Biol. Cybern. 42, 195–204 (1982)

Biological Cybernetics
© Springer-Verlag 1982

The Interpretation of Biological Motion

D. D. Hoffman B E. Flinchbaugh

Artificial Intelligence Laboratory and Department of Psychology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Department of Computer and Information Science, The Ohio State University, Columbus, Ohio, USA

Abstract. The term *biological motion* has been coined by Johansson (1973) to refer to the ambulatory patterns of terrestrial bipeds and quadripeds. In this paper a computational theory of the visual perception of biological motion is proposed. The specific problem addressed is how the three dimensional structure and motions of animal limbs may be computed from the two dimensional motions of their projected images. It is noted that the limbs of animals typically do not move arbitrarily during ambulation. Rather, for anatomical reasons, they typically move in single planes for extended periods of time. This simple anatomical constraint is exploited as the basis for utilizing a "planarity assumption" in the interpretation of biological motion. The analysis proposed is: (1) divide the image into groups of two or three elements each; (2) test each group for pairwise-rigid *planar* motion; (3) combine the results from (2). Fundamental to the analysis are two "structure from planar motion" propositions. The first states that the structure and motion of two points rigidly linked and rotating in a plane is recoverable from three orthographic projections. The second states that the structure and motion of three points forming two hinged rods constrained to move in a plane is recoverable from two orthographic projections. The psychological relevance of the analysis and possible interactions with top down recognition processes are discussed.

1. Introduction

The ambulatory patterns of terrestrial and quadripeds have long born a unique significance for man among the variety of motions extant in his visual world. One's chances of survival in the neighborhood of a potential predator are presumably increased if one can distinguish an aimless meandering from a stealthful stalking or an outright run. More in line with our daily experience, we can quickly infer from the pendulum like motions of the limbs of a human whether he is walking, running, or performing some other motion. We can detect small deviations in gait patterns such as limps. Familiar individuals can often be recognized by the idiosyncracies of their gait.

The term *biological motion* has been coined by Johansson (1973) to refer to this subset of visual motions. In this paper a computational theory[1] for the perception of biological motion is proposed.

In developing this computational theory of biological motion we will also attempt to illustrate a research strategy that has been developed by investigators interested in providing computational descriptions of various aspects of human vision. The strategy may be schematized simply using six steps.

First, human visual information processing is artificially parcelled into provisional independent modules for research tractability[2]. Next a "minimal information display", some highly impoverished visual display which clearly demonstrates a modular human visual ability, is devised. Third, once a minimal information display is found, the information available in the display is accurately and concisely described. Then the nature of the representations built by the visual system in consequence of being presented with the display is

[1] The term *computational theory* is used in the sense proposed by Marr and Poggio (1977). Marr and Poggio observe that to thoroughly understand a complex information processing system involves obtaining descriptions of the system on three relatively independent levels. The top level, the level of the computational theory, describes what is being computed and for what purpose. The second level, that of the algorithm, specifies the nature of the particular algorithm used by the system in implementing the computational theory. The final level involves a description of the choice of hardware used in the system (e.g. neurons versus digital components)

[2] Of course these modules are but interim constructs to be later richly interconnected in an ideally completed computational model of human vision

0340-1200/82/0042/0195/$02.00

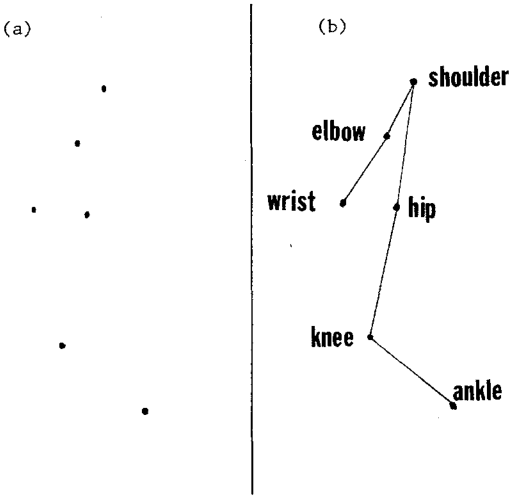

Fig. 1a and b. A single frame of a typical biological motion movie showing a sideview of a person walking. In **a** the dots are shown and in **b** their proper connections illustrated

specified precisely[3]. Fifth, since the information available in the display generally is insufficient in principle to arrive at a unique representation of the type presumably built by the visual system, plausible domain specific constraints about the nature of the world are sought which will allow the construction of such a unique representation. Finally an argument or constructive proof is devised to show that it is in principle possible to build a unique representation of the type desired given the information available in the display along with the a priori constraints about the world.

Once the steps of the computational analysis are completed, specific algorithms are considered for detailed implementations of the computational theory. The implementations provide existence proofs that the theory is internally consistent and also provide running models which can be tested for their psychological reality.

Fortunately the first two steps in building a computational theory for the perception of biological motion have already been completed by Johansson (1973, 1975). He first suggested that the perception of biological motion may be an isolable submodule of visual perception, a module able to build rich descriptions of the structure and motions of animals with recourse only to the projected motions of a limited number of feature points. More specifically, he suggested that the perception of biological motion does not require any visual information about the form of the animal (i.e., the outline due to its occluding contour), its texture, or color.

