Inferring 3D structure
from three points in rigid motion

B. M. Bennett

Department of Mathematics, University of California,

Irvine, California 92717

D. D. Hoffman

Department of Cognitive Science, University of California,

Irvine, California 92717

We prove the following: Given four (or more) orthographic views of three points then (a) the views almost surely have no rigid interpretation but (b) if they do then they almost surely have at most thirty-two rigid interpretations. Part (a) means that the measure of "false targets", viz., the measure of nonrigid motions that project to views having rigid interpretations, is zero. Part (b) means that rigid interpretations, when they exist, are not unique. Uniqueness of interpretation can be obtained if a point is added, but not if views are added. Our proof relies on an upper semicontinuity theorem for proper mappings of complex algebraic varieties. We note some psychophysical motivations of the theory.

1. INTRODUCTION

Although our retinal images are two dimensional, we see the world in three dimensions. Among the sources of information used by human vision to infer the third dimension is visual motion. A large body of theoretical work[1−19] and psychophysical work [20−32] has explored how human vision (or robotic vision systems) might use visual motion to infer depth.

The obvious problem to be overcome in such an inference is the infinite ambiguity of retinal images, whether dynamic or static. Given any collection of images, there are always an uncountable collection of three-dimensional interpretations whose projections onto the retina would lead to the given images[16]. Thus various constraints must be employed to restrict the possible interpretations to a few, or to one.

Among the constraints examined in the literature are rigid motion[1,4,7,8,12,14,17], rigid fixed-axis motion[10,19], nonrigid fixed-axis motion[2], planar rigid motion[10], rigid motion about a vertical axis[15], certain bending motions[12], and rigid motion that conserves angular momentum[3]. This latter constraint led to the following theorem: Given three distinct orthographic projections of three points, (a) the projections are almost surely incompatible with any three-dimensional interpretation in which the points move rigidly and conserve angular momentum, but (b) if the projections are compatible with such an interpretation then, generically, they are compatible with at most two such interpretations.

It was our desire to submit this theorem to psychophysical tests of its psychological plausibility that led to the result presented in this paper. We needed to show human subjects displays of three points undergoing rigid motion that conserves angular momentum, in order to find out if subjects can infer properties of the inertia tensor from such displays, or simply to find out if they could discriminate such displays from displays showing more general rigid motions of three points. We discovered that there are no theorems specifying what can be inferred from the motions of just three points using the constraint of rigidity alone; thus there

was no theoretical work on which we could base "catch" trials consisting of three points undergoing purely rigid motion. The well-known result of Ullman[17] is close to what we needed, but Ullman's theorem requires four noncoplanar points, whereas we needed a result involving only three points. We had reported elsewhere[10] that three orthographic views of three points are always consistent with sixty four rigid interpretations, but this theorem was not sufficient for our purposes because it does not specify conditions under which false targets (nonrigid motions that project to views having rigid interpretations) are rare. Instead it implies that with three views of three points false targets are ubiquitous, making three views of three points useless for inferences of rigid structures. Thus we were forced to investigate whether or not any reasonable inferences of rigid structure could be made from more than three views of three points. In the next section we describe what we found.

2. RIGIDITY THEOREM

In this section we prove the following theorem.

Theorem 1.   Given four (or more) orthographic views of three points then (a) the views almost surely have no rigid interpretation but (b) if they do then they almost surely have at most thirty two rigid interpretations.

Proof. Without loss of generality, we take one of the three points to be the origin of an x, y, z coordinate system with corresponding unit vectors $\hat{e}_x$, $\hat{e}_y$, and $\hat{e}_z$. The positions of the other two points are represented in this coordinate system: the position of point m in view i is the vector $\mathbf{a}_{m,i} = x_{m,i}\hat{e}_x + y_{m,i}\hat{e}_y + z_{m,i}\hat{e}_z$. The condition that the three points move rigidly

over four views, i.e., that they maintain constant interpoint distances in space, is given by the following nine equations:

$$\mathbf{a}_{m,i} \cdot \mathbf{a}_{n,i} = \mathbf{a}_{m,j} \cdot \mathbf{a}_{n,j}, \qquad 1 \leq m \leq n \leq 2, \quad 1 \leq i \leq 3, \quad j = i+1. \tag{2}$$

The six diagonal equations ($m = n$) state that the lengths of the position vectors remain constant over the views. The three off-diagonal equations ($m \neq n$) state that the angle between the two position vectors remains constant over the views. Equations (2) define an algebraic variety E embedded in a Euclidean space $X = \{\mathbf{a}_{m,i} \colon m = 1, 2; \; i = 1, 2, 3, 4\} = \mathbf{R}^{24}$.

