Structure from two orthographic views of rigid motion

B. M. Bennett

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

D. D. Hoffman and J. E. Nicola

Department of Cognitive Science, University of California, Irvine, Irvine, California 92717

C. Prakash

Department of Mathematics, California State University, San Bernardino, California 92407

Received March 8, 1988; accepted January 6, 1989

We study the inference of rigid three-dimensional interpretations for the structure and motion of four or more moving points from but two orthographic views of the points. We develop an algorithm to determine whether image data are compatible with a rigid interpretation. As a corollary of this result we find that the measure of false targets (roughly, nonrigid objects that appear rigid) is zero. We find that if the two views have at least one rigid interpretation, then in fact there is a canonical one-parameter family of rigid interpretations; we show how to compute this family, and we describe precisely how the rigid interpretations vary within it. Since only two views are used, this analysis is relevant also to stereo vision.

1. INTRODUCTION

Psychophysical experiments by Braunstein *et al.*[1] and others have revealed a remarkable capacity of human vision. If one presents subjects with two-dimensional (2-D) motion displays containing as few as two frames (say, videotape frames) of four feature points, subjects can reliably decide whether the 2-D motions of the points have any rigid three-dimensional (3-D) interpretations: i.e., subjects can decide, for each 2-D motion display, whether there exists some rigid motion of the points in $\mathbf{R}^3$ and some choice of viewing direction that would give rise to that display. (Points in $\mathbf{R}^3$ move rigidly only if their 3-D interpoint distances remain constant.) Moreover, when viewing displays that have rigid interpretations, subjects report perceiving a single rigid interpretation.

These abilities, in their full generality, are not explained by current mathematical theories that specify sufficient conditions for the inference of structure from motion. There are diverse reasons for this failure of explanation; among them are that (1) some accounts require more than two frames to reach a unique rigid interpretation; (2) some accounts restrict the kinds of rigid 3-D motion that can be inferred, e.g., allowing only rigid interpretations in which the points move about a single fixed axis in $\mathbf{R}^3$ [rigid, fixed-axis (RFA) interpretations], or in which the points move about a single fixed axis at a constant angular velocity (CAV interpretations), or in which pairs of points move only in a single plane (planar interpretations); and (3) some accounts yield nonrigid interpretations.

For example, among the accounts requiring more than two views is that of Ullman,[2] who proved that, if there is at least one rigid interpretation compatible with a given set of three frames, with each frame containing four points, then there are, in general, at most two rigid interpretations. Among the accounts restricting the kinds of rigid 3-D motion that can be inferred is that of Hoffman and Bennett,[3] who proved that, if there is at least one RFA interpretation compatible with a given set of five frames, with each frame containing two points, then there are, in general, only two RFA interpretations. In the same paper they proved that if there are any CAV interpretations compatible with a given set of four frames, with each frame containing two points, then there are, in general, only two CAV interpretations. In another paper[4] they proved that if there are any RFA interpretations compatible with a particular set three frames, each frame containing three points, then there are, in general, just two RFA interpretations. An algorithm for constructing RFA interpretations was presented by Webb and Aggarwal.[5] Hoffman and Flinchbaugh[6] proved that, if there is at least one planar interpretation compatible with two frames, with each frame containing three points, then there are, in general, only two planar interpretations. Koenderink and van Doorn[7] and Bennett and Hoffman,[8] among others, obtained nonrigid interpretations in their analyses.

In this paper we investigate the two-frame case: we describe formally what rigid 3-D interpretations a vision system, whether biological or machine, can assign to a pair of 2-D frames when each frame contains four or more points.

When a vision system assigns a nonplanar 3-D interpretation to a 2-D display it is, of course, misinterpreting; the only correct interpretations of a 2-D display are themselves 2-D. However, when a vision system assigns a nonplanar 3-D interpretation to a 2-D image it is not, *ipso facto*, misinterpreting. But it is easy to show in this case that infinitely many 3-D interpretations are all equally compatible with any 2-D image. Thus, if a vision system picks one or more interpretations from this infinite collection, the system thereby reveals that its procedure for choosing 3-D interpretations is biased. If one gives a 3-D interpretation, one

0740-3232/89/071052-18$02.00

cannot avoid a bias. This paper can be viewed as an investigation of one bias, viz., rigidity. (The rigidity bias has been discussed extensively in the structure-from-motion literature.[9–13])

Since our analysis in this paper requires only two views, it can also be regarded as relevant to stereo vision.[14–19] In the case of stereo vision one has two distinct views of the same object, with the views taken at the same time; in the case of structure from motion one has two distinct views of the same object, with the views taken at different times. Since only the relative placements of the image plane and the object are relevant to our analysis, the analysis applies both to motion and to stereo vision. For simplicity of exposition, however, we use the structure-from-motion terminology.

From the standpoint of structure from motion, the assignment of rigid interpretations in the two-view case is particularly interesting not only because two views are the minimum number necessary to have motion at all but also because two views are the discrete-time analog of infinitesimal motion. Thus a complete analysis of the assignment of rigid interpretations in this case might provide a basis for a theory of the long-term (or stable) assignment of rigid and semirigid 3-D interpretations, for one might expect that the long-term assignment of rigid or quasi-rigid interpretations would arise from the integration of rigid interpretations in the infinitesimal case.

In Section 5 we exhibit in detail a natural group action on the collection of (almost) all rigid interpretations of two frames. (The group acts by transforming one rigid interpretation into another. If two interpretations are given, there is always a unique group element that effects the transformation.) The group action can be used to compute, *inter alia*, the dimension of the set consisting of (almost) all rigid interpretations (the so-called distinguished interpretations or distinguished configurations). This result, in turn, allows one to compute the dimension of the set of all image data compatible with a rigid interpretation (the so-called distinguished premises). This analysis yields one proof that the set of false targets has a measure of zero; we give a different proof of this fact in Section 2.

It will be helpful to set out some basic notation and terminology before stating our results.

Let $P_0, P_1, \ldots, P_n$ ($n \geq 3$) be points in $\mathbf{R}^3$, and let $Q_0, Q_1, \ldots, Q_n$ be the points obtained from the P_i's by means of a rigid motion of $\mathbf{R}^3$. If we assume that the viewer is using a moving coordinate system in which $P_0 = Q_0 = (0, 0, 0)$, then these rigid-motion data (viz., the P_i's and the Q_i's) are equivalent to the $2n$ vectors $(P_1 - P_0, P_2 - P_0, \ldots, P_n - P_0; Q_1 - Q_0, Q_2 - Q_0, \ldots, Q_n - Q_0)$, in which the last n vectors are obtained from the first n by a rotation about an axis containing the origin in $\mathbf{R}^3$. We call this an instantaneous rotation, since two successive positions of an object represent, as we mentioned above, the discrete-time analog of an instantaneous motion of that object. Thus to infer a (nontranslational) rigid motion of $n + 1$ points from two views is the same thing as to infer an instantaneous rotation of n vectors from two views; for convenience we henceforth formulate everything in the terminology of instantaneous rotation. We are therefore interested in configurations consisting of $2n$ vectors $(\mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1}; \mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2})$, where each $\mathbf{a}_{i,j}$ is in $\mathbf{R}^3$ and where $n \geq 3$. The set of all such configurations may be identified with $(\mathbf{R}^3)^{2n}$; we denote it occasionally by X. If a configuration is an instantaneous rotation, i.e., if the last n vectors are obtained from the first n by a rotation about an axis through the origin, we call it a distinguished configuration; the distinguished configurations form a subset $E \subset (\mathbf{R}^3)^{2n}$. E is a formal representation of the rigidity bias that appears to be used (perhaps in conjunction with other biases) in the interpretation of structure from motion.

We assume that there is a fixed (x, y, z) coordinate system in $\mathbf{R}^3$, in which the viewing direction is the z axis (positive or negative) and the image plane is the x–y plane, so that the viewer, when presented with a configuration, has access to the (x, y) coordinates of the $2n$ vectors. These data, which consist of $2n$ vectors in the x–y plane, are called the image data, or the premise associated to the given configuration. The set of all premises may be identified with $(\mathbf{R}^2)^{2n}$; we denote it by Y. We will typically write a premise in the form $\mathbf{b} = (\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{n,1}; \mathbf{b}_{1,2}, \ldots, \mathbf{b}_{n,2})$, where $\mathbf{b}_{i,j} = (x_{i,j}, y_{i,j})$ are the 2-D coordinates of the ith vector in frame or view j. We will denote by π the map that assigns a premise to each configuration by projecting the configuration onto the (x, y) plane; π simply deletes the z coordinate from each vector of the configuration. The premise associated to a distinguished configuration by π is called a distinguished premise; the distinguished premises form a subset $S \subset (\mathbf{R}^2)^{2n}$. Thus $\pi: (\mathbf{R}^3)^{2n} \to (\mathbf{R}^2)^{2n}$, and $\pi(E) = S$. This structure is shown in Fig. 1.

When presented with a premise, i.e., a vector $\mathbf{b} \in (\mathbf{R}^2)^{2n}$, the most that a viewer with a bias toward rigid interpretations can deduce is, first, whether $\mathbf{b}$ is in S and, if it is, the set $\pi^{-1}(\mathbf{b})$, i.e., the set of all configurations that, when their z coordinates are removed, are equal to $\mathbf{b}$.

However, a viewer with a rigidity bias selects interpretations from the set $\pi^{-1}(\mathbf{b}) \cap E$, supposing such an interpretation to represent the configuration that was actually responsible for the premise $\mathbf{b}$. However, as we mentioned above, the choice of such an interpretation cannot be deductively valid, given the premise $\mathbf{b}$, since there are infinitely many other interpretations that, when their z coordinates are removed, are equal to $\mathbf{b}$. To put this another way, there is no deductively valid way to rule out the possibility of a false target when a distinguished premise $\mathbf{b}$ is presented; a false target is represented by an interpretation in $\pi^{-1}(\mathbf{b})$ that is not in E. (For example, a collection of points on a video screen that simulates the retinal image of a rigid object in $\mathbf{R}^3$ can be a false target for a viewer with a rigidity bias.) Thus, for an observer to infer a unique distinguished configuration in response to a distinguished premise $\mathbf{b}$, an inductive (i.e., not deductively valid) inference is needed. The underlying information structure on which the inductive inferences are superimposed consists of the relevant properties of the map π vis-à-vis the distinguished configurations E and the distinguished premises S. The object of this study is to elucidate these properties.

$$
\begin{array}{lcccl}
\text{configurations} & X = \mathbf{R}^{18} & \supset & E & \text{distinguished configurations} \\
 & \big\downarrow \Pi & & \big\downarrow \Pi & \\
\text{premises} & Y = \mathbf{R}^{12} & \supset & S & \text{distinguished premises}
\end{array}
$$

Fig. 1. Premise and configuration spaces for the inference of instantaneous rotations. The set S of distinguished premises corresponds to those image data that are compatible with an instantaneous-rotation interpretation.

We note in passing that the entire structure consisting of X, Y, E, S, and π together with statistical information about the probabilities assigned to distinct interpretations in the case of multistable perceptions constitutes an example of a formal structure called an observer. We describe this structure in Section 4.

We now summarize our main results, using the terminology of π, E, and S introduced above. These results hold for any $n \geq 3$.

1. In proposition 1 of Section 2, we show that S is contained in the locus of points in $(\mathbf{R}^2)^{2n}$ whose coordinates satisfy certain polynomial equations $f_{p,q,r} = 0$, where $1 \leq p < q < r \leq n$ (so that there are $\binom{n}{3}$ such equations). As a corollary of this result, we find that S has a (Lebesgue) measure of 0 in $(\mathbf{R}^2)^{2n}$; i.e., almost all premises are incompatible with a rigid interpretation.
2. We introduce an additional condition, condition 1 of Section 2, which allows one to determine whether a point that satisfies the polynomial equations is actually in S. Thus the polynomial conditions $f_{p,q,r} = 0$ together with condition 1 give necessary and sufficient conditions for S; in fact, they provide an algorithm to determine whether a premise is in S (i.e., if the two frames of four or more points are compatible with any rigid interpretations).
3. The same computation used to verify condition 1 for a point s of S leads immediately to an algorithm to compute the set $\pi^{-1}(s) \cap E$, viz., the set of all rigid interpretations compatible with s. From this result we find that
4. $\pi^{-1}(s) \cap E$ is a one-parameter family of distinguished configurations.[20]

These results constitute theorem 1 of Section 2. In Section 3 we study the geometry of the one-parameter family of rigid interpretations associated with any given point s of S. To describe the results, we note first that to each distinguished configuration is associated a unique line through the origin in $\mathbf{R}^3$, namely, the axis of the rotation that transforms the P_i's to the Q_i's. We find that

5. As the distinguished configurations vary in $\pi^{-1}(s) \cap E$, their associated axes of rotation sweep out a plane perpendicular to the (x, y) plane; in particular, all these axes project to the same line (call it M) in the image plane.

Thus a natural parameter for the family of rigid interpretations that are compatible with a particular $s \in S$ is the (slant) angle σ that the axis of rotation makes with, say, the positive z axis. The structures in the family then correspond uniquely to values of σ in the union of open intervals $(0, \pi/2) \cup (\pi/2, \pi)$; the structure corresponding to $\sigma \in (0, \pi/2)$ is the reflection in the image plane of the structure corresponding to $\pi - \sigma$. We note that the values 0, $\pi/2$, and π are omitted: there are no actual rigid structures corresponding to these values of σ; instead they correspond to degenerate limits of rigid structures (as we describe below).

The situation may be presented geometrically as follows. A point $\mathbf{b} \in (\mathbf{R}^2)^{2n}$ corresponds to the $2n$ vectors $(\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{n,1}; \mathbf{b}_{1,2}, \ldots, \mathbf{b}_{n,2})$ in $\mathbf{R}^2$. $\mathbf{b}$ is in S if and only if it is possible to find n ellipses $D_1, \ldots, D_n$ with these properties (see Fig. 2):

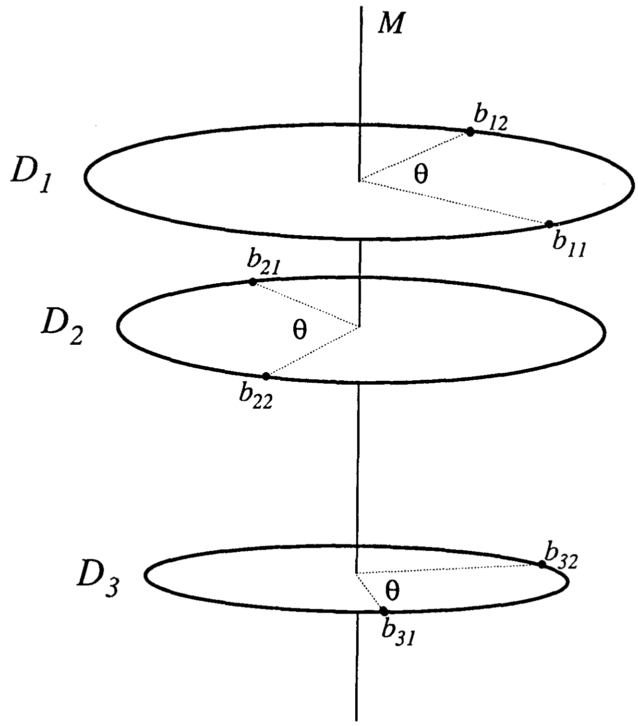

Fig. 2. Ellipses D_i, whose existence as shown means that the premise consisting of the image data $\mathbf{b}_{i,j}$ is distinguished.

a. There exists a line M through the origin of $\mathbf{R}^2$ so that the minor axes of the ellipses D_i all lie on M.

b. All the ellipses have the same eccentricity.

c. For each $i = 1, \ldots, n$, the points $\mathbf{b}_{i,1}$ and $\mathbf{b}_{i,2}$ are on D_i.

d. The elliptical angle θ on D_i determined by the points $\mathbf{b}_{i,1}$ and $\mathbf{b}_{i,2}$ is the same for each $i = 1, \ldots, n$. (By the elliptical angle determined by two points on an ellipse we mean the following: choose any circle in $\mathbf{R}^3$ that projects orthographically to the given ellipse and take the angle subtended at the center of this circle by the two points on the circle that project to the given points on the ellipse.)