2. A Minimal Information Display For Biological Motion

Johansson devised a minimal information display to demonstrate that indeed the visual system can utilize motion information, with no further cues, to infer the correct structure and motion of an animal and often even to recognize which animal is being observed. The display is constructed as follows. Small light bulbs are attached to a subject's body at each of its joints (e.g., ankle, knee, hip, shoulder, elbow, wrist, etc.). The subject is then placed in a dark room and filmed while performing various activities. Single frames of the resulting film look to naive observer's as merely pictures of a few randomly placed dots. But when the film is shown at normal speeds naive observers almost immediately (within 100–1000 ms) see the dots as a person walking, running etc. (see Fig. 1). In fact, the perception is so powerful that it is impossible to force oneself to interpret the dots in any other manner.

The imports of this demonstration are two-fold. The first is psychological. *Humans* have the perceptual ability to utilize the two dimensional motions of feature points to build accurate descriptions of the underlying multi-limbed object. Thus it is of interest to perceptual psychologists how humans perform this conveniently circumscribed task. The second import is on a more general computational level. Since humans perform this perceptual task so reliably and quickly it must in principle be possible to perform. What we have here, in essence, is an existence proof of that fact. Therefore we can be confident that if we carefully characterize the informational input and the perceptual representations which are built in consequence of that informational input, there exists a computational procedure that maps from the former to the latter. Just such a characterization will be attempted next.

3. Characterizing the Information Available

Before a computational solution to the problem of biological motion is possible, one must make explicit the actual information available to the visual system (often called the "proximal stimulus") and the form of the ultimately desired representation. The desired representation will also be called the target representation. The actual information available to the visual system may be called the source representation. In this

[3] The nature of the ultimately desired representations is often inferred by noting what we *see* when shown the display. The desired representation may also be inferred in part by considerations of what in principle should be computed to reach certain goals. Marr and Nishihara (1978), for example, suggest what they call the *3-D model* representation based on considerations of what would be an optimal representation for the purpose of object recognition

section we describe the source representation and in the next section the form of the target representation. The problem will be to find a mapping from the former to the latter.

There appear to be at least four possible characterizations of the source representation (see Fig. 2). These four characterizations arise from decisions about the appropriate models for (a) the nature of the *projection* from the world onto the image plane and (b) the nature of the representation of the *motion* information. Although only one of the four characterizations will be used here, all four merit computational investigation.

The available information will here be characterized as a series of temporally successive orthographic snapshots[4]. In each snapshot what is explicitly represented is the two dimensional coordinates of the projections of the limb joints, such as the ankle, knee, and hip. Motion information is obtained by observing how the coordinates change from frame to frame. The actual coordinate system (e.g., Cartesian, polar coordinates, etc.) used to represent the two dimensional coordinates of the joints is not a concern at this point and will be left undetermined.

4. Defining the Target Representation

Two major considerations are involved when trying to specify a plausible target representation for the interpretation of biological motion. First, what do people perceive when presented with the minimal information display? Second, what information should be made explicit in the representation to facilitate attaining plausible goals of the observer?

The argument from perception is simple. When shown a biological motion display one perceives the three dimensional structure and motion of the limbs. Presumably then one must represent the three dimensional structure and motion of the limbs. This suggests that three dimensional primitives are appropriate for the target representation.

The computational argument is more involved. One plausible utilization of the target representation, though certainly not the only, is in shape recognition. Marr and Nishihara (1978) examine the problem of designing a representation that is in some sense optimal for recognizing shapes. Based on representational design issues and on several criteria for judging the usefulness of a representation for shape recognition they suggest a three dimensional representation based on a shape's natural axes which they call a 3-D model. Marr and Vaina (1980) extend these arguments to the case of recognizing moving shapes.

		PROJECTION	
		Orthographic	Perspective
MOTION	Discrete	discrete orthographic	discrete perspective
	Continuous	continuous orthographic	continuous perspective

Fig. 2. Possible characterizations of the input information to the biological motion module. The choice taken here is orthographic (parallel) projection with discrete motion. The other three characterizations are also viable candidates which should be considered

Based on the argument from perception and on the considerations raised by Marr and Nishihara we suggest that a plausible target representation is a three dimensional description of structure and motion akin to what Marr and Nishihara call a 3-D model. Specifically, what is to be computed is the length (in three dimensions) of each limb segment, the joint angle (in three dimensions) between each limb segment and both its successor and predecessor, and how these angles change over time.

The computational problem may now be precisely formulated. We would like to find a mapping from a finite number of two dimensional orthographic projections of the endpoints of the limbs of a moving animal to a three dimensional representation of the structure and motion of the animal which Marr and Nishihara have called a 3-D model.

5. The Planarity Assumption

Unfortunately there is no unique mapping from a series of frames of a biological motion movie to a 3-D model. The set of candidate three dimensional representations which may consistently be paired with the two dimensional source data is infinite. What we have is a fundamental ambiguity of interpretation.