We take the image plane to be the x, y plane, with the z axis pointing along the line of sight. The orthographic projection onto the image plane of the position vector $\mathbf{a}_{m,i}$ is the vector $\mathbf{b}_{m,i} = x_{m,i}\hat{e}_x + y_{m,i}\hat{e}_y$. Thus

$$\mathbf{a}_{m,i} = \mathbf{b}_{m,i} + z_{m,i}\hat{e}_z. \tag{3}$$

Substituting (3) into (2) gives the nine quadratic equations

$$z_{m,i}z_{n,i} - z_{m,j}z_{n,j} + \mathbf{b}_{m,i} \cdot \mathbf{b}_{n,i} - \mathbf{b}_{m,j} \cdot \mathbf{b}_{n,j} = 0, \qquad 1 \leq m \leq n \leq 2, \quad 1 \leq i \leq 3, \quad j = i+1,$$
$$\tag{4}$$

where the $\mathbf{b}_{m,i}$ are known from the images and the eight $z_{m,i}$ must be found as the solutions to (4). Of course the number of such solutions depends on the "parameters" $\mathbf{b}_{m,i}$. The collection of possible parameters forms a Euclidean space $Y = \{\mathbf{b}_{m,i} \colon m = 1, 2; \; i = 1, 2, 3, 4\} = \mathbf{R}^{16}$. Part (a) of Theorem 1 states that the collection $S \subset Y$ of parameters for which (4) has at least one solution for the $z_{m,i}$ is a set of measure zero for the Lebesgue measure on Y.

The spaces X and Y are naturally related by a projection map $\pi \colon X \to Y$ given by $\{\mathbf{a}_{m,i}\} \mapsto \{\mathbf{b}_{m,i}\}$. Using π, we can characterize the collection of parameters $S \subset Y$ as follows: For $y \in Y - S$ the set $\pi^{-1}(y) \cap E$, i.e., the set of real solutions to (4) for the $z_{m,i}$, is empty; for $y \in S$ the set $\pi^{-1}(y) \cap E$ is nonempty. From this it is clear that $\pi(E) = S$. The situation so far

can be summarized in the following diagram.

$$\text{all 3D motions} = \quad X \quad = \mathbf{R}^{24} \quad \supset \quad E \quad = \text{rigid 3D motions}$$

$$\downarrow \pi \qquad\qquad\qquad \downarrow \pi \qquad\qquad\qquad\qquad (5)$$

$$\text{all 2D displays} = \quad Y \quad = \mathbf{R}^{16} \quad \supset \quad S \quad = \text{`rigid' 2D displays}$$

To prove part (a) of Theorem 1, i.e., to prove that S has Lebesgue measure zero in Y, we use the following key fact. Let $Y_{\mathbf{C}} = \mathbf{C}^{16}$ be the complexification of Y, and $E_{\mathbf{C}}$ the complexification of E. It is shown elsewhere[3] that varieties $E_{\mathbf{C}}$ which arise as complex solution spaces of systems of equations of the type of (2) have the following property: their projective completion in the $z_{m,i}$ variables has no more points than $E_{\mathbf{C}}$ itself. Thus

$$(\dagger) \qquad E_{\mathbf{C}} \text{ is a family of projective varieties parametrized by } Y_{\mathbf{C}}.$$

Let $f \colon Y_{\mathbf{C}} \to \mathbf{N}$ be the function which assigns to each set of parameters $y \in Y_{\mathbf{C}}$ the corresponding number of real and complex solutions in $E_{\mathbf{C}}$, counted with multiplicities, to equations (4). Because of the property ($\dagger$) above, the function f is *upper semicontinuous* in the Zariski topology[3]. In the Zariski topology the closed sets are closed algebraic varieties, i.e., solution sets to systems of polynomial equations. To say that f is upper semicontinuous in the Zariski topology is to say that if at some point $y \in Y_{\mathbf{C}}$ the function f takes the value n, then the locus of points on which f takes values greater than n is a Zariski closed set. Our interest here in these sets is that they have positive codimension in $Y_{\mathbf{C}}$ and therefore are of Lebesgue measure zero in $Y_{\mathbf{C}}$. Thus if we can produce a point $y \in Y_{\mathbf{C}}$ for which f has the value zero, then the set $S_{\mathbf{C}}$ of points in $Y_{\mathbf{C}}$ for which f has a value greater than zero is of Lebesgue measure zero in $Y_{\mathbf{C}}$. This will imply that the set S of points in Y for which f has a value greater than zero is also of Lebesgue measure zero in Y, proving part (a) of Theorem 1.

