

Inferring structure from motion: A homotopy algorithm

Bruce M. Bennett[†] Donald D. Hoffman[‡] Jill E. Nicola[‡] Chetan Prakash^{*}

[†]Mathematics, UC Irvine 92717 [‡]Cognitive Science, UC Irvine 92717

^{*}Mathematics, CSU San Bernardino 92407

ABSTRACT

Theoretical investigations of the inference of three-dimensional structure from image motion often result in systems of coupled nonlinear equations which must be solved to infer the third dimension. If closed form solutions cannot be obtained then various search procedures, such as simulated annealing, are often used. In this paper we discuss a relatively new approach to solving coupled nonlinear systems of equations, an approach based on the so-called "homotopy principle." We discuss this approach in the context of developing an algorithm for inferring structure from motion using an assumption of rigid fixed-axis motion. We also discuss this approach in the more general context of observer theory, a mathematical framework for the field of perception.

INTRODUCTION

Every act of perception is a process of inference; it is a process whereby conclusions are drawn from premises. In the typical case, the conclusions are not deductive consequences of the premises, so that the inference involves an informational leap on the part of the perceiving system.

An exceptionally clear example is a vision system that infers a 3-D structure for an object from a sequence of 2-D images of its movement. In such a case the sequence of 2-D images is the premise of the inference and the perceived 3-D structure is the conclusion. This conclusion is not a deductive consequence of the premise: no principle of logic requires any particular 3-D interpretation. Therefore principles in addition to those of logic must guide this inference.

Several principles have been proposed and studied. In this paper we focus on one, the principle of

rigid fixed-axis (RFA) motion [1-4]. An object undergoes RFA motion just in case it moves rigidly in three dimensions and rotates about a single fixed axis. The RFA principle states that if a sequence of 2-D images is compatible with RFA 3-D interpretations then these interpretations should be assigned a high degree of inductive strength, i.e., they should be regarded as plausible. The RFA principle need not be construed as excluding other principles; rather, it can be coordinated with other principles to create powerful inferential strategies. The RFA principle is useful if two properties obtain for the image sequences used as premises.

1. **Minimal false targets:** Almost all of the sequences of 2-D images that are used as premises are incompatible with any RFA interpretation.
2. **Minimal interpretations:** Of those sequences of 2-D images that are compatible with at least one RFA interpretation, almost all are compatible with but few RFA interpretations.

The first property is desirable because it reduces the probability of false conclusions (section five explains this further using observer theory). The second is desirable because it makes the conclusions more informative. These considerations have led researchers to seek precise conditions under which one or both of these properties obtain. We mention two results of this research, one using the rigidity principle and one using the RFA principle. We then consider the RFA theory in more detail, designing a homotopy algorithm that constructs the RFA interpretations specified by the theory.

Ullman [5] first proved conditions for which both properties hold in the case where one uses a principle of rigid motion (to be distinguished from the RFA

principle, which requires rigidity of motion *and* motion about a single fixed axis). (For more on rigidity and related principles see also [6–17].) He found that if one is given three orthographic views of four points moving arbitrarily in 3-D then, almost surely, there is no rigid interpretation compatible with the views. This addresses the first property. Then he found that if one is given three orthographic views of four non-coplanar points that are moving rigidly then, almost surely, there are two rigid interpretations compatible with the views. The two interpretations are mirror reflections of each other. This addresses the second property.

Hoffman and Bennett [1] gave conditions for which both properties hold in the case where one uses the RFA principle. They found that if one is given three orthographic views of three points moving arbitrarily in 3-D then, almost surely, there is no RFA interpretation compatible with the views. This addresses the first property. Then they found that if one is given three orthographic views of three points that undergo RFA motion then, almost surely, there are just two RFA interpretations compatible with the views. Again, the two interpretations are mirror reflection of each other. This addresses the second property.