To make the nature of this ambifuity clear we introduce the notion of a *pairwise-rigid structure* (see Fig. 3). A pairwise-rigid structure is a set of points moving in space so that each point remains at a constant distance from at least one other point, and no three points are in a rigid configuration. Intuitively a pairwise-rigid structure is a set of rigid rods joined end to end in ball and socket joints with no three rods forming a triangle. Consequently arms and legs qualify as pairwise-rigid structures.

[4] We assume that the correct correspondence of points in the successive snapshots has already been assigned. This problem is discussed in detail by Ullman (1979) and Marr (1981)

It can be shown that an infinite number of different pairwise-rigid structures can give rise to the same sequence of two dimensional projections regardless of the size of the sequence (Flinchbaugh, 1980). The ambiguity derives from the fact that there are an infinite number of rigid interpretations consistent with the motions of two distinct points in an image sequence. Knowledge of the exact position and motion of one of the points does not resolve the ambiguity. Thus even if an interpretation is chosen for one pair of points in a pairwise-rigid structure, an infinity of alternatives still remains for every other pair of points in the structure.

To overcome this ambiguity we need to incorporate plausible constraints about the nature of the world into our interpretation scheme. More specifically, what we would like is a plausible constraint on the *motions* of the limbs of animals because, as we have seen, unless the motion of a pairwise-rigid structure is constrained it cannot be given a unique interpretation.

One candidate motion constraint is the rigidity constraint (Ullman, 1979). Ullman proves that "Given three distinct orthographic views of four non-coplanar points in a rigid configuration, the structure and motion compatible with the three views are uniquely determined." He then proposes an interpretation scheme based on a rigidity assumption which states:

Any set of elements undergoing a 2-D transformation which has a unique interpretation as a rigid body moving in space, should be interpreted as such a body in motion.

The rigidity constraint is sufficient to give a unique interpretation if the object observed is moving rigidly. However the objects of interest here, namely animal limbs, violate the requirement of having four rigidly moving non-coplanar points. All rigidly connected points on a limb are not only coplanar, they are collinear. If a unique interpretation for biological motion is to be obtained, a constraint other than rigidity is required.

We propose to exploit an *anatomical constraint* on the motions of most bipeds and quadripeds as the basis of an interpretation scheme for biological motion. Casual observation reveals that in general the limbs of an ambulating animal do not move about arbitrarily. Rather, for anatomical reasons, each limb tends to move approximately in a single plane for extended periods of time. That is, joints tend to allow rotation more or less about a line. As will be discussed in the next section, this anatomical constraint is sufficient to provide a unique interpretation for biological motion.

Motivated by the observation of this anatomical constraint, the principle we propose for the interpretation of biological motion is what we shall call the *planarity assumption*:[5]

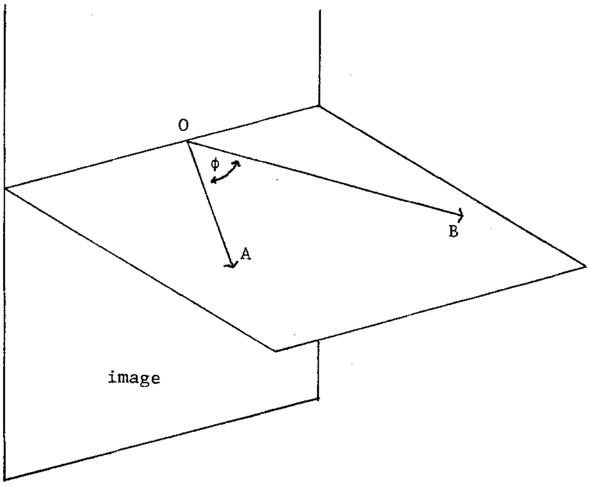

Fig. 3. This figure illustrates a pairwise-rigid structure constrained to move in one plane. This pairwise-rigid structure is composed of two rigid rods (though in general a pairwise-rigid structure can have anywhere from one to an infinity of rigid rods) with endpoints at A and B and a common endpoint at the joint O. The only motion allowed is a change in the angle ϕ, translation in the plane spanned by OA and OB, and rotation in that plane. (In general pairwise-rigid structures are not subject to these motion constraints)

Any set of elements undergoing a 2-D transformation which has a unique interpretation as a pairwise-rigid structure moving in one plane, should be interpreted as such a body in motion.

6. Interpreting Visual Motion Utilizing the Planarity Assumption

The planarity assumption is employed in interpreting visual motion by looking at groups of two or three points and checking if they have a unique interpretation as a pairwise-rigid structure constrained to move in a plane. If not, no interpretation is assigned. If so, the planar interpretation is provisionally accepted as correct.

As is the case with Ullman's rigidity assumption for the recovery of three dimensional structure and motion, the planarity assumption must be shown to be immune to "false targets" and "phantom structures". A false target occurs when a collection of points that does not constitute a pairwise-rigid structure in planar motion gives rise to a series of orthographic pro-

[5] Although the most obvious application of the planarity assumption is in the interpretation of biological motion, we do not intend to imply that utilization of the assumption is restricted to the interpretation of biological motion. Rather we suggest it is a general principle for interpreting visual motion that, like the rigidity principle, is used by the visual system whenever appropriate

jections which are consistent with the planar interpretation. A phantom structure occurs when a collection of points that does constitute a pairwise-rigid structure in planar motion gives rise to a series of orthographic projections which are consistent with more than one planar interpretation. Ullman's proof (1977, Appendix 1) that false targets occur only with probability zero also holds for the planar case. The proof that there can be no phantom structures follows from the following "structure from planar motion" propositions.