It has been shown elsewhere[3] that the six equations of (4) corresponding to the first three views (i.e., for which $1 \le i \le 2$) have, for generic parameters, precisely sixty four complex affine solutions for the six $z_{m,i}$, $(m = 1, 2; i = 1, 2, 3)$. Thus to check that equations (4) have no solutions at all for some parameter point $y \in Y_{\mathbf{C}}$ we can do the following: (1) Find all sixty four solutions to those equations for which $1 \le i \le 2$, (2) substitute these solutions into

the two remaining diagonal equations to compute the corresponding values for $z_{1,4}$ and $z_{2,4}$, (3) substitute these values into the final off diagonal equation to see if it is satisfied, and (4) if the final off diagonal equation is never satisfied we are done. We have done this with the help of Mathematica and found that for a random choice of parameters $\mathbf{b}_{m,i}$ the equations (4) have no solutions when checked in this manner. Thus, since f is upper semicontinuous, we conclude that $S_{\mathbf{C}}$ has measure zero in $Y_{\mathbf{C}}$.

To finish the proof of part (a) of Theorem 1 we must note that $S_{\mathbf{C}}$ having measure zero in $Y_{\mathbf{C}}$ implies that S has measure zero in Y. The technical arguments for this are available elsewhere[10].

We now turn to the proof of part (b) of Theorem 1. We have shown already that S has measure zero in Y. Part (b) states that the function f, restricted to S, generically takes a value of at most thirty two. We can once again use the upper semicontinuity of f, together with a test point, to prove this claim. The idea is to find a point of $S_{\mathbf{C}}$ for which the equations (4) have thirty two solutions, real or complex, for the $z_{m,i}$. By the upper semicontinuity of f this then shows that, except on a Zariski closed proper subset of $S_{\mathbf{C}}$, the function f takes a value of at most thirty two. Since this Zariski closed proper subset is of measure zero within $S_{\mathbf{C}}$, we conclude that generically on $S_{\mathbf{C}}$ there are at most thirty two solutions to equations (4). This will imply that generically on S there are at most thirty two solutions.

There is one important technical issue to be considered before taking this approach. That issue is the irreducibility of $S_{\mathbf{C}}$. Recall that an algebraic variety W is reducible if it can be expressed as a union of distinct subcomponents where no subcomponent is contained in any other and where each component is itself the zero set of a distinct set of polynomials. If $S_{\mathbf{C}}$ is reducible, then we must check a test point on each irreducible component of $S_{\mathbf{C}}$. This issue did not arise in our proof of part (a) because the entire space $Y_{\mathbf{C}} = \mathbf{C}^{16}$ is irreducible. Fortunately $S_{\mathbf{C}}$ is irreducible. The irreducibility of $S_{\mathbf{C}}$ follows from the irreducibility of $E_{\mathbf{C}}$ and the fact that $S_{\mathbf{C}} = \pi(E_{\mathbf{C}})$. To see that $E_{\mathbf{C}}$ is irreducible we first note that $E_{\mathbf{C}}$ is

isomorphic to $\mathbf{C}^6 \times \mathrm{SO}(3, \mathbf{C})^3$; in fact $\{\mathbf{a}_{m,i}\} \in E_{\mathbf{C}}$ iff there exists $\gamma_1, \gamma_2, \gamma_3 \in \mathrm{SO}(3, \mathbf{C})$ such that $\mathbf{a}_{m,k+1} = \gamma_k \mathbf{a}_{m,1}$, $k = 1, 2, 3$. Thus $E_{\mathbf{C}}$ is a product of irreducible varieties and is therefore irreducible.

Hence to show that generically on $S_{\mathbf{C}}$ the number of solutions to equations (4) is at most thirty two, it suffices to find a single test point in $S_{\mathbf{C}}$ for which these equations have thirty two solutions. This we have done, simply by taking three points in $\mathbf{R}^3$, applying three different elements of $\mathrm{SO}(3, \mathbf{R})$ (other than the identity) to the points, and projecting the results onto the x, y plane to obtain a point in $S_{\mathbf{C}}$.

