Inferring the relative three-dimensional positions of two moving points

D. D. Hoffman

School of Social Sciences, University of California, Irvine, California 92717

B. M. Bennett

Department of Mathematics, University of California, Irvine, California 92717

Received June 5, 1984; accepted October 1, 1984

We show that four orthographic projections of two rigidly linked points are compatible with at most four interpretations of the relative three-dimensional positions of the points if the points rotate about a fixed axis—even when the points as a system undergo arbitrary rigid translations. A fifth view (projection) yields a unique interpretation and makes zero the probability that randomly chosen image points will receive a three-dimensional interpretation. Assuming that the points rotate at a constant angular velocity, instead of adding a fifth view, also yields a unique interpretation and makes zero the probability that randomly chosen image points will receive a three-dimensional interpretation.

1. INTRODUCTION

Psychophysical experiments by Johansson[1] and others indicate that human observers can perceive the relative three-dimensional (3-D) positions and motions of moving points from displays containing as few as two points. This ability is not explained by most current theoretical accounts of the recovery of 3-D structure from two-dimensional motion because most accounts require more than two points or place excessive restrictions on the 3-D motions that can be analyzed. Among the analyses requiring more than two points is that of Hoffman and Bennett,[2] who prove that three orthographic projections[3] of three points that rotate rigidly about a fixed axis are compatible with at most one 3-D interpretation (plus an orthographic projection). Similarly, Ullman[4] has shown that three orthographic projections of four noncoplanar points in a rigid configuration are compatible with at most one 3-D interpretation (plus reflection). Among the analyses requiring only two points but placing excessive restrictions on the 3-D motions that can be analyzed is that of Hoffman and Flinchbaugh,[5] who prove that three orthographic projections of two points that are constrained to rotate rigidly in a single plane (not necessarily parallel to the image plane) are compatible with at most one 3-D interpretation (plus reflection). Similarly, Hoffman and Bennett[2] have shown that three views of two rigidly linked points are compatible with at most one interpretation (plus reflection) if the points are constrained to rotate at constant angular velocity about a fixed axis that is parallel to the image plane.

A fundamental problem in the recovery of 3-D structure from image motion is the intrinsic ambiguity: there are always an infinite number of 3-D interpretations compatible with the image motion, regardless of the number of successive frames available and regardless of the type of projection (e.g., orthographic versus perspective). To obtain a unique 3-D interpretation one must always exploit some constraint or restriction on the possible 3-D motions. In this paper we explore the constraint of rigid fixed axis motion, since work by Ullman[4] indicates that rigidity alone is insufficient to obtain a unique 3-D interpretation from the motions of only two points. We show in Section 2 that four views of two points rotating rigidly about a fixed axis (an axis not parallel to or orthogonal to the image plane) and undergoing arbitrary translations (the same translation for each point) are compatible with at most four interpretations of the relative 3-D positions of the points. We then note that a fifth view yields a unique interpretation and makes zero the probability that randomly chosen image points will yield a 3-D interpretation. In Section 3 we show, using upper semicontinuity techniques, that imposing the additional constraint that the points rotate at a constant angular velocity, instead of adding a fifth view, also yields a unique interpretation that is correct with probability one. All the proofs yield closed-form solutions.

The equations studied here are amenable to solution by the techniques of nonlinear programming, making it possible to use them for the design of noise-insensitive algorithms for machine vision systems. Of course, the closed-form solutions presented later in the paper are unsuitable as machine vision algorithms—they are presented only to prove that in fact the equations have a unique solution. However, the equations themselves can be combined into an objective function that is minimized by using any of several nonlinear optimization techniques. An example of this is given by Reuman and Hoffman,[6] who device noise-insensitive algorithms for the equations studied by Hoffman and Flinchbaugh.[5]

2. ARBITRARY ANGULAR ACCELERATION

In this section we prove the following claim:

- Four orthographic projections of two rigidly linked points are compatible with at most four interpretations (plus reflections) of their relative 3-D positions and motions if the