The existence of the ellipses D_i with these properties is equivalent to the existence of an axis L through the origin in $\mathbf{R}^3$ such that (1) L projects to M, (2) the slant of L is determined by the common eccentricity of the family of ellipses, and (3) the points of $\mathbf{b}$ are the projections of the successive positions of n points in $\mathbf{R}^3$ that rotate about L through some elliptical angle θ. This is illustrated in Fig. 3. The results already summarized may be restated geometrically:

6. Given $\mathbf{b} = \{\mathbf{b}_{i,j}\}$, there are $\binom{n}{3}$ fixed polynomials $f_{p,q,r}$ $(1 \leq p < q < r \leq n)$ in the coordinate of the $\mathbf{b}_{i,j}$. If and only if (1) all these $f_{p,q,r}$ vanish and (2) condition 1 of Section 2 is satisfied, then there exists a set of ellipses $D_1, \ldots, D_n$ that satisfy the conditions described above. Moreover, if there is one such set of ellipses, then, in fact, there is a one-parameter family of them. This family can be parameterized by the eccentricity of the ellipses or, equivalently, by the slant, σ, of the axis L. (Recall that the slant of L is the angle between L and the positive z axis.)

Suppose that $\mathbf{b} \in S$ is given. As the parameter σ increases over the interval $(0, \pi/2)$, the eccentricity of the ellipses increases. In fact, as σ approaches $\pi/2$, each ellipse D_i degenerates into a pair of straight lines orthogonal to M through the points $\mathbf{b}_{i,1}$ and $\mathbf{b}_{i,2}$. This limiting case clearly does not correspond to a distinguished configuration.

On the other extreme, as σ approaches 0, each ellipse D_i approaches a circle whose center is the unique point on M equidistant from $\mathbf{b}_{i,1}$ and $\mathbf{b}_{i,2}$. Similarly, this does not correspond to a distinguished configuration. If it did, then, since the ellipses are circles, the axis of rotation of the configuration must be perpendicular to the image plane (i.e., it coincides with the z axis so that $\sigma = 0$), in which case the centers of the circles would project to the origin.

The elliptical angle θ depends on σ: it is strictly increasing as σ goes from 0 to $\pi/2$. The radii of the circles C_i (which project onto the ellipses D_i) also vary with σ. For each $i = 1, \ldots, n$ separately there is a value of σ for which the radius of C_i is minimized, but these values of σ differ for the various i. Thus there is no value of σ for which these radii are minimized simultaneously.

These mathematical properties appear to imply that any given generic σ family contains no mathematically distinguished member. In other words, there appears to be no member of the family that, for purely mathematical reasons (such as minimizing some geometrically meaningful functional), should be a natural perceptual feature point of the family. This raises the question of exactly which, if any, of the possible distinguished configurations is perceived by a human subject who, in response to a distinguished premise, experiences the strong sensation of seeing a rigid object. Such a choice of distinguished configuration is, if one uses only a constraint of rigidity, mathematically gratuitous. This suggests that the perceptual system may be using constraints other than rigidity alone to arrive at its interpretations. It is therefore of particular interest that psychophysical experimentation be focused on these issues. (We are currently working with Braunstein on this question.)

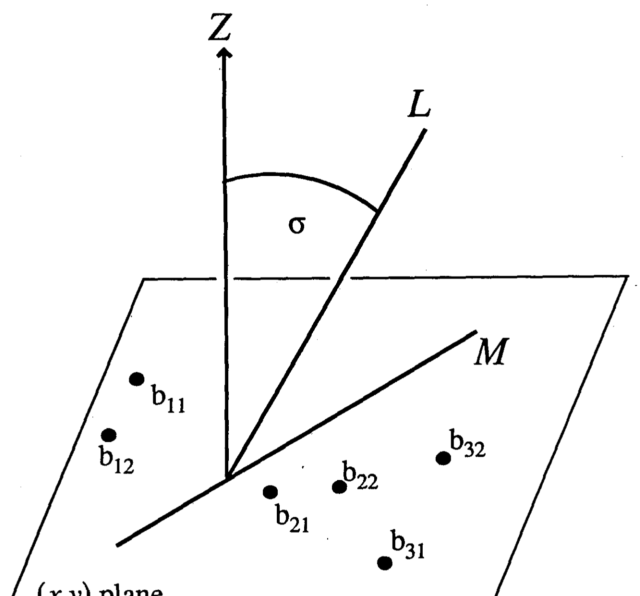

Fig. 3. One-parameter family of rigid structures in $\mathbf{R}^3$ that project to the given distinguished imaged data $\{\mathbf{b}_{i,j}\}$ in the (x, y) plane. These structures are parameterized by the slant σ of their axes, $L(\sigma)$, of rotation; all axes project to the same line M in the (x, y) plane.

In Section 5 we discuss some group-theoretic aspects of the instantaneous rotation observer. Let E' denote the set whose elements are pairs, each of which consists of a distinguished configuration (i.e., an element of E), together with a choice of orientation (i.e., of a positive direction) of the rotation axis of that configuration. Thus there are two elements of E' for each element of E. E' and E have the same dimension. We show that there is a group J of geometric transformations that acts naturally on E' (outside of a small subset E_1' of degenerate configurations of lower dimension than E' itself). In fact, J acts in such a way that anytwo elements of $E' - E_1'$ are related by a unique element of J; in mathematical terminology, $E_0' = E' - E_1'$ is a principal homogeneous space for J. Perceptually, this corresponds to saying that the group J is a particularly natural group of subjective transformations for instantaneous-rotation percepts.

The principal homogeneous space structure on E_0' permits us to compute its dimension and, consequently, the dimensions of E and S. In this way it turns out that the dimension of S is $3n + 2$; i.e., S has $n - 2$ dimensions less than the ambient space $Y = (\mathbf{R}^2)^{2n}$ (in mathematical terminology, S has a codimension of $n - 2$ in Y). For example, when $n = 3$, S is defined by the single polynomial condition $f_{1,2,3} = 0$ so that S is a hypersurface in Y and therefore has a codimension of 1 in Y. However, when $n > 3$ it is not at all obvious from the conditions defining S that S should have a codimension of $n - 2$. The perceptual significance of the fact that the codimension of S in Y increases with n is an open question. One sample hypothesis is that, the larger the codimension is, the rarer is the occurrence of a distinguished premise among all possible premises; hence a distinguished premise might be more striking when it did occur, so that the percept of rigidity might be stronger for large n. We merely mention this issue here for its intrinsic interest; we draw no conclusions about it in this paper.

2. COMPUTATION OF S AND OF $\pi^{-1}(s) \cap E$

We shall use the terminology and notation of Section 1:

$$\mathbf{a} = (\mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1}; \mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2})$$

denotes an element of $(\mathbf{R}^3)^{2n}$, i.e., a configuration consisting of $2n$ vectors $\mathbf{a}_{i,j}$ in $\mathbf{R}^3$. We consider $\mathbf{a}_{i,1}$, $\mathbf{a}_{i,2}$ as the position vectors at successive times of the ith point on some moving object, which may or may not be rigid. We assume a fixed (x, y, z) coordinate system in $\mathbf{R}^3$, and

$$\mathbf{b} = (\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{n,1}; \mathbf{b}_{1,2}, \ldots, \mathbf{b}_{n,2})$$

denotes an element of $(\mathbf{R}^2)^{2n}$. We consider $\mathbf{R}^2$ as the x–y plane in $\mathbf{R}^3$, which is the image plane of a viewer who acquires images by orthographic projection along the z direction. There is a natural map

$$\pi:(\mathbf{R}^3)^{2n} \rightarrow (\mathbf{R}^2)^{2n}$$

so that, if $\mathbf{a}$ and $\mathbf{b}$ are as described above with $\pi(\mathbf{a}) = \mathbf{b}$, then $\mathbf{b}_{i,j}$ is the orthogonal projection of $\mathbf{a}_{i,j}$ onto the x–y plane. In this case, let $(\hat{e}_1, \hat{e}_2, \hat{e}_3)$ be orthogonal unit vectors in the x, y, and z directions, respectively; then we can write

$$\mathbf{a}_{i,j} = x_{i,j}\hat{e}_1 + y_{i,j}\hat{e}_2 + z_{i,j}\hat{e}_3, \tag{1}$$

where $i = 1, \ldots, n$ and $j = 1, 2$. Moreover,

$$\mathbf{b}_{i,j} = x_{i,j}\hat{e}_1 + y_{i,j}\hat{e}_2, \tag{2}$$

with $i = 1, \ldots, n$ and $j = 1, 2$, and we can rewrite Eq. (1) as

$$\mathbf{a}_{i,j} = \mathbf{b}_{i,j} + z_{i,j}\hat{e}_3, \tag{3}$$

where $i = 1, \ldots, n$ and $j = 1, 2$.

We express the assumption of rigid motion by two sets of equations. The first set consists of the equations

$$\mathbf{a}_{l,1} \cdot \mathbf{a}_{m,1} = \mathbf{a}_{l,2} \cdot \mathbf{a}_{m,2}, \qquad 1 \le l, m \le n, \tag{4}$$

where $\cdot$ indicates the dot (scalar) product of vectors. $n + \binom{n}{2}$ of these equations are distinct (take, e.g., $l \le m$). Those n equations described in Eq. (4) for which $l = m$ express the assumption that the distance of each point from the origin does not change from the first view to the second. The $\binom{n}{2}$ equations for which $l \neq m$ then express the condition that the angle between the lth and mth position vectors does not change from the first view to the second. Together, the equations described by Eq. (4) imply that there exists an orthogonal 3×3 matrix that relates the vectors $(\mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1})$ to the vectors $(\mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2})$. Now, in general, an orthogonal matrix represents a motion that involves both a rotation and a reflection; thus we need some additional equations to express that this orthogonal matrix is actually a rotation. For this, it is enough to require that the matrix preserve the scalar triple product of, say, the first three vectors $\mathbf{a}_{1,1}$, $\mathbf{a}_{2,1}$, and $\mathbf{a}_{3,1}$. (If an orthogonal matrix involves a reflection, it will reverse the sign of the scalar triple product of any three linearly independent vectors). Thus we need only include the additional equation

$$(\mathbf{a}_{1,1} \times \mathbf{a}_{2,1}) \cdot \mathbf{a}_{3,1} = (\mathbf{a}_{1,2} \times \mathbf{a}_{2,2}) \cdot \mathbf{a}_{3,2}, \tag{5}$$

which is the same as

$$\det\begin{bmatrix} x_{1,1} & y_{1,1} & z_{1,1} \\ x_{2,1} & y_{2,1} & z_{2,1} \\ x_{3,1} & y_{3,1} & z_{3,1} \end{bmatrix} = \det\begin{bmatrix} x_{1,2} & y_{1,2} & z_{1,2} \\ x_{2,2} & y_{2,2} & z_{2,2} \\ x_{3,2} & y_{3,2} & z_{3,2} \end{bmatrix}.$$

Thus Eq. (5) can be written in the form

$$\psi = 0,$$

where

$$\begin{aligned}\psi = {}& (x_{2,1}y_{3,1} - x_{3,1}y_{2,1})z_{1,1} - (x_{1,1}y_{3,1} - x_{3,1}y_{1,1})z_{2,1} \\ & + (x_{1,1}y_{2,1} - x_{2,1}y_{1,1})z_{3,1} - (x_{2,2}y_{3,2} - x_{3,2}y_{2,2})z_{1,2} \\ & + (x_{1,2}y_{3,2} - x_{3,2}y_{1,2})z_{2,2} - (x_{1,2}y_{2,2} - x_{2,2}y_{1,2})z_{3,2}.\end{aligned} \tag{6}$$

Now, with Eq. (3), Eq. (4) becomes

$$(\mathbf{b}_{l,1} + z_{l,1}\hat{e}_3) \cdot (\mathbf{b}_{m,1} + z_{m,1}\hat{e}_3) = (\mathbf{b}_{l,2} + z_{l,2}\hat{e}_3) \cdot (\mathbf{b}_{m,2} + z_{m,2}\hat{e}_3), \qquad 1 \le l \le m \le n. \tag{7}$$

Expanding and simplifying, we obtain

$$z_{l,1}z_{m,1} = d_{l,m} + z_{l,2}z_{m,2}, \qquad 1 \le l \le m \le n. \tag{8}$$

Equivalently, we may write Eq. (8) in the form

$$\phi_{l,m} = 0, \tag{9}$$

where

$$\phi_{l,m} = z_{l,2}z_{m,2} - z_{l,1}z_{m,1} + d_{l,m}$$

$$d_{l,m} = \mathbf{b}_{l,2} \cdot \mathbf{b}_{m,2} - \mathbf{b}_{l,1} \cdot \mathbf{b}_{m,1}, \qquad 1 \le l \le m \le n.$$

Thus the rigidity assumption on the configuration $\mathbf{a}$ is equivalent to the system of equations

$$\psi = 0, \qquad \phi_{l,m} = 0, \qquad 1 \le l, m \le n \tag{10}$$

in the $x_{i,j}$, $y_{i,y}$, and $z_{i,j}$, for which ψ and the $\phi_{l,m}$ are as in Eqs. (6) and (8). These equations define the distinguished configuration locus E in $(\mathbf{R}^3)^{2n}$. We can also view Eqs. (10) as equations in the unknowns $z_{i,j}$ whose coefficients depend on the premise $\mathbf{b}$. The viewer, given a premise $\mathbf{b}$, would ideally solve the resulting system for the $z_{i,j}$, thereby computing a rigid structure compatible with the premise. We now present the mathematical results that describe to what extent such a computation is possible.