The Structure from Planar Motion Propositions

Proposition 1. *Given three distinct orthographic projections of the two endpoints of a rigid rod which is constrained to rotate in a plane, the structure and motion compatible with the three views are uniquely determined*[6].

Proposition 2. *Given two distinct orthographic projections of the three endpoints of two rigid rods linked in a hinge joint to form a pairwise-rigid structure which is constrained to move in one plane, the structure and motion compatible with the two views are uniquely determined.*

The proofs for these propositions are outlined in appendices one and two respectively. The proofs are constructive and thus provide algorithms for the computation of the structure and motion.

The Interpretation Scheme

The interpretation scheme based on these propositions is as follows. (1) Divide the image into groups of two or three elements each. The appropriate elements for the interpretation of biological motion seem to be the joints of the limbs of an animal, such as the ankle, knee, and hip. (2) Test each group for pairwise-rigid planar motion. For groups of two elements Proposition 1 may be applied. For groups of three elements Proposition 2 may be applied. (3) Combine the results from (2).

Some Potential Objections to the Scheme

Some potential objections to this scheme should be considered. First, it appears that the most this scheme can deliver is the three dimensional structure and motion of the limbs of an animal. The trunk typically violates both the rigidity assumption and the planarity assumption. This may or may not be a serious objection. Two avenues are worth exploring on this problem. First, perhaps further natural constraints in the spirit of the rigidity and planarity assumptions can be found to aid in the bottom up interpretation of trunk structure. In general, bottom up avenues of interpretation should be exhausted before recourse to top down schemes is taken. With this consideration in mind a second interesting possibility exists. Perhaps the limb structure and motion obtained bottom up using the planarity assumption is sufficient to provide a unique index into a stored table of 3-D models of animals. The interpretation of biological motion would then involve an interaction of both bottom up and top down processes. The bottom up processes get the interpretation process off the ground and the top down processes complete the interpretation of those structures which resist bottom up attack.

Another objection to this scheme might be raised. The planarity assumption may work quite nicely when the object observed is performing some repetitive activity such as running, walking, or jogging. But how about more complicated activities? Johansson, for example, has minimal information displays of a dancing couple which we seem able to interpret, though with a bit more difficulty. A good portion of the time the couple is badly violating the planarity assumption when, for example, they spin or turn.

This objection brings up several interesting points. First it should be noted that the planar interpretation scheme does not provide spurious interpretations when the planarity assumption is violated. The scheme can determine when the assumption is valid and when it is not[7]. When it is invalid, no interpretation is made.

To make the second point we divide the dancing sequence into three categories depending upon which assumptions the couple's movements obey. During part of the sequence, for example when the partners step toward or away from each other, their movements conform to the planarity assumption. During these movements the planarity scheme can uniquely determine the three dimensional structure and motions of the dancer's limbs. At other times the dancers spin with many of their limbs held in one position during the spin. Under these conditions the rigidity assumption holds and three dimensional structure may be computed. But there are definitely periods when the motion clearly violates both the rigidity and planarity assumptions. During these periods the bottom up processes proposed so far will simply not be able to give an interpretation. What could happen percep-

[6] Since orthographic projection is used the structure and motion are uniquely determined up to a reflection about the image plane

[7] This is a point that may have escaped some researchers who have objected to the use of elaborate assumptions to aid in the interpretation of the visual world. Generally, schemes based on elaborate assumptions are able to check in a bottom up manner whether or not their assumptions are valid. A second point is worth mentioning. The world is structured. Why shouldn't the visual system exploit that structure in interpreting the visual world?

tually during these periods? There are two possibilities. First, the visual system could utilize the structural information obtained during periods obeying the planarity or rigidity assumptions to interpret the motion during the periods of violation. It can be shown that if the three dimensional structure of a pairwise-rigid object is known, then its motion can be inferred uniquely even when the planarity constraint is violated. The second possibility is simply that no interpretation is made during these periods of violation. From observing these dancing displays it appears that the latter possibility is what often happens. At the moment a dancer starts a spin, we momentarily lose the structure and motion only to regain it later during periods of planar motion.

Linking the Feature Points

One advantage accrues to the planarity scheme somewhat as a side effect. A persistent problem for investigators of biological motion has been to get the correct two dimensional linking of the points. For example, how can we go about linking the ankle point to the knee and the knee to the hip without also introducing an incorrect link between the ankle and hip? Simple solutions like nearest neighbor connections simply do not work. Rashid (1979) sets out specifically to compute the correct two dimensional links based on graph-theoretic cluster analysis of the two dimensional positions and velocities of the points. Webb (1980) starts his analysis of structure from biological motion by assuming that the correct two dimensional links are already known. The planarity scheme does not make the computation of the correct two dimensional linking of the points a specific goal. Instead the three dimensional structure and motion are computed and the two dimensional linkage then falls out incidentally.