0740-3232/85/020350-04$02.00

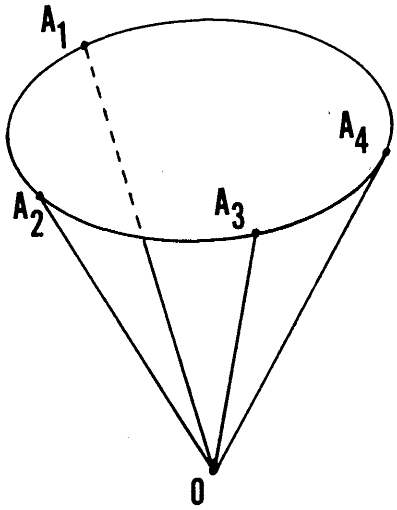

Fig. 1. Geometry underlying the proofs of Sections 2 and 3.

points rotate about a fixed axis—even if the system of points undergoes arbitrary translations. Adding a fifth view yields a unique interpretation and makes zero the probability that randomly chosen points will lead to any interpretation.

Proof. Call the two points O and A. Let $\mathbf{a}_i$ be the vector (in three dimensions) between O and A in view i ($i = 1, \ldots, 4$), and, without loss of generality, let O lie on the origin in each view as shown in Fig. 1. Because the two points are in a rigid configuration we expect that the length of the vector from O to A remains constant over all four views. Consequently we can write

$$\mathbf{a}_1 \cdot \mathbf{a}_1 = \mathbf{a}_2 \cdot \mathbf{a}_2, \tag{2.1a}$$

$$\mathbf{a}_1 \cdot \mathbf{a}_1 = \mathbf{a}_3 \cdot \mathbf{a}_3, \tag{2.1b}$$

$$\mathbf{a}_1 \cdot \mathbf{a}_1 = \mathbf{a}_4 \cdot \mathbf{a}_4. \tag{2.1c}$$

In addition we expect that the vectors $\mathbf{a}_i$ should all lie on a cone whose vertex is at O. Consequently we can write that the heads of the vectors $\mathbf{a}_i$ are coplanar:

$$(\mathbf{a}_1 - \mathbf{a}_2) \cdot [(\mathbf{a}_1 - \mathbf{a}_3) \times (\mathbf{a}_1 - \mathbf{a}_4)] = 0. \tag{2.2}$$

To solve these four equations it is useful to express the $\mathbf{a}_i$'s in terms of components. Let $\mathbf{a}_i = (x_i, y_i, z_i)$. Assume that the line of sight lies along the z axis. Then the x_i's and y_i's are known directly from the views. The four z_i's are unknown and must be solved for.

Equations (2.1) can be expressed in terms of components as

$$z_1^2 - z_2^2 + c_1 = 0, \tag{2.3a}$$

$$z_1^2 - z_3^2 + c_2 = 0, \tag{2.3b}$$

$$z_1^2 - z_4^2 + c_3 = 0. \tag{2.3c}$$

Equation (2.2) can be expressed in terms of components as

$$c_4 z_1 + c_5 z_2 + c_6 z_3 + c_7 z_4 = 0, \tag{2.4}$$

where

$$c_1 = x_1^2 + y_1^2 - x_2^2 - y_2^2, \tag{2.5a}$$

$$c_2 = x_1^2 + y_1^2 - x_3^2 - y_3^2, \tag{2.5b}$$

$$c_3 = x_1^2 + y_1^2 - x_4^2 - y_4^2, \tag{2.5c}$$

$$c_4 = x_3y_4 - x_2y_4 - x_4y_3 + x_2y_3 + x_4y_2 - x_3y_2, \tag{2.5d}$$

$$c_5 = x_1y_4 - x_3y_4 + x_4y_3 - x_1y_3 - x_4y_1 + x_3y_1, \tag{2.5e}$$

$$c_6 = x_2y_4 - x_1y_4 - x_4y_2 + x_1y_2 + x_4y_1 - x_2y_1, \tag{2.5f}$$

$$c_7 = x_1y_3 - x_2y_3 + x_3y_2 - x_1y_2 - x_3y_1 + x_2y_1. \tag{2.5g}$$

Use Eq. (2.3) to eliminate z_2, z_3, and z_4 from Eq. (2.4):

$$c_4 z_1 \pm c_5(z_1^2 + c_1)^{1/2} \pm c_6(z_1^2 + c_2)^{1/2} \pm c_7(z_1^2 + c_3)^{1/2} = 0. \tag{2.6}$$