The RFA result proved by Hoffman and Bennett requires that one solve a system of coupled nonlinear equations to arrive at the RFA interpretations compatible with the given image data. Solving coupled nonlinear equations is not always easy, and can be computationally intensive. The purpose of this paper is two-fold. First we discuss a class of algorithms for solving coupled nonlinear equations, using the RFA result of Hoffman and Bennett for a concrete example. To this end, section two develops the RFA result, culminating in the coupled system that must be solved. Section three introduces the class of algorithms, all exploiting the so-called “homotopy principle.” And section four applies the homotopy approach to the equations of section two to develop an algorithm for inferring RFA interpretations. The second purpose of this paper is to provide a concrete example in support of a mathematical framework for the field of perception, a framework called “observer theory.” To this end, section five briefly reviews the definition of an observer and exhibits the RFA algorithm as a particular example of this more general structure.

THE EQUATIONS

In this section we develop the equations that will be solved, via a homotopy algorithm, to construct the RFA interpretations for three orthographic views of three points.

As in reference 1, let the three points be denoted O , A_1 , and A_2 . Without loss of generality, we take O to be the origin of a cartesian coordinate system. Let \mathbf{a}_{ij} denote the three-dimensional vector between O and A_i in view j ($j = 1, 2, 3$). Let the cartesian coordinates of \mathbf{a}_{ij} with respect to O be denoted by (x_{ij}, y_{ij}, z_{ij}) and assume that the viewer’s line of sight is directed along the positive z -axis. If the motion from view to view is rigid, we expect that the lengths of the vectors \mathbf{a}_{11} , \mathbf{a}_{12} , \mathbf{a}_{13} should be identical, as should be the lengths of the vectors \mathbf{a}_{21} , \mathbf{a}_{22} , \mathbf{a}_{23} . We can therefore write

$$\mathbf{a}_{11} \cdot \mathbf{a}_{11} = \mathbf{a}_{12} \cdot \mathbf{a}_{12}, \quad (1)$$

$$\mathbf{a}_{11} \cdot \mathbf{a}_{11} = \mathbf{a}_{13} \cdot \mathbf{a}_{13}, \quad (2)$$

$$\mathbf{a}_{21} \cdot \mathbf{a}_{21} = \mathbf{a}_{22} \cdot \mathbf{a}_{22}, \quad (3)$$

$$\mathbf{a}_{21} \cdot \mathbf{a}_{21} = \mathbf{a}_{23} \cdot \mathbf{a}_{23}. \quad (4)$$

Moreover, we expect that the angle between the vectors OA_1 and OA_2 should remain constant over all three views. We can therefore write

$$\mathbf{a}_{11} \cdot \mathbf{a}_{21} = \mathbf{a}_{12} \cdot \mathbf{a}_{22}, \quad (5)$$

$$\mathbf{a}_{11} \cdot \mathbf{a}_{21} = \mathbf{a}_{13} \cdot \mathbf{a}_{23}. \quad (6)$$

In terms of components these six equations become

$$\begin{aligned} z_{11}^2 - z_{12}^2 + c_1 &= 0, \\ z_{11}^2 - z_{13}^2 + c_2 &= 0, \\ z_{21}^2 - z_{22}^2 + c_3 &= 0, \\ z_{21}^2 - z_{23}^2 + c_4 &= 0, \\ z_{11}z_{21} - z_{12}z_{22} + c_5 &= 0, \\ z_{11}z_{21} - z_{13}z_{23} + c_6 &= 0, \end{aligned} \quad (7)$$

where

$$\begin{aligned} c_1 &= x_{11}^2 + y_{11}^2 - x_{12}^2 - y_{12}^2, \\ c_2 &= x_{11}^2 + y_{11}^2 - x_{13}^2 - y_{13}^2, \\ c_3 &= x_{21}^2 + y_{21}^2 - x_{22}^2 - y_{22}^2, \\ c_4 &= x_{21}^2 + y_{21}^2 - x_{23}^2 - y_{23}^2, \end{aligned} \quad (8)$$