Let $1 \le p < q < r \le n$ be any three distinct indices. We set

$$f_{p,q,r} = \det\begin{bmatrix} d_{p,p} & d_{p,q} & d_{p,r} \\ d_{q,p} & d_{q,q} & d_{q,r} \\ d_{r,p} & d_{r,q} & d_{r,r} \end{bmatrix}, \tag{11}$$

where the $d_{l,m}$ are as in Eq. (9). (Note that since $d_{l,m} = d_{m,l}$ the above matrix is symmetric.) The $f_{p,q,r}$ may be viewed as polynomials in the $x_{i,j}$ and $y_{i,j}$ and hence as functions of $\mathbf{b} \in (\mathbf{R}^2)^{2n}$. We will first prove the following proposition.

Proposition 1

A necessary condition for the system of Eq. (9): $\phi_{l,m} = 0$ arising from a given premise $\mathbf{b}$ to have a solution in the $z_{i,j}$ is that all the $f_{p,q,r}$, as in Eq. (11), vanish on the coordinates of the $\mathbf{b}_{i,j}$ terms.

In other words, the distinguished premise set S, which is the image by π of the distinguished configuration set $E \subset (\mathbf{R}^3)^{2n}$, is contained in the locus in $(\mathbf{R}^2)^{2n}$ defined by the equations $f_{p,q,r} = 0$. We show below that the criterion of proposition 1 may be augmented to yield a necessary and sufficient condition that is computationally effective. In other words, beginning with proposition 1 we derive an effective computational procedure for determining whether any given premise, i.e., any given collection of image data, is compatible with a rigid 3-D interpretation. Note that such a necessary and sufficient condition that the image data be compatible with a rigid 3-D interpretation is only a necessary (but not sufficient) condition that the image data actually came from a rigid object in $\mathbf{R}^3$. This is because, for $s \in S = \pi(E)$, $\pi^{-1}(s)$ contains configurations not in E as well as configurations in E. In this sense, when the observer infers that a distinguished premise arose from a rigid object the observer is expressing a bias toward rigid interpretations.

Proof of Proposition 1

We want to solve Eq. (9) for the $z_{i,j}$, given values for the $x_{i,j}$ and $y_{i,j}$. Thus we assume that values for the $d_{l,m}$ are given, and we will try to solve the corresponding equations $\phi_{l,m} = 0$ for the z.

To simplify the notation, we shall use w_l in place of $z_{l,1}$ and z_l in place of $z_{l,2}$. With this notation, we want to solve the system

$$\phi_{l,l} = z_l^2 - w_l^2 + d_{l,l} = 0,$$

$$\phi_{l,m} = z_l z_m - w_l w_m + d_{l,m} = 0, \qquad l < m \tag{12}$$

for the z's and w's, where the $d_{l,m}$ are now fixed numbers.

It will be helpful to introduce new coordinates

$$u_l = z_l + w_l, \qquad v_l = z_l - w_l. \tag{13}$$

In terms of these coordinates, Eqs. (12) become

$$\phi_{l,l} = u_l v_l + d_{l,l} = 0,$$

$$\phi_{l,m} = u_l v_m + u_m v_l + 2d_{l,m} = 0, \qquad l < m. \tag{14}$$

Let us assume temporarily that

$$d_{l,l} \neq 0, \qquad l = 1, \ldots, n. \tag{15}$$

It follows from $\phi_{l,l} = 0$ in Eqs. (14) that $v_l \neq 0$. Thus we can divide the equation $\phi_{l,l} = 0$ by v_l to obtain

$$u_l = -d_{l,l}/v_l, \qquad l = 1, \ldots, n. \tag{16}$$

Substituting this into $\phi_{l,m} = 0$ ($l < m$) yields

$$-d_{l,l}v_m/v_l - d_{m,m}v_l/v_m + 2d_{l,m} = 0,$$

which (after multiplication by the nonzero quantity $-v_l v_m$) is equivalent to

$$d_{l,l}v_m^2 + d_{m,m}v_l^2 - 2d_{l,m}v_l v_m = 0. \tag{17}$$

Any solution to Eq. (17) for the v's for which no v_i is 0, together with the corresponding values of the u's from Eq. (16), gives a solution to Eq. (14). We seek necessary and sufficient conditions for such a solution to Eq. (17) to exist.

For this purpose we first note that the left-hand side of Eq. (17) factors:

$$d_{m,m}v_l^2 + d_{l,l}v_m^2 - 2d_{l,m}v_l v_m = d_{m,m}(v_l - \alpha_{l,m}^{+}v_m)(v_l - \alpha_{l,m}^{-}v_m),$$

where

$$\alpha_{l,m}^{+} = [d_{l,m} + (d_{l,m}^2 - d_{l,l}d_{m,m})^{1/2}]/d_{m,m},$$

$$\alpha_{l,m}^{-} = [d_{l,m} - (d_{l,m}^2 - d_{l,l}d_{m,m})^{1/2}]/d_{m,m}. \tag{18}$$

Thus, under our assumption that $d_{l,l} \neq 0$ for all l, Eq. (17) is equivalent to

$$(v_l - \alpha_{l,m}^{+}v_m)(v_l - \alpha_{l,m}^{-}v_m) = 0, \qquad l < m. \tag{19}$$

To say that $(v_1, \ldots, v_n)$ is a solution of this system means that one of the linear factors in each of the $\binom{n}{2}$ forms of Eq. (19) vanishes. In other words, the desired solutions are solutions to a system

$$v_l - \alpha_{l,m}v_m = 0, \qquad l < m, \tag{20}$$

in which no v_i is 0 and in which $\alpha_{l,m} = \alpha_{l,m}^{+}$ or $\alpha_{l,m}^{-}$ [so that there are $2\binom{n}{2}$ such systems in general].

Now, suppose that we are given a solution $v_1, \ldots, v_n$ to Eq. (20) for some choice of the $\alpha_{l,m}$'s. For any three distinct indices, say, $p < q < r \leq n$, we can consider the subsystem of Eq. (20) consisting of the three equations

$$v_p - \alpha_{p,q}v_q = 0,$$

$$v_q - \alpha_{q,r}v_r = 0,$$

$$v_p - \alpha_{p,r}v_r = 0. \tag{21}$$

If these equations are linearly independent, they imply that

$$v_p = 0, \qquad v_q = 0, \qquad v_r = 0;$$

but then these v's cannot be a solution to our original system, for we seek solutions for which no v_i is 0. Thus we find that, under assumption (15), a necessary condition for the original system [Eq. (14)] to have a solution is that, for any $p < q < r$ as above, there exists a choice of $\alpha_{p,q}$, $\alpha_{q,r}$, and $\alpha_{p,r}$ such that Eqs. (21) are linearly dependent.

It is now evident that the linear dependence of Eqs. (21) is equivalent to

$$\alpha_{p,q}\alpha_{q,r} - \alpha_{p,r} = 0. \tag{22}$$

Lemma 1

For given $p < q < r$, the product of all eight left-hand sides of Eq. (22) (obtained by taking all possible choices $\alpha_{p,q} = \alpha_{p,q}^{+}$ or $\alpha_{p,q} = \alpha_{p,q}^{-}$, $\alpha_{q,r} = \alpha_{q,r}^{+}$ or $\alpha_{q,r} = \alpha_{q,r}^{-}$, and $\alpha_{p,r} = \alpha_{p,r}^{+}$ or $\alpha_{p,r} = \alpha_{p,r}^{-}$) is

$$(16d_{p,p}^2/d_{q,q}^2 d_{r,r}^6)f_{p,q,r}.$$

Proof

Lemma 1 may be proved by direct computation. (We first obtained this result by using MACSYMA, but it is possible, if a bit tedious, to do it by hand.) ∎

Therefore, since we are assuming that the $d_{l,l}$ are nonzero, saying that $f_{p,q,r}$ vanishes is equivalent to saying that, for some choice of values $\alpha_{p,q}^{+}$ or $\alpha_{p,q}^{-}$ for $\alpha_{p,q}$, Eqs. (21) are linearly dependent. Hence the vanishing of $f_{p,q,r}$ is necessary for Eq. (17) to have a nonzero solution; so we have proved proposition 1, under the assumption that no $d_{l,l} = 0$.

We now eliminate this assumption. Suppose that some $d_{p,p} = 0$ for some p, and that the Eqs. (14) have a solution $u_1, \ldots, u_n; v_1, \ldots, v_n$. We then have the equation

$$\phi_{p,p} = u_p v_p - d_{p,p} = 0,$$

i.e.,

$$u_p v_p = 0,$$

so that either u_p or v_p must be 0, and the other may be arbitrary. Suppose, then, that $u_p = 0$ and v_p is arbitrary. (Since the equations are symmetric in u and v, the case $v_p = 0$ may be treated analogously.) For any q, the equations

$$\phi_{p,q} = u_p v_q + u_q v_p + 2d_{p,q} = 0, \qquad 1 \leq q \leq n,$$

reduce to

$$u_q v_p + 2d_{p,q} = 0.$$

Now, if v_p is also zero, we see that $d_{p,q} = 0$ for all q. However, in that case, the first row in the matrix in Eq. (11) is a zero row, so that $f_{p,q,r}$ is necessarily zero. If, on the other hand, $v_p \neq 0$, we conclude that, for any q,

$$u_q = -2d_{p,q}/v_p. \tag{23}$$

Now, for any q and r, we can use Eq. (23) to replace u_q and u_r in

$$\phi_{q,r} = u_q v_r + u_r v_q + 2d_{q,r} = 0,$$

and we obtain

$$(-2d_{p,q}/v_p)v_r + (-2d_{p,r}/v_p)v_q + 2d_{q,r} = 0. \quad (24)$$

Now, from the equations described by

$$\phi_{q,q} = u_q v_q + d_{q,q} = 0,$$

we can replace v_q by $-d_{q,q}/u_q$. By using Eq. (23) again, we get

$$v_q = -d_{q,q}/(-2d_{p,q}/v_p) = d_{q,q}v_p/2d_{p,q}. \quad (25)$$

After using Eq. (25) to replace v_r and v_q in Eq. (24), we obtain

$$-d_{p,q}d_{r,r}/d_{p,r} - d_{p,r}d_{q,q}/d_{p,q} + 2d_{q,r} = 0.$$

Multiplying by $d_{p,q}d_{p,r}$, we obtain

$$-d_{p,q}^2 d_{r,r} - d_{p,r}^2 d_{q,q} + 2d_{p,q}d_{p,r}d_{q,r} = 0. \quad (26)$$

Now, recall [from Eq. (11)] that

$$f_{p,q,r} = \det\begin{bmatrix} d_{p,p} & d_{p,q} & d_{p,r} \\ d_{q,p} & d_{q,q} & d_{q,r} \\ d_{r,p} & d_{r,q} & d_{r,r} \end{bmatrix}.$$

Given our assumption that $d_{p,p} = 0$, we can check that $f_{p,q,r}$ is the same as the left-hand side of Eq. (26), i.e., $f_{p,q,r} = 0$.

On the other hand, if p, q, and r is three indices for which none of $d_{p,p}$, $d_{q,q}$, and $d_{r,r}$ is 0, then the original proof (applied to the three-vector subsystem consisting of these indices alone) gives the desired result that $f_{p,q,r} = 0$. Thus in any case all the $f_{p,q,r}$ are 0. This completes the proof of proposition 1. ∎

Beginning with this result, we want to develop a sufficient condition for Eq. (17) to have a nonzero solution. To this end, suppose that all the $f_{p,q,r}$ vanish; then, for each choice of $p < q < r$, there is a choice for the $\alpha_{p,q}$, $\alpha_{q,r}$, and $\alpha_{p,r}$ so that the system of Eqs. (21) is linearly dependent and consequently has a nonzero solution. However, if $n > 3$, there is more than one choice of such p, q, and r. In particular, for a given p and q, there is more than one value of r for which $\alpha_{p,q}$ satisfies Eq. (22). For example, in the case in which $n = 4$, assuming that all the equations $f_{p,q,r} = 0$ are satisfied, we will have both

$$\alpha_{1,2}\alpha_{2,3} = \alpha_{1,3}$$

for some choice of values of $\alpha_{1,2}$, $\alpha_{2,3}$, and $\alpha_{1,3}$ and

$$\alpha_{1,2}\alpha_{2,4} = \alpha_{1,4}$$

for some choice of values of $\alpha_{1,2}$, $\alpha_{2,4}$, and $\alpha_{1,4}$. However, the choice of value of $\alpha_{1,2}$ for which the first equation holds may not be the same as the choice of $\alpha_{1,2}$ for which the second equation holds. In other words, when $n \geq 4$, the fact that $f_{p,q,r} = 0$ for all $p < q < r$ does not imply that there is a single choice of value for each $\alpha_{p,q}$ with which Eq. (22) holds for all r. It might occur, for example that

$$\alpha_{1,2}^{+}\alpha_{2,3}^{+} = \alpha_{1,3}^{-},$$

whereas

$$\alpha_{1,2}^{-}\alpha_{2,4}^{+} = \alpha_{1,4}^{-};$$

in this case there is no single choice for $\alpha_{1,2}$ so that, with this choice, Eqs. (21) are linearly dependent for both $(p, q, r) = (1, 2, 3)$ and $(p, q, r) = (1, 2, 4)$. We cannot expect, then, that there is any choice of the $\alpha_{p,q}$, depending only on p and q and not on some particular (p, q, r) triple, so that for these fixed choices all the six equations of the form of Eq. (20) for $1 \leq l < m \leq 4$ have a common nonzero solution in v_1, v_2, v_3, and v_4.

Thus, assuming that all the $f_{p,q,r}$ vanish, we consider the following additional condition.

Condition 1

For the given set $\{\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{n,2}\}$ of image vectors, and for each pair of indices $1 \leq l, m \leq n$, there is a single fixed choice for $\alpha_{l,m}$ (namely, $\alpha_{l,m}^{+}$ or $\alpha_{l,m}^{-}$) so that for each $p < q < r$ the Eqs. (22) hold for the given fixed choices of $\alpha_{p,q}$, $\alpha_{q,r}$, and $\alpha_{p,r}$.

If condition 1 holds, consider the corresponding system of equations of the type of Eq. (20) (in which the choices of the coefficients $\alpha_{l,m}$ are the given fixed ones); then any of the $\binom{n}{2}$ equations in this system [Eq. (20)] is a consequence of just the equations

$$\begin{aligned} v_1 - \alpha_{1,2}v_2 &= 0, \\ v_1 - \alpha_{1,3}v_3 &= 0, \\ &\vdots \\ v_1 - \alpha_{1,n}v_n &= 0. \end{aligned} \quad (27)$$

This is a system of only $n - 1$ (independent) linear equations in n unknowns, so it has a one-dimensional (1-D) set of solutions; in particular, it has a nonzero solution $v_1, \ldots, v_n$. However, we note that, because of the particular form of Eq. (20), if one of the v_i is 0, so are all the others, so that in a nonzero solution none of the v_i equals 0. Thus we obtain a solution to Eqs. (14), in fact, a 1-D family of such solutions.

Now we note that Eqs. (14) remain unchanged if we exchange the u and v coordinates, i.e., if we exchange u_l and u_l for each $l = 1, \ldots, n$. Therefore the solution set is invariant under such an interchange. Thus, corresponding to the 1-D family of solutions above, there is another such family that is obtained by exchanging the u and v coordinates in the first family. Thus we make the following proposition.