Recalling that $x \pm y = 0$ if $(x + y)(x - y) = 0$, we can rewrite Eq. (2.6) as a product of eight polynomials. Expanding and simplifying this product (with the help of MACSYMA) we find that

$$d_1 z_1^8 + d_2 z_1^6 + d_3 z_1^4 + d_4 z_1^2 + d_5 = 0, \tag{2.7}$$

where the d_i's are functions entirely of the image data (x's and y's). Equation (2.7) is nonhomogeneous of fourth degree in z_1^2 and can be solved in closed form for z_1^2. Knowing z_1, one can solve for the remaining z_i's by using Eqs. (2.3). This shows that in general four views of two points spinning rigidly about a fixed axis are compatible with at most four interpretations plus reflections.

A fifth view yields a unique interpretation (plus reflection) for the following reason. The tips of the vectors $\mathbf{a}_i$ (i.e., the points A_i) all lie on some circle in $\mathcal{R}^3$. This circle projects to an ellipse in the image. Five points uniquely determine an ellipse. The ellipse, in turn, determines two circles in $\mathcal{R}^3$ that are reflections of each other. Note that the projection of the point O must lie on the line passing through the minor axis of the ellipse for there to be an interpretation. The probability of this happening for randomly chosen points in the plane is zero. Consequently the probability is zero that randomly chosen points will lead to a 3-D interpretation.

3. CONSTANT ANGULAR VELOCITY

In this section we prove the following claim:

- Given four orthographic projections of two points rotating rigidly and at a constant angular velocity about a fixed axis, there is a unique interpretation (plus reflection) for the 3-D structure and motion that is compatible with the projections. Furthermore the probability is zero that four views of two points chosen at random will lead to an interpretation.

Referring again to Fig. 1, it is clear that the constraint of constant angular velocity can be expressed by the following equations:

$$\mathbf{a}_1 \cdot \mathbf{a}_2 = \mathbf{a}_2 \cdot \mathbf{a}_3, \tag{3.1a}$$

$$\mathbf{a}_1 \cdot \mathbf{a}_2 = \mathbf{a}_3 \cdot \mathbf{a}_4. \tag{3.1b}$$

In terms of components these become

$$z_1 z_2 - z_2 z_3 + d_6 = 0, \tag{3.2a}$$

$$z_1 z_2 - z_3 z_4 + d_7 = 0, \tag{3.2b}$$

where

$$d_6 = x_1x_2 + y_1y_2 - x_2x_3 - y_2y_3, \quad (3.3a)$$

$$d_7 = x_1x_2 + y_1y_2 - x_3x_4 - y_3y_4. \quad (3.3b)$$

Equations (3.2), (2.3), and (2.4) give a system of six polynomial equations in the four variables z_i. In this section we prove that generically this system of equations has no solutions, thus demonstrating that the probability is zero that (eight) randomly chosen image points will have a 3-D interpretation. The proof consists in providing a single choice of four views of two points [i.e., eight (x, y) pairs] for which the system has no solutions. Next we prove that if the system of equations has solutions, then generically the number of solutions is one (plus a reflection), thus demonstrating that the 3-D interpretation is unique. Again the proof consists in providing a single choice of four views of two points for which the system has one solution. Proof by concrete example in this fashion is licensed in both cases by the upper semicontinuity of the number of solutions to a parameterized family of algebraic equations in projective space, i.e., the fact that the number of solutions is an upper semicontinuous function of the parameters for the Zariski topology (see Appendix A for a brief explanation of upper semicontinuity).[7-10]

Parenthetically, we should point out an easy way to find solutions common to Eqs. (2.3), (2.4), and (3.2). Use Eqs. (2.3a) and (2.3b) to eliminate z_2 and z_3 from Eq. (3.2a):

$$\pm z_1(z_1^2 + c_1)^{1/2} \pm [(z_1^2 + c_1)(z_1^2 + c_2)]^{1/2} + d_6 = 0. \quad (3.4)$$