$$c_5 = x_{11}x_{21} + y_{11}y_{21} - x_{12}x_{22} - y_{12}y_{22},$$

$$c_6 = x_{11}x_{21} + y_{11}y_{21} - x_{13}x_{23} - y_{13}y_{23}.$$

Equations 7 express the rigidity aspect of the RFA principle. To express that the motion must also be about a fixed axis, we observe that for fixed-axis motion the plane defined by the three positions of point A_1 is parallel to the plane defined by the three positions of point A_2 . Thus we can write

$$(\mathbf{a}_{11} - \mathbf{a}_{12}) \cdot [(\mathbf{a}_{11} - \mathbf{a}_{13}) \times (\mathbf{a}_{21} - \mathbf{a}_{22})] = 0,$$

$$(\mathbf{a}_{11} - \mathbf{a}_{12}) \cdot [(\mathbf{a}_{11} - \mathbf{a}_{13}) \times (\mathbf{a}_{21} - \mathbf{a}_{23})] = 0. \quad (9)$$

In terms of components, equations 9 become

$$a_1 z_{11} + a_2 z_{12} + a_3 z_{13} + a_4 z_{21} + a_5 z_{22} = 0,$$

$$a_6 z_{11} + a_7 z_{12} + a_8 z_{13} + a_4 z_{21} + a_5 z_{23} = 0, \quad (10)$$

where

$$a_1 = (x_{12} - x_{13})(y_{21} - y_{22})$$

$$- (x_{21} - x_{22})(y_{12} - y_{13}),$$

$$a_2 = (x_{21} - x_{22})(y_{11} - y_{13})$$

$$- (x_{11} - x_{13})(y_{21} - y_{22}),$$

$$a_3 = (x_{11} - x_{12})(y_{21} - y_{22})$$

$$- (x_{21} - x_{22})(y_{11} - y_{12}),$$

$$a_4 = (x_{11} - x_{12})(y_{11} - y_{13})$$

$$- (x_{11} - x_{13})(y_{11} - y_{12}), \quad (11)$$

$$a_5 = -a_4,$$

$$a_6 = (x_{12} - x_{13})(y_{21} - y_{23})$$

$$- (x_{21} - x_{23})(y_{12} - y_{13}),$$

$$a_7 = (x_{21} - x_{23})(y_{11} - y_{13})$$

$$- (x_{11} - x_{13})(y_{21} - y_{23}),$$

$$a_8 = (x_{11} - x_{12})(y_{21} - y_{23})$$

$$- (x_{21} - x_{23})(y_{11} - y_{12}).$$

Hoffman and Bennett [1] find that equations 7 have, in general, 64 solutions. Furthermore, they find that equations 10 eliminate, generically, all but two of these 64 solutions. Therefore, to find the RFA interpretations appropriate to a given set of image data we can solve equations 7 and then eliminate all but two interpretations using equations 10. Or we can solve equa-

tions 7 and 10 together as a system. A convenient method for solving either system is provided by the homotopy principle.

THE HOMOTOPY PRINCIPLE

The method we use to obtain all solutions to a system of nonlinear equations is called the "homotopy" or "path-following" method [18-20]. In this section we summarize the main ideas and results of this method as it pertains to solving equations 7, or 7 and 10. For a comprehensive survey of path-following methods, we refer the reader to the book by Garcia and Zangwill [18].

The idea behind the homotopy method is this. Suppose we wish to solve the system of equations

$$F(x) = 0,$$

where $x \in \mathbf{R}^m$ and $F: \mathbf{R}^m \rightarrow \mathbf{R}^m$. Suppose there is a "simpler" system, $G(x) = 0$, whose solutions we have already obtained. Suppose x_0 is such a solution, i.e., $G(x_0) = 0$. We contrive a path $\bar{x}(t) \in \mathbf{R}^m$, $t \in [0, 1]$, which starts at x_0 and ends at a solution of $F(x) = 0$. We do this by enlisting yet another function, the "homotopy" $H: \mathbf{R}^{m+1} \rightarrow \mathbf{R}^m$, such that