Proposition 2

Necessary and sufficient conditions for the existence of a solution to Eqs. (14) are that all the $f_{p,q,r}$ vanish and, moreover, that condition 1 be satisfied. In this case there are two 1-D families of such solutions, which are related by exchanging the u and v coordinates.

We remark that it is a relatively inexpensive computation to check condition 1, given that the $f_{p,q,r}$ terms vanish. In fact, for each p, q, and r it can be determined explicitly which choice of $\alpha_{p,q}$ makes Eq. (22) work, and then it becomes evident whether condition 1 holds. Even for large n, one can envisage a parallel processor that performs the necessary verifications easily.

We believe that it is possible to express both of the conditions of proposition 2 as a single set of polynomial conditions similar to (but more complex than) the vanishing of the $f_{p,q,r}$; however, we have not yet found such conditions. In any case, for computational purposes the conditions of proposition 2 are effective, for to solve the system of Eq. (20) explicitly (i.e., to produce explicitly the one-parameter family of solutions) one must use actual values of the $\alpha_{p,q}$, so the verification of condition 1 is already necessary for the computation.

Theorem 1

Necessary and sufficient conditions for the existence of a rigid interpretation for the image data **b** are that the $f_{p,q,r}$ vanish and that condition 1 be satisfied. In this case there is (generically) a unique 1-D family of interpretations.

Proof

A rigid interpretation for the image data is a solution to Eqs. (10), consisting of the equations $\phi_{l,m} = 0$ and the equation $\psi = 0$. According to proposition 2, the conditions in question are necessary and sufficient for the existence of solutions to the equations $\phi_{l,m} = 0$ alone, and there are then two one-parameter families of such solutions. The proof of theorem 1 is accomplished by showing that, of these two one-parameter families, the solutions in exactly one of them always satisfy the additional equation $\psi = 0$.

To see this, recall that $\psi = 0$ means

$$\det\begin{bmatrix} x_{1,1} & y_{1,1} & z_{1,1} \\ x_{2,1} & y_{2,1} & z_{2,1} \\ x_{3,1} & y_{3,1} & z_{3,1} \end{bmatrix} = \det\begin{bmatrix} x_{1,2} & y_{1,2} & z_{1,2} \\ x_{2,2} & y_{2,2} & z_{2,2} \\ x_{3,2} & y_{3,2} & z_{3,2} \end{bmatrix}.$$

On the other hand, when we introduced the equation $\psi = 0$, we noted that the equations $\phi_{l,m} = 0$ [for $(l, m) = (1, 2), (1, 3), (2, 3)$] already imply that these determinants differ at most by a sign. Therefore, if the determinants are not equal, they become equal if each $z_{l,1}$ is replaced by $-z_{l,1}$ (for $l = 1, \ldots, n$) and each $z_{l,2}$ is left as is. Now, this new set of z's is still a solution to the equations $\phi_{l,m} = 0$ [cf. the original form, Eq. (9), of these equations]. Moreover, we check that this transformation of the z's, when expressed in terms of the (u, v) coordinates, corresponds to the exchanging of the u and v. Therefore, from proposition 2, the transformation corresponds to moving from one of the two one-parameter families of solutions to the $\phi_{l,m} = 0$ to the other. Now the values of the determinants in the expression above are clearly continuous functions of the entries and hence remain the same as we move within any one of the one-parameter families. It follows that one of these families consists of solutions that satisfy the determinant equation and the other does not. Thus we conclude the proof. ∎

3. GEOMETRY OF $\pi^{-1}(s) \cap E$

We begin with a point s of S, and we consider the one-parameter family that contains all those rigid structures compatible with the image data **b** represented by s. Algebraically, the members of this family correspond to solutions of the linear Eqs. (27) for $(v_1, \ldots, v_n)$: namely, for every such solution we define $(u_1, \ldots, u_n)$ by $u_l = -d_{l,l}/v_l$, and then (considering here, for simplicity, only the case in which $d_{l,l} \neq 0$, for all l) we get the unknowns w_1 and z_1 (which are simplified notations for the original depth coordinates $z_{l,1}$ and $z_{l,2}$) from

$$z_l = (u_l + v_l)/2, \qquad w_l = (u_l - v_l)/2. \tag{28}$$

Now since Eqs. (27) are homogeneous and linear, if $(v_1, \ldots, v_n)$ is one solution and if t is any nonzero number, then $(tv_1, \ldots, tv_n)$ is another solution, and in fact all solutions are obtained in this manner for some t. Thus all solutions are obtained from a given one by a scaling of the v's, but this does not correspond to a scaling (in the usual sense of the word) of the associated rigid configurations. In fact, when the v's scale by the factor t, the u's scale by $1/t$, so that the original z coordinates of the points in the configuration change [by Eqs. (28)] from $[(u + v)/2, (u - v)/2]$ to

$$(u + t^2v)/2t, \qquad (u - t^2v)/2t. \tag{29}$$

We will show that this corresponds to a nontrivial variation of the geometry of the configurations; the variation, however, occurs within constraints that are clearly describable, as we show explicitly below.

Before analyzing this variation for general t, we can see immediately that the change that results from multiplying the v's by $t = -1$ corresponds to the reflection of the configuration in the image (x, y) plane. In fact, it is evident from expressions (29) that the effect of $t = -1$ is simply to change the signs of the z coördinates. Thus the one-parameter family of interpretations here contains the image-plane reflection ambiguity that is ubiquitous in structure-from-motion theories based on orthographic projection. Traditionally in such theories, for appropriate image data the 3-D interpretation is unique up to this reflection ambiguity; this occurs, for example, in the cases of rigid interpretations based on three views of four points[2] and fixed-axis rotation interpretations based on three views of three points.[4] In the present case, the one-parameter family of interpretations has two connected components, namely, $t > 0$ and $t < 0$; these sets are disconnected because $t = 0$ does not correspond to a solution of our equations (if $d_{l,l} \neq 0$). Thus the usual twofold reflection ambiguity expresses itself here as two connected components of a 1-D ambiguity.

We now present the general analysis. To review our notation, suppose that s is a point of S that corresponds to image data consisting of $2n$ vectors $\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{n,1}; \mathbf{b}_{1,2}, \ldots, \mathbf{b}_{n,2}$ in $\mathbf{R}^2$. We denote the coordinates of $\mathbf{b}_{i,j}$ by $(x_{i,j}, y_{i,j})$. A rigid interpretation for these data consists of $2n$ vectors $\mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1}; \mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2}$ in $\mathbf{R}^3$, such that the coordinates $(x_{i,j}, y_{i,j}, z_{i,j})$ of $\mathbf{a}_{i,j}$ satisfy Eqs. (10); thus the rigid interpretation arises as a solution of Eqs. (10) for the $z_{i,j}$'s when the $(x_{i,j}, y_{i,j})$ are given. We also use the notation $\mathbf{a}_{i,j} = \mathbf{b}_{i,j} + z_{i,j}\hat{e}_3$, where $\hat{e}_3$ is the unit vector in the z direction. For simplicity we denote $z_{l,2}$ by z_l and $z_{l,1}$ by w_l, and we let $v_l = z_l - w_l$. We further recall [Eq. (22)] that, in terms of the v's, the existence of the solution is equivalent to the existence, for each l and m, of numbers $\alpha_{l,m}$ such that $v_l = \alpha_{l,m}v_m$ and such that $\alpha_{p,q}\alpha_{q,r} = \alpha_{p,r}$ for each p, q, and r.

Now, since $\mathbf{a}_{i,1}$ and $\mathbf{a}_{i,2}$ are the successive positions of the ith point in the configuration, it follows that for any $1 \leq l < m \leq n$ the axis of rotation of the configuration is collinear with the vector

$$(\mathbf{a}_{l,2} - \mathbf{a}_{l,1}) \times (\mathbf{a}_{m,2} - \mathbf{a}_{m,1}) = [(\mathbf{b}_{l,2} + z_{l,2}\hat{e}_3) - (\mathbf{b}_{l,1} + z_{l,1}\hat{e}_3)] \times [(\mathbf{b}_{m,2} + z_{m,2}\hat{e}_3) - (\mathbf{b}_{m,1} + z_{m,1}\hat{e}_3)].$$

From our notation $v_l = z_l - w_l = z_{l,2} - z_{l,1}$ and $v_m = z_{m,2} - z_{m,1}$, this becomes

$$[(\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) + v_l\hat{e}_3] \times [(\mathbf{b}_{m,2} - \mathbf{b}_{m,1}) + v_m\hat{e}_3];$$

since $v_l = \alpha_{l,m}v_m$, we may write it thus:

$$[(\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) + \alpha_{l,m}v_m\hat{e}_3] \times [(\mathbf{b}_{m,2} - \mathbf{b}_{m,1}) + v_m\hat{e}_3].$$

Expanding this expression, we obtain

$$[(\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) \times (\mathbf{b}_{m,2} - \mathbf{b}_{m,1})] + [(\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) \times v_m\hat{e}_3] + [\alpha_{l,m}v_m\hat{e}_3 \times (\mathbf{b}_{m,2} - \mathbf{b}_{m,1})].$$

Let us divide and multiply the second term in this sum by $\alpha_{l,m}$:

$$[(\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) \times (\mathbf{b}_{m,2} - \mathbf{b}_{m,1})] + [(1/\alpha_{l,m})(\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) \times (\alpha_{l,m}v_m\hat{e}_3)] + [\alpha_{l,m}v_m\hat{e}_3 \times (\mathbf{b}_{m,2} - \mathbf{b}_{m,1})].$$

After rearranging, we obtain

$$\alpha_{l,m}v_m\hat{e}_3 \times [(\mathbf{b}_{m,2} - \mathbf{b}_{m,1}) - (1/\alpha_{l,m})(\mathbf{b}_{l,2} - \mathbf{b}_{l,1})] + [(\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) \times (\mathbf{b}_{m,2} - \mathbf{b}_{m,1})]. \quad (30)$$

Note that for $l < m$ we may define $\alpha_{m,l} = 1/\alpha_{l,m}$; this is consistent with Eq. (20). We may then denote

$$\begin{aligned}\mathbf{p} &= \hat{e}_3 \times [(\mathbf{b}_{m,2} - \mathbf{b}_{m,1}) - (1/\alpha_{l,m})(\mathbf{b}_{l,2} - \mathbf{b}_{l,1})] \\ &= \hat{e}_3 \times [(\mathbf{b}_{m,2} - \mathbf{b}_{m,1}) - \alpha_{m,l}(\mathbf{b}_{l,2} - \mathbf{b}_{l,1})], \\ \mathbf{c} &= (\mathbf{b}_{l,2} - \mathbf{b}_{l,1}) \times (\mathbf{b}_{m,2} - \mathbf{b}_{m,1}),\end{aligned} \quad (31)$$

where $\mathbf{p}$ and $\mathbf{c}$ are fixed vectors that depend only on the image data; $\mathbf{p}$ lies in the (x, y) plane, and $\mathbf{c}$ is collinear with the z axis. Thus, after replacing $\alpha_{l,m}v_m$ by v_l in expression (30), we obtain the following result: The axis of rotation of the configuration is collinear with the vector

$$\mathbf{c} + v_l\mathbf{p}. \quad (32)$$

Of course expression (32) varies with v_l, which represents the particular interpretation. We know that the set of all such interpretations is obtained, as v_l assumes all possible nonzero real values. We thus obtain

Theorem 2

Let s be a point of S corresponding to some distinguished image data $\mathbf{b} = \{\mathbf{b}_{i,j}\}$. Choose any numbers $1 \le l < m \le n$, and let the vectors $\mathbf{p}$ and $\mathbf{c}$ be as in Eqs. (31); then, for any $t \ne 0$ in $\mathbf{R}$, the rigid interpretation corresponding to $v_l = t$ is a configuration whose axis of rotation L is collinear with the vector

$$\mathbf{c} + t\mathbf{p}. \quad (33)$$

All the axes are lines through the origin in the plane determined by $\mathbf{p}$ and $\mathbf{c}$; this is a plane perpendicular to the x–y plane (see Fig. 4). As t varies, all possible such lines are realized, except for the line perpendicular to the (x, y) plane (which corresponds to the value $t = 0$, which is not permissible) and the line in the (x, y) plane (which corresponds to $t = \pm\infty$). In Fig. 5 we illustrate several cases of configurations corresponding to various values of t for fixed image data; the configuration is specified by the ellipses that are the projections into the (x, y) plane of the 3-D circles of rotation.

We now make various computations that are concerned with how the geometry of the configurations varies with t. First, we summarize briefly the results thus far: we are interested in rigid interpretations $\mathbf{a}$ consisting of the $2n$ vectors $\mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1}; \mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2}$ in $\mathbf{R}^3$. With our usual notation, we suppose that $\mathbf{a}_{l,i} = (x_{l,i}, y_{l,i}, z_{l,i})$, and we denote $v_l = z_{l,2} - z_{l,1}$, $u_l = z_{l,2} + z_{l,1}$. If our image data (which consist of the x and y coordinates) are distinguished, i.e., if the data lie in the locus S in $(\mathbf{R}^2)^{2n}$, then we can find numbers $\alpha_{l,m}$ as in Eqs. (18): for each (l, m), $v_l = \alpha_{l,m}v_m$. There is then a one-parameter family of possible rigid interpretations that corresponds to the one-parameter family of solutions of the system of linear equations [see Eq. (20)] for the v_l's. If $(v_1, \ldots, v_n)$ is a solution, so is $(tv_1, \ldots, tv_n)$ for any nonzero t. Thus the set of interpretations is parameterized naturally by nonzero values of v_l for any fixed l. However, we adopt the convention of letting $v_1 = t$. We can then view each v_l as a function of t, namely, $v_l(t) = \alpha_{l,1}t$; $v_1(1) = 1$.

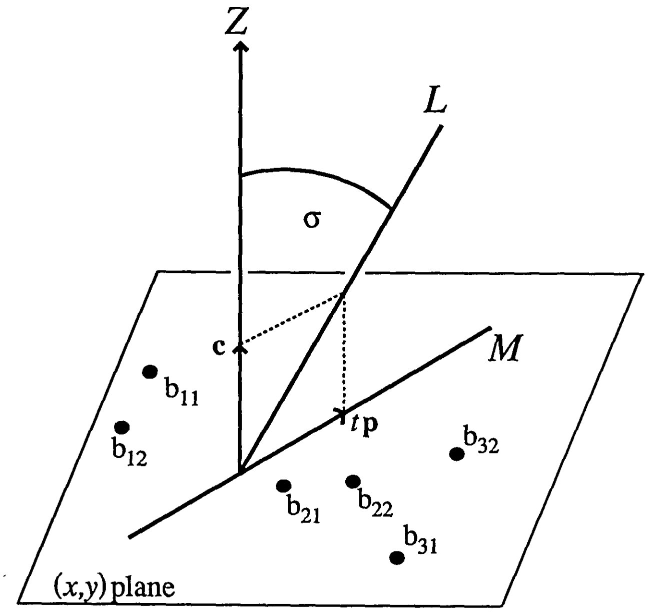

Fig. 4. Axis of rotation of the rigid interpretation (for the distinguished premise **b**), which corresponds to the parameter value t and is collinear with the vector $\mathbf{c} + t\mathbf{p}$: all the axes lie in the plane determined by $\mathbf{c}$ and $\mathbf{p}$.