A simple example may help to see this. Suppose we have several views of just three feature points: the ankle, knee, and hip. We would like to determine if there is a unique interpretation of these points as a pairwise-rigid structure in planar motion using the second proposition that two views of three points is sufficient for our purpose. First we submit the three points to an implementation of Proposition 2 with the ankle tagged as the provisional pivot point. The routine returns with no interpretation. Next we tag the hip as the provisional joint of a pairwise-rigid structure in planar motion. Again no interpretation is returned. Finally we ask if there is an interpretation with the knee as the pivot point. The routine returns the three dimensional distance between the knee and hip, between the knee and ankle, the motion of these limbs, and the plane of the motion. Consequently we know that this is the correct interpretation, and we know the correct three dimensional structure and motion. But note that we also know, as a side effect, there is no link between the ankle and hip feature points.

A Psychophysical Prediction

Some previous algorithms require that each limb be seen at least once in its full extension so that its projected length is the same as its length in three dimensions. The planarity scheme clearly predicts that it is not necessary to see *any* of the limbs in maximal extension to infer the correct structure and motion. The critical psychophysical experiment on this issue is trivial. One simply views a biological motion display where none of the limbs reaches maximal extension. When this is done the perception of the biological motion is not at all reduced.

Summary

The visual interpretation of biological motion has been investigated using a computational approach. Anatomical constraints on how the limbs of animals typically move during ambulation were exploited as the basis for an interpretation scheme based on an assumption of planar motion. Two "structure from planar motion" propositions were proved, providing explicit computational methods for implementing the planarity scheme[8].

Appendix 1. The Structure from Planar Motion Proposition for Two Points

Proposition. *Given three distinct orthographic projections of the two endpoints of a rigid rod which is constrained to move in a plane, the structure and motion compatible with the three views are uniquely determined (up to a reflection about the image plane).*

Proof. Let O, A_1, A_2, and A_3 be the endpoints of the rigid rod in frames one through three respectively (see Fig. 4). Let $\mathbf{a}_i$ be the vector from O to A_i in frame i. Let the coordinates of $\mathbf{a}_i$ be (x_i, y_i, z_i). Under orthographic projection the x and y coordinates of each vector are unaltered and the z coordinates are lost completely. Thus the problem consists of recovering the three unknown z coordinates. We first show that there are at most four solutions (i.e., two solutions plus their reflections) for the z coordinates given three views, and then show that there is a unique solution.

[8] An implementation of the planarity scheme in a simple local network is developed in Hoffman (1981)

Note that in Fig. 4 the reference point O does not translate over the three views. This does not imply a loss of generality. Two types of translation are possible. The first, translation in depth, is in principle unrecoverable under orthographic projection. The second, translation parallel to the image plane, yields projected translations identical to the translation of the object in the world. Since these translations are trivially recovered, they are ignored in this analysis.

From the fact that the length (in three dimensions, not in the image) of $\mathbf{a}$ is invariant over the three views we obtain the two equations[9]

$$\|\mathbf{a}_1\| = \|\mathbf{a}_2\|\,, \tag{1}$$

$$\|\mathbf{a}_1\| = \|\mathbf{a}_3\|\,. \tag{2}$$

Three vectors lie in a plane if and only if their triple scalar product is zero. From the planarity constraint we obtain the equation[10]

$$[\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3] = 0\,. \tag{3}$$

Equations (1)–(3) may be expanded into polynomial equations in terms of their z coordinates giving:

$$z_1^2 - z_2^2 + k_1 = 0\,, \tag{4}$$

$$\mathrm{z}_1^2 - \mathrm{z}_3^2 + \mathrm{k}_2 = 0\,, \tag{5}$$

$$k_3 z_1 + k_4 z_2 + k_5 z_3 = 0\,. \tag{6}$$

The k's in these equations are expressions entirely in the x and y coordinates of the position vectors[11]. Since these quantities are available directly from the orthographic projections they are lumped together into constants. The goal here is to solve these three equations for the three z coordinates.

The solution space for the three z coordinates in three views can be visualized as the mutual intersection points of two hyperboloid sheets and one plane passing through the origin. This is illustrated in Fig. 5.

The simple fact that we have three equations and three unknown here does not mean that this system has a finite number of solutions. To ascertain if there are a finite number of solutions we apply the inverse function theorem. This theorem allows us to conclude that wherever the Jacobian of these equations is nonsingular the mapping defined by the equations is locally one to one and onto (i.e., a local diffeomorphism). Consequently any roots at points where the Jacobian is nonsingular are isolated and not part of a continuum of solutions.