Expanding and simplifying Eq. (3.4) gives

$$d_8z^4 + d_9z^2 + d_{10} = 0, \quad (3.5)$$

where

$$d_8 = c_2^2 - 4d_6^2, \quad (3.6a)$$

$$d_9 = c_1c_2^2 - 2c_1d_6^2 - c_2d_6^2, \quad (3.6b)$$

$$d_{10} = d_6^4 - 2c_1c_2d_6^2 + c_1^2c_2^2. \quad (3.6c)$$

Equation (3.5) is a second-degree nonhomogeneous equation in z_1^2 and gives two solutions (plus reflections) for z_1. Having explicit values for z_1, it is an easy matter to compute the remaining z_i's using Eqs. (2.3) and then to check if the solutions satisfy Eqs. (2.4) and (3.2b). We wrote a computer program that does just this and then used it to analyze the two test cases discussed below. This ends the parenthetical remark.

To prove that generically Eqs. (2.3), (2.4), and (3.2) have no solutions we note that they have no solutions for the following choice of image data:

$$x_1 = 5.280761, \quad y_1 = 8.863270,$$

$$x_2 = 4.523088, \quad y_2 = 8.457234,$$

$$x_3 = 3.169634, \quad y_3 = 6.894401,$$

$$x_4 = 2.165064, \quad y_4 = 4.500000. \quad (3.7)$$

This proves that the probability of "false targets," i.e., the probability that randomly moving points will be assigned a 3-D interpretation by Eqs. (2.3), (2.4), and (3.2), is zero.

To prove that if these equations have solutions then generically they have but one solution (plus reflection), we note that they have but one solution for the following set of image data:

$$x_1 = 3.968492, \quad y_1 = 8.457234,$$

$$x_2 = 2.415874, \quad y_2 = 6.894401,$$

$$x_3 = 1.263478, \quad y_3 = 4.500000,$$

$$x_4 = 0.6503015, \quad y_4 = 1.562834. \quad (3.8)$$

For these data the common solutions are $(z_1, z_2, z_3, z_4) = (-6.53653, -8.75390, -10.3997, -11.2754)$ and $(6.53653, 8.75390, 10.3997, 11.2754)$. This proves the uniqueness of the 3-D interpretation under these equations. (See Appendix A for more detail.)

4. CONCLUSION

We conclude that four views of two points undergoing arbitrary 3-D translations are compatible with at most four interpretations of the relative 3-D positions of the points if the points are rotating rigidly about a fixed axis. A fifth view yields a unique interpretation and makes zero the probability that randomly chosen points will receive a 3-D interpretation. Assuming that the points rotate at a constant angular velocity, instead of adding a fifth point, also yields a unique interpetation that is correct with probability one.

It should be reiterated that the examples examined in Section 3 are not merely for illustration but actually constitute rigorous proofs because of the upper semicontinuity result cited in that section.

APPENDIX A

Our technique of proof for the claim of Section 3 is based on the *principle of upper semicontinuity*, which may be stated for our purposes as follows:

Let S be a system of algebraic (polynomial) equations in complex projective space of arbitrary dimension. Suppose that the coefficients of the equations in S depend algebraically on some parameters, which vary in a complex space C^n. Then the function assigning to each point $P \in C^n$ (i.e., to each set of values of the parameters) the number $N(P)$ of solutions (including multiplicities) to the equations S for that choice of parameter values is upper semicontinuous in the Zariski topology on C^n.

In the Zariski topology the closed sets are algebraic varieties (solution sets of polynomial equations). Recall that a function is upper semicontinuous if the locus of points where it assumes a value greater than or equal to some given value is a closed set. Hence the upper semicontinuity principle translates into the following: Given any integer m, the set T_m of points $P \in C^n$, where $N(P) \geq m$ is the solution set of a family of polynomial equations.

The importance of this principle here follows from the fact that *proper* Zariski closed subsets of C^n (proper algebraic varieties in C^n) have measure zero in C^n, and similarly those points on the variety having real coordinates form a measure zero subset of the set of all points in C^n with real coordinates, i.e., of $R^n \subset C^n$. Thus the probability is zero that a randomly chosen real point of C^n will lie on a given proper algebraic variety.