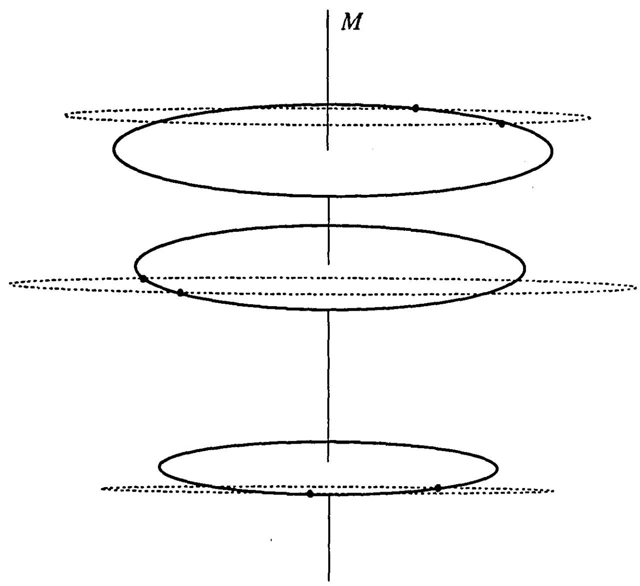

Fig. 5. Ellipses for the interpretations corresponding to two distinct values of t (for the same image data $\{\mathbf{b}_{i,j}\}$).

Axis of Rotation

For any $l \neq m$ we can define vectors $\mathbf{p}$ and $\mathbf{c}$ as in Eqs. (31), and then for any nonzero value of v_l there is a unique interpretation $\mathbf{a}$ whose axis of rotation is collinear with the vector $v_l\mathbf{p} + \mathbf{c}$. If we use (l', m') instead of (l, m), the corresponding vectors $\mathbf{p}'$ and $\mathbf{c}'$ are different, but $v_{l'}\mathbf{p}' + \mathbf{c}'$ will be collinear with $v_l\mathbf{p} + \mathbf{c}$ [provided that $v_{l'}$ and v_l are components of the same solution vector $v_1, \ldots, v_n$ of Eq. (20)].

Thus, in order to have a standard representation for the axis of rotation, we can let $l = 1$ and $m = 2$ in theorem 2, and we obtain that, for some given image data $\mathbf{b}_{l,i} = (x_{l,i}, y_{l,i})$, the axis of rotation of the configuration for the parameter value $v_1 = t$ is collinear with the vector [recalling Eqs. (31)]

$$t\hat{e}_3[(\mathbf{b}_{2,2} - \mathbf{b}_{2,1}) - \alpha_{2,1}(\mathbf{b}_{1,2} - \mathbf{b}_{1,1})] + [(\mathbf{b}_{1,2} - \mathbf{b}_{1,1}) \times (\mathbf{b}_{2,2} - \mathbf{b}_{2,1})] = tA\hat{e}_1 - tB\hat{e}_2 + C\hat{e}_3,$$

where

$$\begin{aligned} A &= \alpha_{2,1}(y_{1,2} - y_{1,1}) - (y_{2,2} - y_{2,1}), \\ B &= \alpha_{2,1}(x_{1,2} - x_{1,1}) - (x_{2,2} - x_{2,1}), \\ C &= (x_{1,2} - x_{1,1})(y_{2,2} - y_{2,1}) - (x_{2,2} - x_{2,1})(y_{1,2} - y_{1,1}). \end{aligned} \tag{34}$$

We will denote the unit vector in this direction by $\hat{r}(t)$, so that

$$\hat{r}(t) = \frac{tA\hat{e}_1 - tB\hat{e}_2 + C\hat{e}_3}{[t^2(A^2 + B^2) + C^2]^{1/2}}. \tag{35}$$

Centers of Rotation

Now suppose that $\mathbf{a}(t)$ is the interpretation corresponding to the parameter value $v_1 = t$; then the center of the rotation of the point $\mathbf{a}_{l,1}(t)$ [into the point $\mathbf{a}_{l,2}(t)$] is the projection of $\mathbf{a}_{l,1}(t)$ onto the axis of rotation, i.e.,

$$[\mathbf{a}_{l,1}(t) \cdot \hat{r}(t)]\hat{r}(t). \tag{36}$$

We now compute this. We have

$$\mathbf{a}_{l,1}(t) = [x_{l,1}, y_{l,1}, z_{l,1}(t)] = \{x_{l,1}, y_{l,1}, [u_l(t) - v_l(t)]/2\}.$$

Let us denote

$$\bar{u}_l = u_l(1),$$

$$\bar{v}_l = v_l(1);$$

then from expressions (29) we have

$$[u_l(1) - v_l]/2 = (\bar{u}_l - t^2\bar{v}_l)/2t,$$

so we may write

$$\mathbf{a}_{l,1} = [x_{l,1}, y_{l,1}, (\bar{u}_l - t^2\bar{v}_l)/2t]. \tag{37}$$

Then, using the expression for $\hat{r}(t)$ given in Eq. (35), we obtain the following proposition.

Proposition 3

With our usual notation, if $\mathbf{a}(t)$ is the rigid interpretation corresponding to the parameter value $v_1 = t$, then for any $l = 1, \ldots, n$ the center of rotation of $\mathbf{a}_{l,1}(t)$ into $\mathbf{a}_{l,2}(t)$ is the point $(c_1^{(l)}(t), c_2^{(l)}(t), c_3^{(l)}(t)]$ of $\mathbf{R}^3$ given by

$$\begin{aligned} c_1^{(l)}(t) &= \frac{2At^2(Ax_{l,1} - By_{l,1}) + AC\bar{u}_l - ACt^2\bar{v}_l}{2[t^2(A^2 + B^2) + C^2]}, \\ c_2^{(l)}(t) &= \frac{-2Bt^2(Ax_{l,1} - By_{l,1}) - BC\bar{u}_l + BCt^2\bar{v}_l}{2[t^2(A^2 + B^2) + C^2]}, \\ c_3^{(l)}(t) &= \frac{2Ct^2(Ax_{l,1} - By_{l,1}) + C^2\bar{u}_l - C^2t^2\bar{v}_l}{2t[t^2(A^2 + B^2) + C^2]}, \end{aligned} \tag{38}$$

where A, B, and C are as in Eqs. (34) and where $\bar{u}_l = u_l(1)$ and $\bar{v}_l = v_l(1)$.

It follows that the projection into the image (x, y) plane of the lth center of rotation is the point $[c_1^{(l)}(t), c_2^{(l)}(t)]$ given by

$$\left\{\frac{2At^2(Ax_{l,1} - By_{l,1}) + AC\bar{u}_l - ACt^2\bar{v}_l}{2[t^2(A^2 + B^2) + C^2]}, \frac{-2Bt^2(Ax_{l,1} - By_{l,1}) - BC\bar{u}_l + BCt^2\bar{v}_l}{2[t^2(A^2 + B^2) + C^2]}\right\}. \tag{39}$$

The dependence of this point on the parameter t is illustrated in Fig. 6. The point $[c_1^{(l)}(t), c_2^{(l)}(t)]$ is the center of the

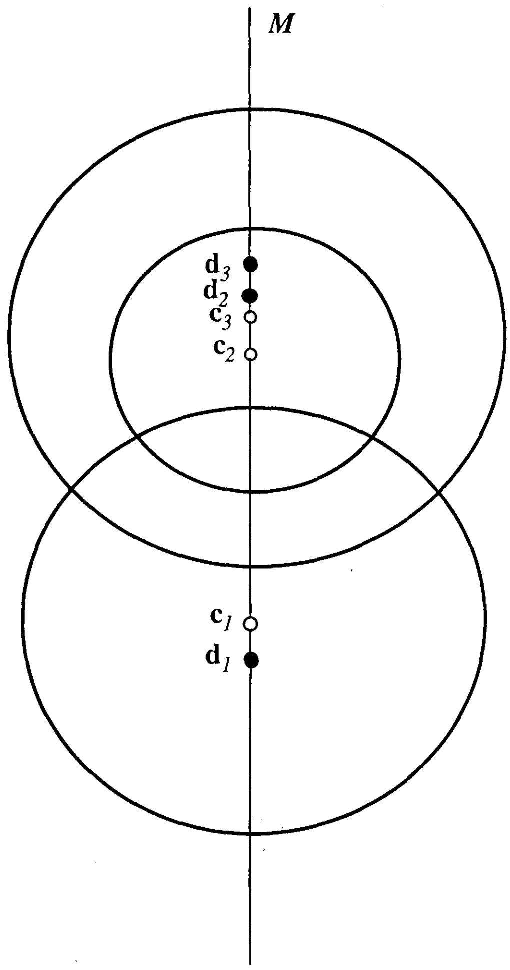

Fig. 6. The actual centers of the ellipses for an interpretation corresponding to a small value of t ($t = 0.25$) are the $\mathbf{c}_i$, marked with filled circles; the limiting projected centers as $t \to \infty$ are the $\mathbf{d}_i$, marked with open circles. Note that the ellipses are almost circular for this value of t but there is still a clear discrepancy between the $\mathbf{c}_i$ and the $\mathbf{d}_i$.

ellipse in the image plane that is the projection of the circle of rotation in $\mathbf{R}^3$ of the point $\mathbf{a}_{l,1}(t)$ into the point $\mathbf{a}_{l,2}(t)$; of course, for any t the points $\mathbf{a}_{l,1}(t)$ and $\mathbf{a}_{l,2}(t)$ project to the image points $\mathbf{b}_{l,1} = (x_{l,1}, y_{l,1})$ and $\mathbf{b}_{l,2} = (x_{l,2}\, y_{l,2})$. As t varies, the projected centers are constrained to move on the line M generated by $\mathbf{p}$ [as in Eqs. (31)], since this line is the common projection of the axes of rotation for all t.

We note that for generic image data (for which $d_{l,l} \neq 0$), the image points $\mathbf{b}_{l,1}$ and $\mathbf{b}_{l,2}$ cannot be equidistant from the projected center of rotation $[c_1^{(l)}(t), c_2^{(l)}(t)]$ for any permissible, i.e., nonzero, value of t. In fact, if they are equidistant from this projected center, then the ellipse is a circle so that the axis of rotation points in the z direction; but this corresponds to $t = 0$, a contradiction. However, we have the following proposition.

Proposition 4

As $t \to 0$, the axis direction $\hat{r}(t)$ approaches the z axis, and the center of rotation $[c_1^{(l)}(t), c_2^{(l)}(t), c_3^{(l)}(t)]$ (for any l) approaches ∞ along this axis. However the limiting position of the projected center of rotation [expression (39)] is the unique point on the line M that is equidistant from the image points $\mathbf{b}_{l,1}$ and $\mathbf{b}_{l,2}$.

Proof

That the axis direction approaches the z axis as $t \to 0$ follows from theorem 2. That the center of rotation goes to ∞ as $t \to 0$ follows from the fact that $c_3^{(l)}(t) \to \infty$ as $t \to 0$; this fact is immediate from Eqs. (38).

To prove the latter assertion, we will consider the case in which $l = 1$; this is sufficient, since the points can always be renumbered so that the former lth point becomes the first point in the new ordering. It is then evident from expression (39) that the limiting position of the projected center is

$$\mathbf{d}_1 = (A\bar{u}_1/2C, -B\bar{u}_1/2C). \tag{40}$$

We then compute

$$|\mathbf{b}_{1,1} - \mathbf{d}_1|^2 - |\mathbf{b}_{1,2} - \mathbf{d}_1|^2 = [(x_{1,1} - A\bar{u}_1/2C)^2 + (y_{1,1} + B\bar{u}_1/2C)^2] - [(x_{1,2} - A\bar{u}_1/2C)^2 + (y_{1,2} + B\bar{u}_1/2C)^2].$$

Expanding and simplifying, we may write this expression in the form

$$({x_{1,1}}^2 - {x_{1,2}}^2) + ({y_{1,1}}^2 - {y_{1,2}}^2) + \frac{(x_{1,2} - x_{1,1})A\bar{u}_1 - (y_{1,2} - y_{1,1})B\bar{u}_1}{C}. \tag{41}$$

If we substitute in for A and B by using Eqs. (34), the numerator of the second term in expression (41) becomes

$$\bar{u}_1(x_{1,2} - x_{1,1})[\alpha_{2,1}(y_{1,2} - y_{1,1}) - (y_{2,2} - y_{2,1})] - \bar{u}_1(y_{1,2} - y_{1,1})[\alpha_{2,1}(x_{1,2} - x_{1,1}) - (x_{2,2} - x_{2,1})].$$

When we expand this, the $\bar{u}_1\alpha_{2,1}$ terms cancel, and the expression is simply

$$-\bar{u}_1[(x_{1,2} - x_{1,1})(y_{2,2} - y_{2,1}) - (x_{2,2} - x_{2,1})(y_{1,2} - y_{1,1})] = -\bar{u}_1 C.$$

Hence the second term in expression (41) is just $-\bar{u}_1$; i.e., expression (41) is

$$({x_{1,1}}^2 - {x_{1,2}}^2) + ({y_{1,1}}^2 - {y_{1,2}}^2) - \bar{u}_1.$$

Noting that $\bar{v}_1 = v_1(1) = 1$ by convention, we may write this as follows:

$$({x_{1,1}}^2 - {x_{1,2}}^2) + ({y_{1,1}}^2 - {y_{1,2}}^2) - \bar{u}_1\bar{v}_1.$$

But in configuration $\mathbf{a}(1)$ we have $\bar{u}_1 = z_{1,2} + z_{1,1}$, $\bar{v}_1 = z_{1,2} - z_{1,1}$, so that $\bar{u}_1\bar{v}_1 = {z_{1,1}}^2 - {z_{1,2}}^2$. Thus expression (41) is

$${x_{1,1}}^2 + {y_{1,1}}^2 + {z_{1,1}}^2 - {x_{1,2}}^2 - {y_{1,2}}^2 - {z_{1,2}}^2 = |\mathbf{a}_{1,1}(1)|^2 - |\mathbf{a}_{1,2}(1)|^2 = 0,$$

since configuration $\mathbf{a}(1)$ is rigid. This concludes the proof of proposition 4. ∎

Angle of Rotation

We know that, for each $t \neq 0$, there is a rotation through some angle $\theta(t)$ about an axis collinear with the vector $\hat{r}(t)$; this rotation carries $\mathbf{a}_{l,1}(t)$ into $\mathbf{a}_{l,2}(t)$ for each $l = 1, \ldots, n$. We now study the variation of the angle $\theta(t)$ with the parameter t.

Let $R(t)$ denote the matrix of the rotation in question for the parameter value t. We can then express the angle $\theta(t)$ of this rotation by the formula

$$\cos[\theta(t)] = \frac{\mathrm{Tr}\, R(t) - 1}{2}, \tag{42}$$

where Tr denotes the trace of the matrix, i.e., the sum of its diagonal elements. (Since the trace of a linear map is independent of the coordinate system, this formula may be proved in general simply by verifying it in the case in which the axis of rotation is, say, the z axis.)