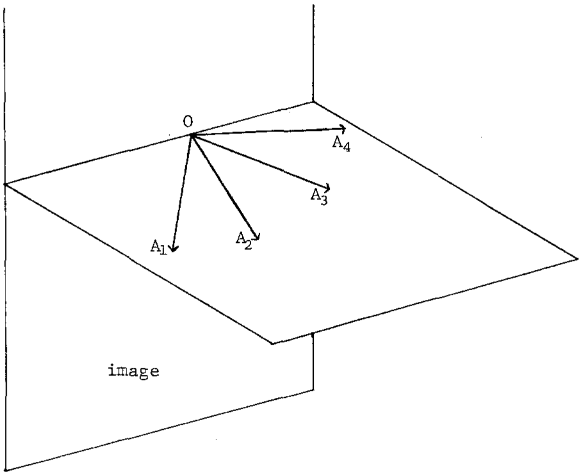

Fig. 4. Geometry underlying the proof of Theorem 1

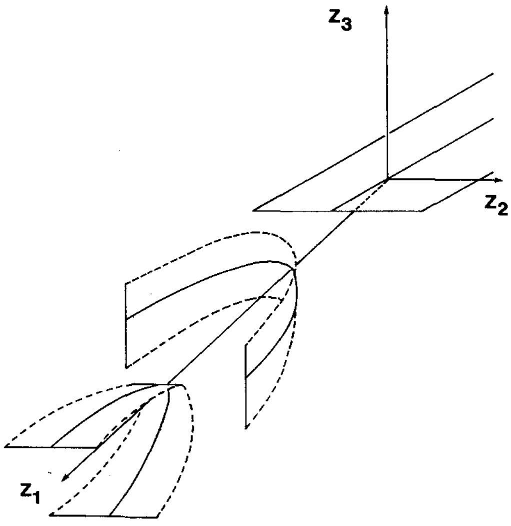

Fig. 5. The solution space for the coordinates z_1, z_2, and z_3. The solution can be seen here to be the mutual intersection of two hyperboloid sheets rotated ninety degrees with respect to each other and a plane passing through the origin. The asymptotic lines of the hyperboloid sheets are always at forty five degrees with respect to the z_1 axis. (The limbs of the hyperboloids on the other side of the z_2z_3 plane are not shown)

The determinant of the Jacobian of (1)–(3) is:

$$\begin{vmatrix} 2z_1 & -2z_2 & 0 \\ 2z_1 & 0 & -2z_3 \\ k_3 & k_4 & k_5 \end{vmatrix}\,.$$

[9] The notation $\|\mathbf{a}_1\|$ is vector shorthand for the length of the vector $\mathbf{a}_1$. In terms of the components of $\mathbf{a}_1$ this length may be expressed $\sqrt{x_1^2+y_1^2+z_1^2}$

[10] The triple scalar product of three vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ is indicated by the shorthand $[\mathbf{a}_1\mathbf{a}_2\mathbf{a}_3]$. Taking the triple scalar product involves first taking the vector cross product of $\mathbf{a}_2$ and $\mathbf{a}_3$ and then taking the dot product of the resulting vector with $\mathbf{a}_1$. Intuitively the triple scalar product gives the volume of the parallelepiped formed by the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$

[11] The actual expressions for the k's are $k_1 = x_1^2 + y_1^2 - x_2^2 - y_2^2$, $k_2 = x_1^2 + y_1^2 - x_3^2 - y_3^2$, $k_3 = x_2y_3 - x_3y_2$, $k_4 = x_3y_1 - x_1y_3$, $k_5 = x_1y_2 - x_2y_1$

This Jacobian has rank three[12]. If (4)–(6) involved transcendental functions the most we could conclude from this Jacobian test would be that the set of solutions was of measure zero. However (4)–(6) are polynomials. Consequently we can assert that the system of equations has but a finite set of solutions in general. By Bezout's theorem[13] we know that the sum of the multiplicities of the solutions does not exceed the product of the degrees of the equations, which in this case is four. This can be seen geometrically from Fig. 5.

We have shown that there are at most four real solutions given three views of the two points. These four solutions come in two pairs, with the two members of a given pair being the reflections about the image plane of each other.

We now prove the solution is unique up to a reflection[14]. Solve (6) for z_1, substitute into (4) and (5) and simplify.

$$(k_4^2-k_3^2)z_2^2+k_5^2z_3^2+2k_4k_5z_2z_3+k_1k_3^2=0\,, \qquad (7)$$

$$k_4^2z_2^2+(k_5^2-k_3^2)z_3^2+2k_4k_5z_2z_3+k_2k_3^2=0\,. \qquad (8)$$

Multiply (7) by k_2 and (8) by k_1. Subtract (8) from (7). Divide the result by z_3^2 and let $x=z_2/z_3$.

$$\begin{aligned}&[k_2(k_4^2-k_3^2)-k_1k_4^2]x^2+2k_4k_5(k_2-k_1)x\\ &\quad+[k_2k_5^2-k_1(k_5^2-k_3^2)]=0\,. \end{aligned} \qquad (9)$$

Solve (9) for x.

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}\,. \qquad (10)$$

Where $a=k_2(k_4^2-k_3^2)-k_1k_4^2$, $b=2k_4k_5(k_2-k_1)$, and $c=k_2k_5^2-k_1(k_5^2-k_3^2)$.

Before continuing we establish one claim.