Two issues remain before the proof of part (b) is complete. First we must be concerned with the multiplicity of each of the thirty two solutions obtained in the manner just described. If we want to use the upper semicontinuity of f to show that generically on $S_{\mathbf{C}}$ there are at most thirty two solutions then we need to show that each of the thirty two solutions we obtain for a particular test point has multiplicity one; it is the number of solutions counted with multiplicities, and not just the number of solutions alone, that is upper semicontinuous. One can check the multiplicities of the solutions in the obvious way: check the Jacobian of the equations (4) when evaluated at each of the thirty two solutions to see that each is nonsingular. This we have done with the help of Mathematica.

We have shown that generically on $S_{\mathbf{C}}$ the function f is at most thirty two. The final issue in the proof of part (b) is to note that this implies that generically on S the function f is also at most thirty two. The technical arguments for this are available elsewhere[10]. ∎

3. OBSERVER REPRESENTATION

Theorem 1 defines a class of visual inferences. The premises for these inferences are four

views of three points; the conclusions describe 3D structures in rigid motion. The abstract form of these inferences can be described as follows. The set of possible premises is the set $Y = \mathbf{R}^{16}$ of all possible three views of two vectors in the plane. The set of possible conclusions is the set $X = \mathbf{R}^{24}$ of all possible 3D interpretations for elements of Y. Those interpretations satisfying the rigidity constraint form a Lebesgue measure zero subset E of X. The conclusions X and premises Y are related by a function π given by $(x_{ij}, y_{ij}, z_{ij}) \mapsto (x_{ij}, y_{ij})$, and for each premise $y \in Y$ the set $\pi^{-1}(\{y\})$ is the set of all 3D conclusions compatible with the premise y. Those premises y which have at least one compatible conclusion that satisfies the rigidity constraint form a subset S of Y. Clearly $S = \pi(E)$. Moreover, S has Lebesgue measure zero in Y. Thus for most $y \in Y$ none of the compatible conclusions satisfy the rigidity constraint, and hence the probability of false targets for this inference is zero. For premises $s \in S$ the number of compatible rigid interpretations is, generically, thirty two. Therefore the conclusion associated to such an s is best thought of as a probability measure, say η_s, supported on these thirty two rigid interpretations. The weight given to a particular interpretation by this measure can be thought of as the frequency with which that interpretation is perceived, given that one is viewing the display s.

Thus the inference of structure from rigid motion examined here is specified by a six-tuple (X, Y, E, S, π, η). This six-tuple precisely satisfies the definition of *observer* given in observer theory[33,34]. According to the *observer thesis*[33,34] every perceptual capacity, whether instantiated in neurons or in silicon, can be described as an instance of a single formal structure, viz., the observer:

Definition 6.   An *observer* is a six-tuple (X, Y, E, S, π, η) where

1. X and Y are measurable spaces. E is an event of X. S is an event of Y. Points of X and Y are measurable.

2. π is a measurable map from X onto Y such that $\pi(E) = S$.

3. η is a markovian kernel that associates to each point s of S a probability measure on E

which gives the set $\pi^{-1}(s) \cap E$ a probability of one.

The present theory of structure from rigid motion is a specific example in support of the observer thesis.

ACKNOWLEDGEMENTS

REFERENCES

1. J. Aloimonos and A. Bandyopadhyay, "Perception of structure from motion: lower bound results," Univ. of Rochester Dept. of Comput. Sci. TR 158, (1985).

2. B. M. Bennett and D. D. Hoffman, "The computation of structure from fixed-axis motion: nonrigid structures," Biol. Cybern. **51,** 293–300 (1985).

3. B. M. Bennett, D. D. Hoffman, J. S. Kim, and S. N. Richman, "Inferring 3D structure from image motion: The constraint of Poinsot motion," Int. J. Comput. Vis. (in press).

4. B.M. Bennett, D.D. Hoffman, J.E. Nicola, and C. Prakash, "Structure from two orthographic views of rigid motion," J. Opt. Soc. Am. A, **6,** 1052–1069 (1989).

5. E. H. Carlton and R. N. Shepard, "Psychologically simple motions as geodesic paths: I. Asymmetric objects," J. Math. Psych. **34,** 127–188 (1990).

6. E. H. Carlton and R. N. Shepard, "Psychologically simple motions as geodesic paths: II. Symmetric objects," J. Math. Psych. **34,** 189–228, (1990).