In our case, the system of equations S is the system consisting of Eqs. (3.2), (2.3), and (2.4), and the parameter space is C^8 with coordinates (x_i, y_i) $i = 1, \ldots, 4$. From the upper semicontinuity principle we can conclude that the locus $V \in C^8$ consisting of those (x_i, y_i) for which Eqs. (3.2), (2.3), and (2.4) admit of at least one solution is an algebraic variety in C^8. Therefore to show that the probability of false targets is zero, we only have to show that V is a *proper* subvariety of C^8, i.e., it suffices to produce one point in C^8 that is not in V. This was done in Section 3.

The proof of uniqueness also uses upper semicontinuity techniques. Let T_m and V be as above. We have $T_0 = C^8$, and T_1 is the variety in C^8 that we called V. In our case, since solutions come in pairs corresponding to reflection about the image plane, $T_1 = T_2$, $T_3 = T_4$, and so forth. To prove that the probability of unique interpretation is one, we want to show the following: If $V(R)$ denotes the set of points in V with real coordinates, and $T_4(R)$ denotes the set of points in T_4 with real coordinates, then $T_4(R)$ has measure zero in $V(R)$.

As with false targets, the proof is based on producing a point in $V(R)$ that is not in $T_4(R)$—which was done in Section 3. This implies (by the upper semicontinuity principle) that T_4 is a proper algebraic subvariety of V. One cannot conclude from this alone that T_4 has measure zero in V, for it is *a priori* possible that V may be *reducible*, i.e., it may consist of several components, say of equal dimension, one or more of which constitute T_4. It can be shown in our case, using the approach of Hoffman and Bennett,[2] that all components of V other than T_4 have dimension less than T_4, so that the uniqueness claim does follow from the test point produced in Section 3. Note that this consideration does not arise in the false targets proof, for there we are starting with $T_0 = C^8$, which is irreducible.

ACKNOWLEDGMENTS

We thank J. Beusman for useful discussions and the Fairchild Laboratory for Artificial Intelligence Research (FLAIR) for support. This paper is based on research supported by the National Science Foundation under grant no. IST-8413560.

REFERENCES

1. G. Johansson, "Visual motion perception," Sci. Am. **232,** 76–88 (1975).
2. D. Hoffman and B. Bennett, "The computation of structure from fixed axis motion: rigid structures" (submitted to Comput. Vision Graphics Image Process, 1984). Available upon request.
3. Orthographic projection takes the 3-D coordinates of a point into its 2-D coordinates by the map $(x, y, z) \rightarrow (x, y)$. Perspective projection, in contrast, takes the 3-D coordinates of a point into its 2-D coordinates by the map $(x, y, z) \rightarrow (x/z, y/z)$. 3-D translations are analyzed trivially under orthographic projection since translations parallel to the image plane are congruent in the image and in three dimensions, and translations in depth have no affect on the image.
4. S. Ullman, "The interpretation of structure from motion," Proc. R. Soc. London Ser. B **203,** 405–426 (1979).
5. D. Hoffman and B. Flinchbaugh, "The interpretation of biological motion," Biol. Cybernet. **42,** 195–204 (1982). (Also available as MIT AI Memo 592.)
6. S. Reuman and D. Hoffman, "Regularities of nature: their role in the interpretation of visual motion," in *From Pixels to Predicates: Recent Advances in Computational Vision*, A. Pentland, ed. (Ablex Publishing, in press).
7. A. Grothendieck, "Éléments de géometrie algebrique. III. Étude cohomologique des faisceaux cohérents," Publications Mathématiques des l'Institut des Hautes Études Scientifiques **11** (1961).
8. M. Atiyah-Macdonald, *Introduction to Commutative Algebra* (Addison-Wesley, Reading, Mass., 1969).
9. R. Hartshorne, *Algebraic Geometry* (Springer-Verlag, New York, 1977).
10. B. Bennett, "Semi-continuity and perception" (in preparation).