Suppose that $\mathbf{a}(t)$ is a rigid interpretation, consisting of points $\mathbf{a}_{l,i}(t) = (x_{l,i}, y_{l,i}, z_{l,i})$, $l = 1, \ldots, n$; $i = 1, 2$; note that $z_{l,i}$ depends on t, but we exclude this from the notation for the sake of simplicity. The matrix $R(t)$ then is specified uniquely by

$$\begin{bmatrix} x_{l,2} \\ y_{l,2} \\ z_{l,2} \end{bmatrix} = R(t) \begin{bmatrix} x_{l,1} \\ y_{l,1} \\ z_{l,1} \end{bmatrix}$$

for $l = 1, 2, 3$. Thus, if we let

$$T_i(t) = \begin{bmatrix} x_{1,i} & x_{2,i} & x_{3,i} \\ y_{1,i} & y_{2,i} & y_{3,i} \\ z_{1,i} & z_{2,i} & z_{3,i} \end{bmatrix}, \qquad i = 1, 2,$$

then $R(t)$ is specified uniquely by the matrix equation $R(t)T_1(t) = T_2(t)$; i.e.,

$$R(t) = T_2(t){T_1}^{-1}(t). \tag{43}$$

From this, we compute directly that

$$\mathrm{Tr}[R(t)] = \frac{\begin{bmatrix} (x_{1,1}y_{2,1} - x_{2,1}y_{1,1})z_{3,2} + (x_{1,1}y_{2,2} - x_{2,2}y_{1,1})z_{3,1} + (x_{1,2}y_{2,1} - x_{2,1}y_{1,2})z_{3,1} \\ + (x_{3,1}y_{1,1} - x_{1,1}y_{3,1})z_{2,2} + (x_{3,1}y_{1,2} - x_{1,2}y_{3,1})z_{2,1} + (x_{3,2}y_{1,1} - x_{1,1}y_{3,2})z_{2,1} \\ + (x_{2,1}y_{3,1} - x_{3,1}y_{2,1})z_{1,2} + (x_{2,1}y_{3,2} - x_{3,2}y_{2,1})z_{1,1} + (x_{2,2}y_{3,1} - x_{3,1}y_{2,2})z_{1,1} \end{bmatrix}}{(x_{1,1}y_{2,1} - x_{2,1}y_{1,1})z_{3,1} + (x_{3,1}y_{1,1} - x_{1,1}y_{3,1})z_{2,1} + (x_{2,1}y_{3,1} - x_{3,1}y_{2,1})z_{1,1}}.$$

To simplify the notation, let us write this in the form

$$\frac{\begin{pmatrix} F_{1,1}z_{3,2} + F_{1,2}z_{3,1} + F_{1,3}z_{3,1} \\ + F_{2,1}z_{2,2} + F_{2,2}z_{2,1} + F_{2,3}z_{2,1} \\ + F_{3,1}z_{1,2} + F_{3,2}z_{1,1} + F_{3,3}z_{1,1} \end{pmatrix}}{D_3 z_{3,1} + D_2 z_{2,1} + D_1 z_{1,1}}, \tag{44}$$

where $F_{1,1} = (x_{1,1}y_{2,1} - x_{2,1}y_{1,1})$, $F_{1,2} = (x_{1,1}y_{2,2} - x_{2,2}y_{1,1})$, etc. and $D_3 = (x_{1,1}y_{2,1} - x_{2,1}y_{1,1})$, etc. We now express the dependence on t by substituting $z_{l,1} = (\bar{u}_l - t^2\bar{v}_l)/2t$, $z_{l,2} = (\bar{u}_l + t^2\bar{v}_l)/2t$ into expression (44), where, as stated above, $\bar{u}_l = u_l(1)$, $\bar{v}_l = v_l(1)$. We get

$$\frac{\begin{bmatrix} F_{1,1}(\bar{u}_3 + t^2\bar{v}_3)/2t + F_{1,2}(\bar{u}_3 - t^2\bar{v}_3)/2t + F_{1,3}(\bar{u}_3 - t^2\bar{v}_3)/2t \\ + F_{2,1}(\bar{u}_2 + t^2\bar{v}_2)/2t + F_{2,2}(\bar{u}_2 - t^2\bar{v}_2)/2t + F_{2,3}(\bar{u}_2 - t^2\bar{v}_2)/2t \\ + F_{3,1}(\bar{u}_1 + t^2\bar{v}_1)/2t + F_{3,2}(\bar{u}_1 - t^2\bar{v}_1)/2t + F_{3,3}(\bar{u}_1 - t^2\bar{v}_1)/2t \end{bmatrix}}{D_3(\bar{u}_3 - t^2\bar{v}_3)/2t + D_2(\bar{u}_2 - t^2\bar{v}_2)/2t + D_1(\bar{u}_1 - t^2\bar{v}_1)/2t}. \tag{45}$$

Here we may collect terms and cancel the common factor 1/2t in the numerator and denominator, to obtain

$$\frac{\begin{bmatrix} (F_{1,1} + F_{1,2} + F_{1,3})\bar{u}_3 + (F_{2,1} + F_{2,2} + F_{2,3})\bar{u}_2 + (F_{3,1} + F_{3,2} + F_{3,3})\bar{u}_1 \\ + (F_{1,1} - F_{1,2} - F_{1,3})t^2\bar{v}_3 + (F_{2,1} - F_{2,2} - F_{2,3})t^2\bar{v}_2 + (F_{3,1} - F_{3,2} - F_{3,3})t^2\bar{v}_1 \end{bmatrix}}{(D_1\bar{u}_1 + D_2\bar{u}_2 + D_3\bar{u}_3) - t^2(D_1\bar{v}_1 + D_2\bar{v}_2 + D_3\bar{v}_3)}. \tag{46}$$

Thus we have found that

$$\mathrm{Tr}\, R(t) = (A + t^2B)/(C + t^2D), \tag{47}$$

where

$$A = (F_{1,1} + F_{1,2} + F_{1,3})\bar{u}_3 + (F_{2,1} + F_{2,2} + F_{2,3})\bar{u}_2 + (F_{3,1} + F_{3,2} + F_{3,3})\bar{u}_1,$$

$$B = (F_{1,1} - F_{1,2} - F_{1,3})\bar{v}_3 + (F_{2,1} - F_{2,2} - F_{2,3})\bar{v}_2 + (F_{3,1} - F_{3,2} - F_{3,3})\bar{v}_1,$$

$$C = D_1\bar{u}_1 + D_2\bar{u}_2 + D_3\bar{u}_3,$$

$$D = -(D_1\bar{v}_1 + D_2\bar{v}_2 + D_3\bar{v}_3).$$

Note that the quantities A, B, C, and D depend only on the image data (and not on t).

Finally, we obtain

$$\cos[\theta(t)] = \frac{(A + t^2B)/(C + t^2D) - 1}{2}, \tag{48}$$

where A, B, C, and D are independent of t.

Our goal here is to determine whether the geometric quantity $\theta(t)$ attains any extrema in our one-parameter family of configurations. Accordingly, we differentiate the right-hand side of Eq. (48) with respect to t and set the result equal to 0. For this purpose we may simply differentiate $(A + t^2B)/(C + t^2D)$ and set it equal to 0, to obtain

$$2tB(C + t^2D) - 2tD(A + t^2B) = 0,$$

which simplifies to become

$$t(BC - AD) = 0. \tag{49}$$

Thus, for generic image data compatible with rigid interpretations, the only possible extremum occurs when $t = 0$, but this is not a permissible value of t, since it corresponds to no actual rigid structure. However, it is a minimum of the function of t given in Eq. (48), and as t goes to $\pm\infty$ the angle $\theta(t)$ increases monotonically to 180°. In fact, we make the following proposition.

Proposition 5

As t increases through positive or negative values, the angle $\theta(t)$ is strictly increasing: it attains no extrema in the family of rigid interpretations for the given image data. The limiting value of $\theta(t)$ as $t \to 0$ may be described as follows: For any $l = 1, \ldots, n$, it is the angle subtended at the limit position as $t \to 0$ of the lth projected center by the image points $\mathbf{b}_{l,1}$ and $\mathbf{b}_{l,2}$ (these points are equidistant from that limit position by proposition 4).

Radii of Rotation

The radius of rotation $r^{(l)}(t)$ of the lth point in the configuration $\mathbf{a}(t)$ is the distance of, say, $\mathbf{a}_{l,1}(t)$ from the center of rotation $[c_1^{(l)}(t), c_2^{(l)}(t), c_3^{(l)}(t)]$. Thus

$$r^{(l)}(t)^2 = [c_1^{(l)}(t) - x_{l,1}]^2 + [c_2^{(l)}(t) - y_{l,1}]^2 + \left[c_3^{(l)}(t) - \frac{\bar{u}_l - t^2\bar{v}_l}{2t}\right]^2. \tag{50}$$

We have derived an expression for the $c_j^{(l)}(t)$ terms in Eqs. (38), so we can insert this into Eq. (50) to get a computable expression for $r^{(l)}(t)^2$ as a function of t. We find that this function has a local minimum at $t = 0$ and has several other local minima at certain nonzero values of t. However, there is not, in general, a single nonzero value of t that is simultaneously a local minimum of the functions $r^{(l)}(t)$ for the different indices $l = 1, \ldots, n$; this may be seen by checking a generic example.

We may summarize the preceding analyses as follows: The one-parameter family of rigid interpretations (for some given distinguished image data) has no feature points: there is no member of the family for which there is attained an extremum of the angle of rotation. Similarly there is no structure in the family for which the radii of rotation of all the points in the configuration have simultaneous extrema. However, in the limiting case corresponding to $t = 0$, these quantities are minimized. This case does not represent an actual rigid structure but is a virtual structure, which displays the limiting geometric properties of the family, and in that sense it may be regarded as a token of the family.

The fact that there is no mathematically singular member of the family itself means that, of all the possible rigid interpretations compatible with the image data, there is no canonical choice of an actual rigid interpretation for a subject who is presented with those data. Yet there is no doubt that the subject sees a rigid object, in the sense that the subjective experience of rigidity is highly correlated with the event that the image data permit a rigid interpretation (i.e., the event that the image data lies in the distinguished premise set S.) This raises the question of which, if any, of the actual rigid structures is being seen. Moreover, since the virtual structure corresponding to the limiting case in which $t = 0$ is so strongly mathematically singular with regard to the family, we might ask whether it plays any role in the perceptual representation.

In this regard it would be desirable to design, if possible, a psychophysical experiment that would produce significant statistics for the image-plane projection of the centers of rotation perceived by the subject. For example, the subjects' task might be to locate each of the n centers of rotation for a given image configuration on the display screen. One interesting possibility is that the perceived centers of rotation may be those that correspond to the limiting value $t = 0$, even though this would be incompatible, in general, with the holistic percept of rigid rotation (unless the image display is consistent with an interpretation of rotation about an axis parallel to the viewing direction, and we may assume that this is not the case). However, if we are to draw any conclusions here, the experiment must produce data samples with rather small variance, and it remains to be seen whether this can be accomplished.

On the other hand, regardless of whether subjects can precisely locate individual centers of rotation, it is a reasonable hypothesis that a given subject will display a characteristic preference for perceiving certain members of the family. In other words, we would expect that each subject has an interpretation kernel; i.e., for each set of image data there is a probability distribution on the set of rigid interpretations that are compatible with those image data. This is the distribution of the rigid interpretations "output" by that subject while viewing those image data. The study of this hypothesis again poses a problem for the design of appropriate psychophysical experiments, but results from pilot studies by Braunstein[21] are at least broadly suggestive. For example, it is evident that the structures that correspond to large values of t (and therefore involve large angles of rotation) are least likely to be seen, without effort, by most subjects.

4. OBSERVER THEORY

In this section we recall briefly the formal definition of an observer, in the sense of observer theory (cf. Refs. 22–24). This definition states the structure common to all perceptual inferences, regardless of their modality. Our goal in this section is limited to indicating how the structure-from-motion inference treated in this paper is one instance of the general observer definition. More-detailed expositions of the motivation and philosophy of observer theory, and some of its further developments, may be found in the papers cited above.

An observer is a sextuple $\mathcal{O} = (X, Y, E, S, \pi, \eta)$ satisfying the following conditions (see Fig. 7):

(a) X and Y are measurable spaces.[25] X denotes the configuration space of $\mathcal{O}$, and Y denotes its premise space.

(b) E is a measurable subset of X, called the set of distinguished configurations; S is a measurable subset of Y, called the set of distinguished premises.

(c) $\pi: X \to Y$ is a measurable, surjective function[26] such that $\pi(E) = S$. π is called the perspective map of the observer.

(d) η is the interpretation kernel of $\mathcal{O}$. It provides, for each $s \in S$, a way to assign probabilities to measurable sets in $\pi^{-1}(s) \cap E$; if $A \subset \pi^{-1}(s) \cap E$, the corresponding probability

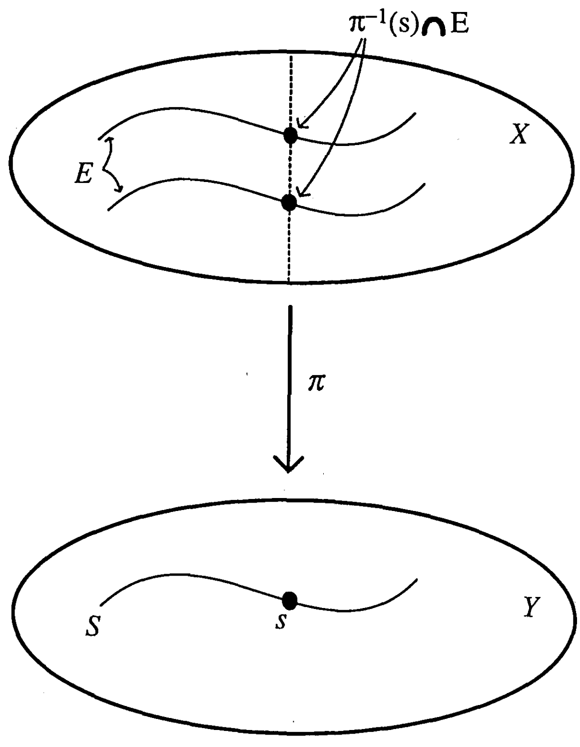

Fig. 7. Schematic of an observer. Y is the space of premises. S is the distinguished premises. X is the configuration space. E is the distinguished configurations. For each premise $s \in S$, the conclusion of the observer is a probability measure $\eta(s, \cdot)$ supported on $\pi^{-1}(s) \cap E$ (which happens to consist of two points in the case illustrated).

is denoted $\eta(s, A)$. We denote by $\eta(s, \cdot)$ this assignment $A \rightarrow \eta(s, A)$; thus $\eta(s, \cdot)$ is a probability measure on $\pi^{-1}(s) \cap E$.