Claim. *Provided the plane of rotation of the rod is not parallel to the image plane and that none of the three projected images of the rod are collinear, at most one of the solutions for x is valid.*

Proof. Let the length of the rod be ϱ. Let the projected lengths of the rod in views 2 and 3 be, respectively, r_2 and r_3. Then $z_2=\pm\sqrt{\varrho^2-r_2^2}$ and $z_3=\pm\sqrt{\varrho^2-r_3^2}$. Consequently

$$x=\frac{z_2}{z_3}=\pm\frac{\sqrt{\varrho^2-r_2^2}}{\sqrt{\varrho^2-r_3^2}}\,. \qquad (11)$$

Thus if x has two solutions, then these two solutions must have the same absolute value and opposite sign if both are to be valid. From (10) we conclude that x will have two valid solutions only when

$$\frac{-b+\sqrt{b^2-4ac}}{2a}=-\left[\frac{-b-\sqrt{b^2-4ac}}{2a}\right] \qquad (12)$$

which is true only when $b=0$. From the equation for b in (10) we see this implies the following degenerate conditions: $k_4=0$, $k_5=0$, $k_2=k_1$. k_4 can be interpreted as the dot product of the projected image of the rod in view three with a vector orthogonal to the projected image of the rod in view one. k_5 can be interpreted as the dot product of the projected image of the rod in view one with a vector orthogonal to the projected image of the rod in view two. Thus k_4 or k_5 is zero only if the appropriate projected images of the rod are collinear. $k_1=k_2$ implies that $r_2=r_3$. This can happen if the plane of rotation of the rod is parallel to the image plane or if the projected images in view two and three are collinear. Thus, except for these degenerate conditions, x must have a unique valid solution. Q.E.D.

Substitute z_3x for z_2 in (7). This can be done since $x=z_2/z_3$. Note that x is now one of two known values.

$$(k_4^2-k_3^2)x^2z_3^2+k_5^2z_3^2+2k_4k_5xz_3^2+k_1k_3^2=0\,. \qquad (13)$$

The solution for z_3 is

$$z_3=\pm\sqrt{\frac{-k_1k_3^2}{(k_4^2-k_3^2)x^2+k_5^2+2k_4k_5x}}\,. \qquad (14)$$

The solutions for z_2 and z_1 follow immediately: $z_2=xz_3$, $z_1=\sqrt{z_3^2-k_2}$. By the claim we know that only one of the two values of x is valid, except in degenerate cases. Thus the solutions for z_1, z_2, z_3 are unique up to a reflection.

In practice we can find solutions for the z's using both values of x and reject the pair of solutions which is either imaginary or which violates the conditions established in the claim. Q.E.D.

Appendix 2. The Structure from Planar Motion Proposition for Three Points

Proposition. *Given two distinct orthographic projections of the three endpoints of two rigid rods linked to form a pairwise-rigid structure which is constrained to move in a*

[12] The actual expressions for the k's in terms of x's and y's must be used when determining the rank of the Jacobian. Otherwise hidden dependencies among the variables may escape notice. One can find all the degenerate cases (i.e., cases when the Jacobian drops rank and no unique solution is possible) by factoring the determinant of the Jacobian and setting the factors equal to zero

[13] For a nontechnical discussion of the inverse function theorem and Bezout's theorem see Richards et al. (1981)

[14] Horn (1981, personal communication) first proved uniqueness of the solution. He noted that the two points in planar motion trace out a circle in space. This circle maps into an ellipse with known center under orthographic projection. Three points on the ellipse determine its three parameters – the major and minor axes, and the angle of the major axis. He made a similar construction for the case of two views of three points

plane, the structure and motion compatible with the two views are uniquely determined (up to a reflection about the image plane).

Outline of Proof. Let O, A_i, B_i be the endpoints of the two rigid rods (which form a joint at O) in frame i where $i = 1, 2$ (see Fig. 6). Let $\mathbf{a}_i$ be the vector from O to A_i and $\mathbf{b}_i$ be the vector from O to B_i. Let the coordinates of $\mathbf{a}_i$ be (x_{ai}, y_{ai}, z_{ai}). Let the coordinates of $\mathbf{b}_i$ be (x_{bi}, y_{bi}, z_{bi}). Under orthographic projection the x and y coordinates of each vector remain unaltered and the z coordinates are lost completely. Thus the problem consists of recovering the four unknown coordinates z_{ai} and z_{bi}. We first show that there are but a finite number of solutions for the z coordinates given only two views, and then show that the solution is actually unique up to a reflection.

From the fact that the lengths (in three dimensions, not in the image) of $\mathbf{a}$ and $\mathbf{b}$ remain invariant over the two views we obtain the two equations[15]

$$\|\mathbf{a}_1\| = \|\mathbf{a}_2\|, \tag{1}$$

$$\|\mathbf{b}_1\| = \|\mathbf{b}_2\|. \tag{2}$$

Three vectors lie in a plane if and only if their triple scalar product is zero. From the planarity constraint we obtain the two equations:

$$[\mathbf{a}_1\mathbf{b}_1\mathbf{a}_2] = 0, \tag{3}$$

$$[\mathbf{a}_1\mathbf{b}_1\mathbf{b}_2] = 0, \tag{4}$$

Equations (1)–(4) may be expanded into polynomial equations in terms of their four z coordinates giving:

$$z_{a1}^2 - z_{a2}^2 + k_1 = 0, \tag{5}$$

$$z_{b1}^2 - z_{b2}^2 + k_2 = 0, \tag{6}$$

$$k_3 z_{a1} + k_4 z_{b1} + k_5 z_{a2} = 0, \tag{7}$$

$$k_6 z_{a1} + k_7 z_{b1} + k_8 z_{b2} = 0. \tag{8}$$

The k's in these equations are expressions entirely in the x and y coordinates of the position vectors. Since these quantities are available directly from the orthographic projections they are lumped together into constants. The goal here is to solve these four equations for the four z coordinates.

