Computation, grant 79–23110-MCS, and by the AFOSR under an Image Understanding contract F49620–83–C–0135.

REFERENCES

1. D. C. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, 1982).
2. J. E. W. Mayhew and H. C. Longuet-Higgins, "A computational model of binocular depth perception," Nature **297,** 376–378 (1982), have shown that, in principle, the use of vertical disparity can overcome this problem, but the resolution required appears to be physiologically implausible.
3. D. D. Hoffman and B. E. Flinchbaugh, "The interpretation of biological motion," Biol. Cybernet. **42,** 195–204 (1982).
4. A. Bobick, "A hybrid approach to structure-from-motion," in *Proceedings of the ACM Siggraph/Sigart Workshop on Motion* (Association for Computing Machinery, New York, 1982), pp. 91–109.
5. S. Ullman, *The Interpretation of Visual Motion* (MIT U. Press, Cambridge, Mass., 1979).
6. K. Prazdny, "Egomotion and relative depth map from optical flow," Biol. Cybernet. **36,** 87–102 (1980).
7. R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," Tech. Rep. R-921 (University of Illinois Coordinated Science Laboratory, Urbana, Ill., 1981).
8. E. Levison and R. Sekuler, "A two-dimensional analysis of direction-specific adaptation," Vision Res. **20,** 103–107 (1980).
9. C. A. Burbeck and D. Regan, "Independence of orientation and size in spatial discriminations," J. Opt. Soc. Am. **73,** 1691–1694 (1983).
10. C. Wheatstone, "Contributions to the physiology of vision," Phil. Trans. B. **13,** 371–394 (1838).
11. W. Richards, "Response functions for sine and square-wave modulations of disparity," J. Opt. Soc. Am. **62,** 907–911 (1972).
12. K. I. Beverley and D. Regan, "Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space," J. Physiol. **235,** 17–29 (1973).
13. K. I. Beverley and D. Regan, "The relation between discrimination and sensitivity in the perception of motion in depth," J. Physiol. **249,** 387–398 (1975).
14. D. Regan and K. I. Beverley, "Looming detectors in the human visual pathway," Vision Res. **18,** 415–421 (1978).
15. D. Regan and K. I. Beverley, "Visual responses to changing size and to sideways motion for different directions of motion in depth: linearization of visual responses," J. Opt. Soc. Am. **70,** 1289–1296 (1980).
16. M. Cynader and D. Regan, "Neurons in cat parastriate cortex sensitive to the direction of motion in three-dimensional space," J. Physiol. **274,** 549–569 (1978).
17. D. Regan and M. Cynader, "Neurons in cat visual cortex tuned to the direction of motion in depth: effect of stimulus speed," Invest. Opthalmol. **22,** 535–550 (1982).
18. D. Regan, K. I. Beverley, and M. Cynader, "The visual perception of motion in depth," Sci. Amer. **241,** 136–151 (1978).
19. D. Regan and K. I. Beverley, "Binocular and monocular stimuli for motion in depth: changing disparity and changing size feed the same motion-in-depth stage," Vision. Res. **19,** 1331–1342 (1979).
20. W. Richards and H. Lieberman, "A correlation between stereo ability and the recovery of structure-from-motion," Am. J. Optom. (to be published).
21. Vertical head motion with stereopsis appears no better than head motion with monocular viewing.
22. W. E. L. Grimson, "Binocular shading and visual surface reconstruction," A. I. Memo No. 697 (Massachusetts Institute of Technology, Cambridge, Mass., 1982).
23. W. A. Richards, J. M. Rubin, and D. D. Hoffman, "Equation counting and the interpretation of sensory data," Perception **11,** 557–576 (1982); A. I. Memo No. 614 (Massachusetts Institute of Technology, Cambridge, Mass., 1981).