Intuitively, the premise space Y is a mathematical representation of the set of proximal stimuli for the observer's inferences. The configuration space X provides the syntax for the observer's perceptual interpretations; the single configurations, represented by the elements of X, are the atoms of the syntax. The perspective map π determines what configurations are compatible with a given premise: the configurations compatible with the premise y in Y are those in $\pi^{-1}(y)$. The observer is biased: when presented with a premise y not in S, the observer concludes only that there is no compatible distinguished interpretation. This is a deductively valid inference. If the premise is in S, say, s, then its conclusion is the probability measure $\eta(s, \cdot)$. In this way nonzero probabilities are assigned only to interpretations in $\pi^{-1}(s) \cap E$, and in this sense E expresses the observer's bias.

We note that, in the case of a premise s in S, the observer's inference fails to be deductively valid for two reasons. First, the observer will assign probabilities only to interpretations compatible with s that are also in E. However, it is possible that the premise s arose from interaction of the observer with a state of affairs that corresponds to a configuration in $\pi^{-1}(s) - E$, i.e., a nondistinguished configuration that is nevertheless compatible with the distinguished premise s. Such a state of affairs (or the configuration to which it corresponds) is a false target. The second reason why, given the distinguished premise, the inference fails to be deductively valid is that there is, in general, not just one point in $\pi^{-1}(s) \cap E$ that is assigned a positive measure by $\eta(s, \cdot)$. In fact, the inference is, in general, fundamentally probabilistic and not simply a result of noise: there is a fundamental perceptual uncertainty that corresponds to the fact that the map π (even when restricted to E) is many to one.

In this paper we study a particular observer $\mathcal{O}$, namely, the instantaneous-rotation observer. The reader will note that we have been using the observer notation and terminology from the outset. For this particular $\mathcal{O}$, we have $X = (\mathbf{R}^3)^{2n}$ and $Y = (\mathbf{R}^2)^{2n}$, and $\pi: X \rightarrow Y$ corresponds to the orthogonal projection of each copy of $\mathbf{R}^3$ onto its (x, y) plane $\mathbf{R}^2$. [The measurable structure on X and Y may be taken to be the usual one for Euclidean spaces: $\sigma(X)$ is the smallest σ algebra containing all spheres.] If we view an element of X as two n-tuples of vectors in $\mathbf{R}^3$, then the distinguished configurations, E, consist of those for which the second n-tuple of points is related to the first n-tuple by a rotation about an axis through the origin. We can then define the set of distinguished premises $S \subset Y$ to be the image $\pi(E)$ of E in Y. In Section 2 of this paper we have found equations for E, and then in Section 3 we have deduced effective conditions to determine whether a given premise y is in S; these conditions are a representation of the first part of the observer's inference procedure, namely, the decision about whether a premise is distinguished. In Section 3 we have analyzed the set $\pi^{-1}(s) \cap E$: we determined that this set is parameterized by nonzero real numbers t, etc.

According to the observer definition, the observer's conclusion, given the premise s, is a probability measure on this set; the various conclusions corresponding to different points s of S are collected in the interpretation kernel η. We have noted in the last part of Section 3 that there is in general no choice of a single element in $\pi^{-1}(s)$ that for purely mathematical reasons is destined to serve as a canonical interpretation, given the premise s. In other words, there is no canonical choice for an interpretation kernel η in which the probability measure $\eta(s, \cdot)$ assigns a probability of 1 to a particular point of $\pi^{-1}(s) \cap E$ and a probability of 0 to all other points in $\pi^{-1}(s) \cap E$. For this reason, the interpretation kernel η, such as it may be, contains nontrivial information about the structure of the observer. The hypothetical psychophysical studies mentioned briefly at the end of Section 3 have as their goal the determination of η for individual subjects or classes of subjects. In any case the psychophysical studies already undertaken (in collaboration with Braunstein[21]) suggest strongly that our abstract observer is, in fact, instantiated in the human perceptual system. (This does not mean, however, that there necessarily exists in the brain a neural network that solves the algebraic equations that define E and S.)

5. GROUP ACTION

We now introduce a new observer $\mathcal{O}'$ whose premise space Y is the same as that of our instantaneous-rotation observer $\mathcal{O}$ but whose configuration space X' and distinguished configurations E' are natural augmentations of the corresponding sets X and E for $\mathcal{O}$. This new observer is of particular interest, because there is here a well-behaved group action (defined almost everywhere) on X' and E'. It is tempting to assert that this group action models a perceptually natural group of mental geometric transformations of the configurations. Here we content ourselves with describing the mathematical structure and using it to obtain results about the geometry of our original observer $\mathcal{O}$. In particular this approach enables us to compute the dimensions of E and S, information that is not easy to obtain directly, even from the explicit algebraic conditions defining E and S that we have derived in Sections 2 and 3 above.

Let $n \geq 3$ be given. We use our usual notation: $X = (\mathbf{R}^3)^{2n}$, $Y = (\mathbf{R}^2)^{2n}$; elements of X are denoted by x or by $\mathbf{a} = (\mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1}; \mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2})$; elements of Y are denoted $\mathbf{b} = (\mathbf{b}_{1,1}, \ldots, \mathbf{b}_{n,1}; \mathbf{b}_{1,2}, \ldots, \mathbf{b}_{n,2})$; $\pi: X \rightarrow Y$ is defined by $\pi(\mathbf{a}) = \mathbf{b}$, where $\mathbf{b}_{l,i}$ is the projection of $\mathbf{a}_{l,i}$ onto the (x, y) plane.

Terminology

An axis in $\mathbf{R}^3$ is an oriented line $\mathcal{A}$ through the origin, i.e., a line with its positive direction specified. We denote the set of such axes by $\mathbf{A}$.

The set $\mathbf{A}$ of axes corresponds to the set of points on the unit sphere $\mathbf{S}^2$ centered at the origin: each such point determines a line through the origin, whose positive direction is taken to be the direction from the origin to the point. Let

$$X' = \mathbf{A} \times X = \{(\mathcal{A}, x) | \mathcal{A} \in \mathbf{A}, x \in X\}. \tag{51}$$

X' is the configuration space of new observer $\mathcal{O}'$; an element $x' = (\mathcal{A}, x)$ of X' now is called a configuration. We call $\mathcal{A}$ the reference axis of the configuration x'.

We have a natural map

$$f: X' \rightarrow X$$

defined by

$$f(\mathcal{A}, x) = x; \tag{52}$$

$$\begin{tikzcd} X' \arrow[d, "p"] & \supset & E' \arrow[d, "p"] \\ Y & \supset & S \end{tikzcd} \qquad p = \pi \circ f$$

Fig. 8. X' is the configuration space, and E' is the distinguished configuration space, of the observer $\mathcal{O}'$. The premise space Y and distinguished premise space S for $\mathcal{O}'$ are the same as for the original instantaneous-rotation observer $\mathcal{O}$.

then we can define

$$p: X' \to Y$$

by

$$p = \pi \circ f, \qquad \text{i.e., } p(\mathcal{A}, x) = \pi(x). \tag{53}$$

p is the perspective map of $\mathcal{O}'$; Y is the premise space of $\mathcal{O}'$ as well as of $\mathcal{O}$.

Let E' be the set of those elements of X' in which the two n-tuples of points $(\mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1})$ and $(\mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2})$ of $\mathbf{R}^3$ are related by a rigid rotation of $\mathbf{R}^3$ about the given axis $\mathcal{A}$. Thus

$$E' = \{(\mathcal{A}; \mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1}; \mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2}) \in X' | \sigma \mathbf{a}_{i,1} = \mathbf{a}_{i,2};$$
$$1 \le i \le n; \quad \text{where } \sigma \in SO(3, \mathbf{R}) \text{ is a rotation about } \mathcal{A}\}. \tag{54}$$

E' is the distinguished configuration space of $\mathcal{O}'$. Thus we now use the terms distinguished configurations or instantaneous rotations to refer to the elements of E'. Note that

$$f(E') = E,$$

where f is as in Eq. (52). We note that while f: $X' \to X$ is infinite to one, its restriction to E' is (Lebesgue) generically two to one. For given a point $\mathbf{a}$ of E for which the vectors $\mathbf{a}_{1,1}, \mathbf{a}_{2,1}, \ldots, \mathbf{a}_{n,1}$ are linearly independent, the rotation σ of $\mathbf{R}^3$ such that $\sigma \mathbf{a}_{l,1} = \mathbf{a}_{l,2}$ for all $l = 1, \ldots, n$ is determined uniquely (since $n \ge 3$). Hence the axis of σ is determined uniquely as a linear subspace of $\mathbf{R}^3$ and therefore up to orientation as an element of $\mathbf{A}$. Thus for each $e \in E$ there are exactly two $\mathcal{A}$'s such that $(\mathcal{A}, e) \in E'$; these $\mathcal{A}$'s correspond to the same line through the origin but have opposite orientations.

Figure 8 illustrates schematically the relations among X', Y, E', S, and p. These spaces and maps together define the observer $\mathcal{O}'$ (except for the interpretation kernel as defined in Section 4).

Given an axis $\mathcal{A}$ and a point $\mathbf{a} \in \mathbf{R}^3$ that does not lie on $\mathcal{A}$, we shall denote this by $\mathbf{a} \notin \mathcal{A}$. We consider subsets

$$X_0' = \{[\mathcal{A}; (\mathbf{a}_{i,j})] \in X' | \mathbf{a}_{i,j} \notin \mathcal{A} \ \forall i, j\},$$
$$E_0' = E' \cap X_0'. \tag{55}$$

These sets emerge as principal homogeneous spaces for certain groups. For any element of E_0', the rotation σ corresponding to it [as in Eq. (54)] is determined uniquely by the axis $\mathcal{A}$ and any single pair $(\mathbf{a}_{i,1}, \mathbf{a}_{i,2})$. Note that a distinguished configuration in which some $\mathbf{a}_{i,j}$ lies on the axis $\mathcal{A}$ is a limit of distinguished configurations in which no point lies on the axis. Thus E' is the closure of E_0'.

We think of the configurations as corresponding to two successive positions of n points (plus origin) moving arbitrarily in $\mathbf{R}^3$, together with a choice of reference axis $\mathcal{A}$; successivity refers to some particular discrete time scale. In this sense, the set E' of augmented instantaneous rotations consists of those configurations that are in fact (rigid) rotations about their reference axis. We reiterate that every rigid motion of $n + 1$ points, one of which is fixed at the origin, corresponds to exactly two elements of E': $(\mathcal{A}, \mathbf{a})$ and $(\mathcal{A}', \mathbf{a})$, where $\mathcal{A}$ and $\mathcal{A}'$ are of opposite orientations.

Now, since $f(E') = E$, it follows that $p(E') = \pi(E) = S$, where $S \subset Y$ is the same set of distinguished premises as before; S consists of those image data (i.e., two views of n points) that are compatible with at least one rigid interpretation. We let

$$S_0 = p(E_0'). \tag{56}$$

We note that, since E' is the closure of E_0', S is the closure of S_0.

For the purpose of our group action we first describe X_0' and E_0' in cylindrical coordinates. We represent each element $x' \in X_0'$ in the form

$$x' = \begin{pmatrix} \mathcal{A} & \begin{matrix} \hat{v}, r_{2,1}, \ldots, r_{n,1} & h_{1,1}, \ldots, h_{n,1} & l_{1,1}, \ldots, l_{n,1} \\ r_{1,2}, r_{2,2}, \ldots, r_{n,2} & h_{1,2}, \ldots, h_{n,2} & l_{1,2}, \ldots, l_{n,2} \end{matrix} \end{pmatrix}, \tag{57}$$

where we have fixed a coordinate system in $\mathbf{R}^3$ and where the symbols are defined as follows:

$\mathcal{A}$ is the reference axis of x',
$\hat{v}$ is a unit vector at the origin, perpendicular to $\mathcal{A}$,
$r_{i,j}$ are angles with $0 \le r_{i,j} < 2\pi$,

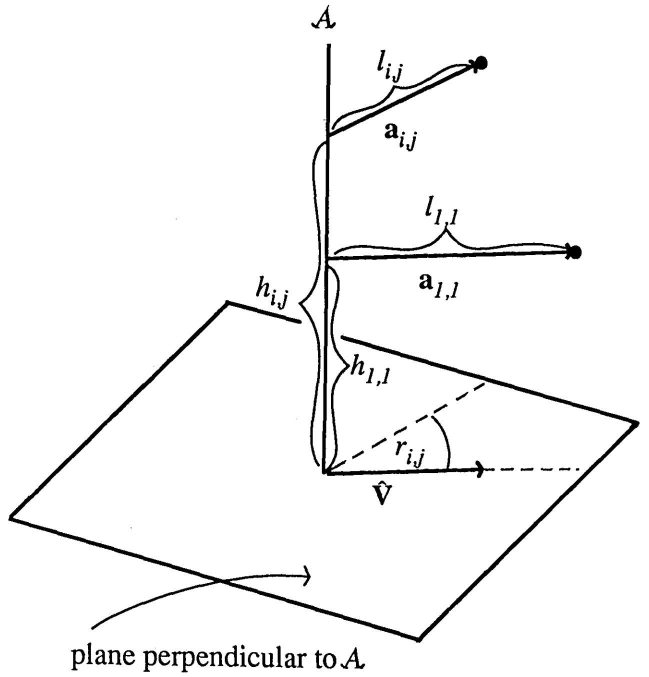

Fig. 9. Cylindrical coordinates for X. $\hat{v}$ is the projection of $\mathbf{a}_{1,1}$. The dashed line at an angle $r_{i,j}$ to $\hat{v}$ is the projection of $\mathbf{a}_{i,j}$.

$h_{i,j}$ are arbitrary real numbers,
$l_{i,j}$ are positive real numbers.

x' in Eq. (57) corresponds to the element $[\mathcal{A}, (\mathbf{a}_{i,j})]$ in X_0' defined as follows: let L denote the plane perpendicular to $\mathcal{A}$. Then $\hat{v}$ is the unit vector in L in the direction of the projection in L of $\mathbf{a}_{1,1}$. $r_{i,j}$ is then the angle between $\hat{v}$ and the projection in L of $\mathbf{a}_{i,j}$. $h_{i,j}$ and $l_{i,j}$ are as illustrated in Fig. 9; in particular, $l_{i,j}$ is the perpendicular distance from $\mathbf{a}_{i,j}$ to $\mathcal{A}$. This is illustrated in Fig. 9. Notice that $\hat{v}$ is well defined, since $\mathbf{a}_{1,1} \notin \mathcal{A}$. Similarly, the $r_{i,j}$ are well defined, since $\mathbf{a}_{i,j} \notin \mathcal{A}$. The augmented instantaneous rotations may be described within X' in a natural way as the solution set of equations that are linear in these coordinates, as described in the proposition below.

Proposition 6
E_0' is the subset of X_0' consisting of those elements x' whose representation in the form of Eq. (57) has the following properties:

(1) $r_{1,2} = r_{2,2} - r_{2,1} = \ldots = r_{n,2} - r_{n,1}$.
(2) $h_{i,1} = h_{i,2}$ for each $i = 1, \ldots, n$.
(3) $l_{i,1} = l_{i,2}$ for each $i = 1, \ldots, n$.