7. O. D. Faugeras and S. Maybank, "Motion from point matches: Multiplicity of solutions," Inter. J. Comput. Vis. **4,** 225–246 (1990).

8. N. Grzywacz and E. Hildreth, "Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations," J. Opt. Soc. Am. A **4,** 503–518 (1987).

9. D. D. Hoffman and B. M. Bennett, "Inferring the relative three-dimensional positions of two moving points," J. Opt. Soc. Am. A **2,** 350–353 (1985).

10. . D. Hoffman and B. M. Bennett, "The computation of structure from fixed-axis motion: rigid structures," Biol. Cybern. **54,** 71–83 (1986).

11. D. D. Hoffman and B. E. Flinchbaugh, "The interpretation of biological motion," Biol. Cybern. **42,** 197–204 (1982).

12. T. Huang and C. Lee, "Motion and structure from orthographic projections," IEEE Trans. Patt. Anal. Mach. Intell. **11,** 536–540 (1989).

13. J. Koenderink and A. van Doorn, "Depth and shape from differential perspective in the presence of bending deformations," J. Opt. Soc. Am. A **3,** 242–249 (1986).

14. J. Koenderink and A. van Doorn, "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer," Optica Acta **22,** 773–791 (1975).

15. E. Kruppa, "Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung," Akademie der Wissenschaften in Wien: Mathematisch-naturwissenschaftliche Klasse Sitzungsberichte **122,** 1939–1948 (1913).

16. H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two perspective projections," Nature **293,** 133–135 (1981).

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

18. A. Waxman and K. Wohn, "Contour evolution, neighborhood deformation, and image flow: Textured surfaces in motion," in *Image Understanding 1985–1986,* W. Richards and S. Ullman, eds. (Ablex, Norwood, New Jersey, 1987), pp. 72–98.

19. J. A. Webb and J. K. Aggarwal, "Structure from motion of rigid and jointed objects," Artif. Intell. **19,** 107–130 (1982).

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

21. M. Braunstein, D. Hoffman, L. Shapiro, G. Andersen and B. Bennett, "Minimum points and views for the recovery of three-dimensional structure," J. Exper. Psychol.: Human Percept. and Perfor. **13,** 335–343 (1987).

22. J. J. Gibson and E. J. Gibson, "Continuous perspective transformations and the perception of rigid motion," J. Exper. Psychol. **54,** 129–138 (1957).

23. D.L. Gilden, D.R. Proffitt, "Understanding collision dynamics," J. Exper. Psychol.: Human Percept. Perform. **15,** 372–383 (1989).

24. B. Green, "Figure coherence in the kinetic depth effect," J. Exper. Psychol. **62,** 272–282 (1961).

25. M.K. Kaiser, D.R. Proffitt, "The development of sensitivity to causally relevant dynamic information," Child Development **55,** 1614–1624 (1984).

26. M.K. Kaiser, D.R. Proffitt, "Observers' sensitivity to dynamic anomalies in collisions," Percept. Psychophys. **42,** 275–280 (1987).

27. J. S. Lappin, J. F. Donner and B. Kottas, "Minimal conditions for the visual detection of structure and motion in three dimensions," Science **209,** 717–719 (1980).

28. D.R. Proffitt, D. Gilden, M.K. Kaiser, S. Whelan, "The effect of configural orientation on perceived trajectory in apparent motion," Percept. Psychophys. **45,** 465–474 (1988).

29. V. S. Ramachandran, S. Cobb and D. Rogers-Ramachandran, "Perception of 3-D structure from motion: The role of velocity gradients and segmentation boundaries," Percept. Psychophys. **44,** 390–393 (1988).

30. J. T. Todd, R. A. Akerstrom, F. D. Reichel and W. Hayes, "Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors," Percept. Psychophys. **43,** 179–188 (1988).

31. J.T. Todd, W.H. Warren, "Visual perception of relative mass in dynamic events," Perception **11,** 325–335 (1982).

32. H. Wallach and D O'Connell, "The kinetic depth effect," J. Exper. Psychol. **45,** 205–217 (1953).

33. B.M. Bennett, D.D. Hoffman, C. Prakash, *Observer mechanics,* (Academic Press, New York, 1989).

34. B.M. Bennett, D.D. Hoffman, C. Prakash, "Unity of perception," Cognition **38,** 295–334 (1991).