The simple fact that there are four equations and four unknowns does not imply that this system has a finite number of solutions. To ascertain if there are a finite number of solutions we apply the inverse function theorem. This theorem lets us conclude that wherever the Jacobian of these equations is nonsingular the mapping defined by the equations is locally one to one and onto (i.e., a local diffeomorphism). This means that any roots at points where the Jacobian is nonsingular are isolated and not part of a continuum of solutions.

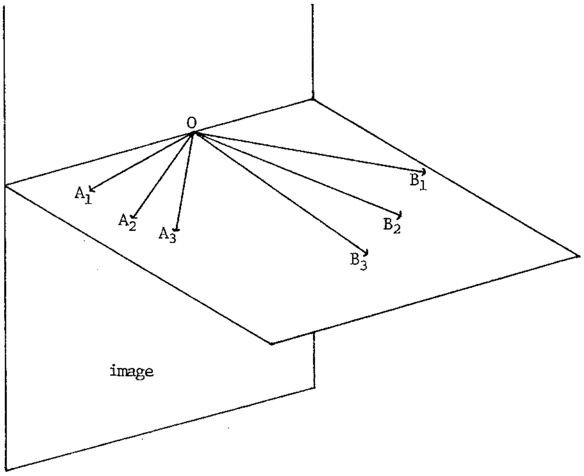

Fig. 6. Geometry underlying the proof of Theorem 2

The determinant of the Jacobian of these four equations is:

$$\begin{vmatrix} 2z_{a1} & -2z_{a2} & 0 & 0 \\ 0 & 0 & 2z_{b1} & -2z_{b2} \\ k_3 & k_5 & k_4 & 0 \\ k_6 & 0 & k_7 & k_8 \end{vmatrix}.$$

This Jacobian has rank four. If (5)–(8) involved transcendental functions the most we could conclude from this Jacobian test would be that the set of solutions was at most of measure zero. However (5)–(8) are polynomials. Consequently we can assert that the system of equations has but a finite set of solutions in general. By Bezout's theorem[16] we know that the sum of the multiplicities of the solutions does not exceed the product of the degrees of the equations, which in this case is four.

We have shown that there are at most four real solutions given two views of the three points. These four solutions come in two pairs, with the two members of a given pair being the reflections about the image plane of each other. The proof that the solution is unique is almost identical to that given in appendix one and will not be reiterated here.

Acknowledgements. We thank W. Richards, S. Ullman, B. Horn, and D. Marr for helpful discussions and ideas, and J. Rubin and A. Pentland for careful readings of earlier drafts. We also thank the entire vision group of the MIT Artificial Intelligence Laboratory for helpful feedback during in house seminars on this topic.

[15] See the footnotes in Appendix 1 for an explanation of the vector notation used in these equations

[16] For a nontechnical discussion of the inverse function theorem and Bezout's theorem see Richards et al. (1981)

This report describes research done at the Artificial Intelligence Laboratory and Psychology Department of the Massachusetts Institute of Technology. Support for this research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643 and by NSF and AFOSR under grant 7923110-MCS. D. D. Hoffman was also supported by a Hughes Aircraft Company graduate fellowship.

References

Flinchbaugh, B.E.: Visual interpretation of pairwise-rigid structure, forthcoming (1981)

Hoffman, D.: Visual motion perception as hypothesis testing, MIT AI Memo 630 (1981)

Johansson, G.: Visual perception of biological motion and a model for its analysis. Percept. Psychophys. **14**, 201–211 (1973)

Johansson, G.: Visual motion perception. Sci. Am. 76–88 (1975)

Marr, D.: Vision: a computational investigation into the human representation and processing of visual information. San Francisco: W.H. Freeman 1981

Marr, D., Nishihara, H.K.: Representation and recognition of the spatial organization of three dimensional shapes. MIT AI Memo 416 (1977)

Marr, D., Poggio, T.: From understanding computation to understanding neural circuitry. Neurosci. Res. Program Bull. **15**, 470–488 (1977)

Marr, D., Vaina, L.: Representation and recognition of the movement of shapes. MIT AI Memo 597 (1980)

Rashid, R.: Towards a system for the interpretation of moving light displays. University of Rochester Computer Science TR53 (1979)

Richards, W., Rubin, J., Hoffman, D.: Solving problems in natural computation. MIT AI Memo 614 (1981)

Ullman, S.: The interpretation of visual motion. MIT PhD Thesis (1977)

Ullman, S.: The interpretation of visual motion. Cambridge: MIT Press 1979

Webb, J.: Static analysis of moving jointed objects. Proc. AAAI, 35–37 (1980)

Received: May 29, 1981

Dr. D. D. Hoffman
Artificial Intelligence Laboratory
and Department of Psychology
Massachusetts Institute of Technology
E10-012
Cambridge, MA 02139
USA