To see this, let e' denote an element of X_0' for which these conditions are satisfied. Let us denote by θ the common value of $r_{2,2} - r_{2,1}, \ldots, r_{n,2} - r_{n,1}$. For each $i = 1, \ldots, n$ denote by h_i and l_i the common values of $h_{i,1} = h_{i,2}$ and $l_{i,1} = l_{i,2}$. If we also drop the second subscripts on the $r_{2,1}, \ldots, r_{n,1}$, then we can write e' in the form

$$e' = (\mathcal{A}, \hat{v}, r_2, \ldots, r_n, \theta, h_1, \ldots, h_n, l_1, \ldots, l_n). \tag{58}$$

As before, let $\mathbf{a}_{i,1}$ be the vector of $\mathbf{R}^3$ whose cylindrical coordinates relative to $\mathcal{A}$ are (r_i, h_i, l_i), where the angle r_1 is measured with respect to $\hat{v}$ (so that $r_1 = 0$). Let σ denote the rotation about the axis $\mathcal{A}$ through the angle θ; then

$$e' = (\mathcal{A}, \mathbf{a}_{1,1}, \ldots, \mathbf{a}_{n,1}; \mathbf{a}_{1,2}, \ldots, \mathbf{a}_{n,2}) \in E'. \tag{59}$$

We now introduce groups G and J for which X_0' and E_0', respectively, are principal homogeneous (although for our application we use only that E_0' is principal homogeneous for J):

$$G = SO(3, \mathbf{R}) \times (\mathbf{S}^1)^{n-1} \times (\mathbf{S}^1)^n \times \mathbf{R}^n \times \mathbf{R}^n \times (\mathbf{R}^*)^n \times (\mathbf{R}^*)^n,$$
$$J = SO(3, \mathbf{R}) \times (\mathbf{S}^1)^{n-1} \times \mathbf{S}^1 \times \mathbf{R}^n \times (\mathbf{R}^*)^n. \tag{60}$$

$\mathbf{S}^1$ is the circle group, namely, the additive group $\mathbf{R}/2\pi\mathbf{Z}$; $\mathbf{R}$ is the additive group of real numbers, and $\mathbf{R}^*$ is the multiplicative group of positive real numbers. We denote elements γ of G in the form

$$\gamma = \begin{pmatrix} & & \alpha_{2,1}, \ldots, \alpha_{n,1} & \zeta_{1,1}, \ldots, \zeta_{n,1} & \lambda_{1,1}, \ldots, \lambda_{n,1} \\ \beta & & & & \\ & \alpha_{1,2} & \alpha_{2,2}, \ldots, \alpha_{n,2} & \zeta_{1,2}, \ldots, \zeta_{n,2} & \lambda_{1,2}, \ldots, \lambda_{n,2} \end{pmatrix}, \tag{61}$$

with $\beta \in SO(3, \mathbf{R})$, the α terms in $\mathbf{S}^1$, the ζ terms in $\mathbf{R}$, and the λ terms in $\mathbf{R}^*$. We write elements j of J in the form

$$j = (\beta, \alpha_2, \ldots, \alpha_n, \delta, \zeta_1, \ldots, \zeta_n, \lambda_1, \ldots, \lambda_n). \tag{62}$$

We view J as a subgroup of G by identifying j in Eq. (62) with the element γ of G given by

$$\gamma = \begin{pmatrix} & & \alpha_2, \ldots, \alpha_n & \zeta_1, \ldots, \zeta_n & \lambda_1, \ldots, \lambda_n \\ \beta & & & & \\ & \delta & \alpha_2 + \delta, \ldots, \alpha_n + \delta & \zeta_1, \ldots, \zeta_n & \lambda_1, \ldots, \lambda_n \end{pmatrix}.$$

We now describe the action of G on X_0'. Let $x' \in X_0'$ be as in Eq. (57) and $\gamma \in G$ as in Eq. (61); then

$$\gamma x' = \begin{pmatrix} & \beta\hat{v} & (r_{2,1} + \alpha_{2,1}), \ldots, (r_{n,1} + \alpha_{n,1}) & (h_{1,1} + \zeta_{1,1}), \ldots, (h_{n,1} + \zeta_{n,1}) & \lambda_{1,1}l_{1,1}, \ldots, \lambda_{n,1}l_{n,1} \\ \beta\mathcal{A} & & & & \\ & r_{1,2} + \alpha_{1,2} & (r_{2,2} + \alpha_{2,2}), \ldots, (r_{n,2} + \alpha_{n,2}) & (h_{1,2} + \zeta_{1,2}), \ldots, (h_{n,2} + \zeta_{n,2}) & \lambda_{1,2}l_{1,2}, \ldots, \lambda_{n,2}l_{n,2} \end{pmatrix}. \tag{63}$$

Here $\beta\mathcal{A}$ and $\beta\hat{v}$ denote the axis and the vector in $\mathbf{R}^3$ that are the images of $\mathcal{A}$ and $\hat{v}$ under the rotation β. This induces an action of J on E_0' that may then be described as follows: If e' is as in Eq. (58) and j is as in Eq. (62), then

$$je' = [\beta\mathcal{A}, \beta\hat{v}, (r_2 + \alpha_2), \ldots, (r_n + \alpha_n),$$
$$\theta + \delta, (h_1 + \zeta_1), \ldots, (h_n + \zeta_n), \lambda_1 l_1, \ldots, \lambda_n l_n). \tag{64}$$

We can now see that, for any pair $(x', \bar{x}')$ of elements of X_0', there is a unique $\gamma \in G$ such that $\bar{x}' = \gamma x'$. Suppose that x' is as in Eq. (57) and $\bar{x}''$ has components $\bar{\mathcal{A}}, \bar{\hat{v}}, \bar{d}_{2,1}$, etc. For any two pairs $(\mathcal{A}, \hat{v})$ and $(\bar{\mathcal{A}}, \bar{\hat{v}})$, each consisting of an oriented axis and a unit vector orthogonal to it, there is a unique β such that $(\bar{\mathcal{A}}, \bar{\hat{v}}) = (\beta\mathcal{A}, \beta\hat{v})$. This gives us the coordinate β of the required $\gamma \in G$. From Eq. (63) and Fig. 1 it is clear that the remaining coordinates of γ are fixed by the requirements that

$$\alpha_{i,j} = \bar{d}_{i,j} - r_{i,j} \qquad (\text{mod } 2\pi),$$
$$\zeta_{i,j} = \bar{h}_{i,j} - h_{i,j},$$
$$\lambda_{i,j} = \bar{l}_{i,j}/l_{i,j}.$$

$\lambda_{i,j}$ is a well-defined element of $\mathbf{R}^*$, since neither $l_{i,j}$ nor $\bar{l}_{i,j}$ is zero. Therefore X_0' is a principal homogeneous space for G, and E_0' is a principal homogeneous space for J.

We can also use the principal homogeneous structure of E_0' to make dimension calculations for E' (and S). First, observe that $\dim(E' - E_0')$ is less than $\dim(E_0')$. This is because when we restrict some $\mathbf{a}_{i,j}$ to lie on the axis $\mathcal{A}$ of the configuration we are eliminating one parameter of variation in the configuration, namely, the distance $l_{i,j}$ of $\mathbf{a}_{i,j}$ from $\mathcal{A}$. Thus $\dim(E') = \dim(E_0')$. Now, from the principal homogeneous space structure, it follows that the dimension of E_0' is the same as the dimension of J, which is $3n + 3$, as can be seen from Eqs. (60) by counting the dimensions of the groups in the expression for J as a product. We conclude that

$$\dim(E') = 3n + 3. \tag{65}$$

We now use the following fact:

If $g: Z \to W$ is any differentiable map, and if for all $w \in W$ the dimension of $g^{-1}(w)$ is constant, say, d, then $\dim(Z) = \dim(W) + d$.

Let us apply this fact to the restriction of the map f to E', i.e., $f_{E'}: E' \to E$. We know that, for all $e \in E$, $f^{-1}(e)$ consists of two points, so that $d = \dim[f_{E'}^{-1}(e)] = 0$. It follows that $\dim(E') = \dim(E)$. Thus from Eq. (65) we get

$$\dim(E) = 3n + 3. \tag{66}$$

Let us now apply this fact again to the restriction of π to E, i.e., π_E: $E \to S$. We have shown in Sections 2 and 3 that $\pi^{-1}(s)$ has a dimension of 1 for all $s \in S$. Therefore $\dim(E) = \dim(S) + 1$; i.e.,

$$\dim(S) = 3n + 2. \tag{67}$$

Finally, we note that the dimension of $Y = (\mathbf{R}^2)^{2n}$ is $4n$, so that the codimension of S in Y [i.e., $\dim(Y) - \dim(S)$] is $n - 2$:

$$\mathrm{codim}_Y(S) = n - 2. \tag{68}$$

It follows that for $n \geq 3$ the dimension of S is strictly less than that of Y, so that S has a measure of zero in the sense of the Lebesgue measure on Y. If we take this Lebesgue measure to be a natural unbiased measure on Y, this implies that the natural measure of false targets is zero.

In particular, $\mathrm{codim}_Y(S)$ increases with n. An interesting question is whether this is related to changes in strength of percept, if any exists, as n increases. In other words, as the number of points in the premise display increases, does the special event of a distinguished premise become more striking, and, if so, is this related to the fact that this special event becomes rarer (in the sense of larger codimension) as n increases? We merely wish to raise this issue here; we will not comment further on it in this paper.

We summarize. We have defined all but the interpretation kernel of an augmented instantaneous-rotation observer. The observer is augmented in that the axis $\mathcal{A}$ is included explicitly as part of the configuration. By so augmenting the configurations we are essentially able to exhibit X' and E' as principal homogeneous spaces for the groups G and J ("essentially" signifies that we have deleted the degenerate subsets $X' - X_0'$ and $E' - E_0'$); this also has the effect of linearizing the equations for E' in X'. The introduction of $\mathcal{A}$ into the configuration is thus attractive for mathematical reasons, but it raises an interesting question about the perceptual salience of the axis of rotation. Does the human visual system recover the axis of rotation in the relevant displays of structure from motion? If so, how precisely? Casual inspection suggests that the axis is perceptually salient, but this question deserves careful experimental investigation.

ACKNOWLEDGMENTS

We thank M. Braunstein for useful discussions and F. Pollick and L. Shapiro for assistance in the psychophysical experiments that led to this theory. This study was supported by National Science Foundation grant IRI-8700924 and by U.S. Office of Naval Research contracts N00014-85-K-0529 and N00014-88-K-0354.

REFERENCES AND NOTES

1. M. Braunstein, D. Hoffman, L. Shapiro, G. Andersen, and B. Bennett, "Minimum points and views for the recovery of three-dimensional structure," J. Exp. Psychol. **13**, 335–343 (1987).
2. S. Ullman, *The Interpretation of Visual Motion* (MIT Press, Cambridge, Mass., 1979).
3. D. Hoffman and B. Bennett, "Inferring the relative three-dimensional positions of two moving points," J. Opt. Soc. Am. A **2**, 350–353 (1985).
4. D. Hoffman and B. Bennett, "The computation of structure from fixed-axis motion: rigid structures," Biol. Cybern. **54**, 71–83 (1986).
5. J. Webb and J. Aggarwal, "Structure from motion of rigid and jointed objects," Artif. Intell. **19**, 107–130 (1982).
6. D. Hoffman and B. Flinchbaugh, "The interpretation of biological motion," Biol. Cybern. **42**, 197–204 (1982).
7. 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).
8. B. Bennett and D. Hoffman, "The computation of structure from fixed-axis motion: nonrigid structures," Biol. Cybern. **51**, 293–300 (1985).
9. B. Green, "Figure coherence in the kinetic depth effect," J. Exp. Psychol. **62**, 272–282 (1961).
10. J. J. Gibson and E. J. Gibson, "Continuous perspective transformations and the perception of rigid motion," J. Exp. Psychol. **54**, 129–138 (1957).
11. C. Hay, "Optical motions and space perception—an extension of Gibson's analysis," Psychol. Rev. **73**, 550–565 (1966).
12. H. Wallach and D. O'Connell, "The kinetic depth effect," J. Exp. Psychol. **45**, 205–217 (1953).
13. S. Ullman, "The interpretation of structure from motion," Proc. R. Soc. London Ser. B **203**, 405–426 (1979).
14. W. E. L. Grimson, "A computer implementation of a theory of human stereo vision," Philos. Trans. R. Soc. London Ser. B **292**, 217–253 (1980).
15. J. Koenderink and A. van Doorn, "Geometry of binocular vision and a model for stereopsis," Biol. Cybern. **21**, 29–35 (1976).
16. H. C. Longuet-Higgins, "The role of the vertical dimension in stereoscopic vision," Perception **11**, 377–386 (1982).
17. D. Marr and T. Poggio, "A computational theory of human stereo vision," Proc. R. Soc. London Ser. B **204**, 301–328 (1979).
18. J. Mayhew, "The interpretation of stereo-disparity information: the computation of surface orientation and depth," Perception **11**, 387–403 (1982).
19. W. Richards, "Structure from stereo and motion," Artifical Intelligence Lab. Memo. 731 (Massachusetts Institute of Technology, Cambridge, Mass., 1983).
20. While the present paper was under review, we were informed that several researchers have noted independently that, in effect, infinitely many rigid interpretations are compatible with any given point of S. Among the papers are those of T. Huang and C. Lee, "Motion and structure from orthographic projections," IEEE Trans. Pattern Anal. Machine Intell. (to be published) and S. Ullman, "Recent computational studies in the interpretation of structure from motion," in *Human and Machine Vision*, J. Beck, B. Hope, and A. Rosenfeld, eds. (Academic, New York, 1986), pp. 459–480. Ullman's result is for continuous time.
21. M. Braunstein, Department of Cognitive Science, University of California, Irvine, Irvine, California 92717 (personal communication, 1988).
22. B. Bennett, D. Hoffman, and C. Prakash, "Perception and computation," in *Proceedings of the IEEE First International Conference on Computer Vision*, M. Brady, ed. (Institute of Electrical and Electronics Engineers, New York, 1987), pp. 356–364.
23. B. Bennett, D. Hoffman, and C. Prakash, "New mathematical foundations for the field of perception," submitted to Behav. Brain Sci.

24. B. Bennett, D. Hoffman, and C. Prakash, *Observer Mechanics* (Academic, New York, to be published).

25. A measurable space X is a set, together with a specified collection of subsets of X, denoted $\sigma(X)$, that is closed under complementation and countable union and intersection (a collection of sets with these properties is called a σ algebra). The sets in $\sigma(X)$ are called the measurable subsets of X. Intuitively the measurable subsets correspond to those events to which probabilities may, in principle, be assigned. However, no particular such assignment need be specified as part of the definition of measurability itself.

26. To say that π is measurable means that, for each measurable set A of Y, its inverse image $\pi^{-1}(A)$ is a measurable set of X. To say that π is surjective means that, for every $y \in Y$, there is at least one $x \in X$ such that $\pi(x) = y$.