$$H(x, 0) = G(x) \quad \text{and} \quad H(x, 1) = F(x). \quad (12)$$

We then stipulate that, at each t , the path $\bar{x}(t)$ be a solution to the system $H(\cdot, t) = 0$:

$$H(\bar{x}(t), t) = 0, \quad \forall t \in [0, 1]. \quad (13)$$

Colloquially, as t increases from 0 to 1, the simple system $G(x) = 0$ is "bent" into the desired system $F(x) = 0$, and $\bar{x}(t)$ is a path of solutions to progressively changing systems of equations. Hence the terms "homotopy" and "path following." If we can specify G and H so that the paths $\bar{x}(t)$ may be computed, then we can solve the desired system. The homotopies discussed in Garcia and Zangwill are of the form

$$H(x, t) = (1 - t)G(x) + tF(x). \quad (14)$$

This convex combination of G and F , also called "linear" homotopy, is the one we shall use.

In order to put this idea into practice, we need to investigate certain questions, such as:

- (i). Under what conditions does the set of all solutions (x, t) to $H(x, t) = 0$ consist of a union of smooth paths?
- (ii). How may we compute the paths?
- (iii). How can we ensure that the paths in (i) will end at a solution to $F(x) = 0$?
- (iv). For which F can *all* solutions to $F(x) = 0$ be obtained by a judicious choice of G ?

As our purpose is utilitarian rather than expository, we refer the reader to the book by Garcia and Zangwill for general answers to these questions (as well as a host of applications to nonlinear problems). Here we will restrict ourselves to the instance of $F = (F_k)_{k=1}^m$ being composed of functions F_k which are *polynomials* on \mathbf{R}^m .

It turns out that in order to obtain all solutions, we will need to complexify, as in [21]. That is, we will think of our functions F, G as being defined not on \mathbf{R}^m but on \mathbf{C}^m , by replacing the variables $x_j \in \mathbf{R}$ ($x = (x_j)_{j=1}^m$) by $z_j \in \mathbf{C}$. We accordingly think of our paths as lying in \mathbf{C}^m . Notice that in doing so the coefficients of the polynomials F_k remain real.

It is convenient, for computational purposes, to make the following replacements into real and imaginary parts:

- (15a) $\mathbf{C}^m \rightarrow \mathbf{R}^{2m}$ by $z_j = u_{2j-1} + iu_{2j}$. Write $u = (u_l)_{l=1}^{2m}$, $z = (z_j)_{j=1}^m$, and $n = 2m$.
- (15b) If $H_k(z) = \bar{H}_{2k-1}(u, t) + i\bar{H}_{2k}(u, t)$, where $\bar{H}_{2k-1}, \bar{H}_{2k}$ are real and imaginary parts respectively, replace $H(z)$ by $\bar{H}(u, t)$, where

$$\bar{H}: \mathbf{R}^{2m} \rightarrow \mathbf{R}^{2m}.$$

We call this return to real numbers the “expansion” of the complex-variable case. The two are completely equivalent for the sake of path following because $H(z, t) = 0$ if and only if, for the u corresponding to z , $\bar{H}(u, t) = 0$.

Assuming, then, that we have complexified, we may give precise answers to the questions (i) through (iv) above.

Theorem 1. Let $F: \mathbf{C}^m \rightarrow \mathbf{C}^m$, such that its com-

ponent F_k is a polynomial on \mathbf{C}^m of degree d_k with real coefficients, $1 \leq k \leq m$. Let the components of $G: \mathbf{C}^m \rightarrow \mathbf{C}^m$ be

$$G_k(z) = z^{d_k+1} - 1. \quad (16)$$

Let H be as in equation (14) and let $D = \prod_{k=1}^n (d_k + 1)$.

- (i). The set of all (z, t) satisfying $H(z, t) = 0$ may be sorted into exactly D smooth paths.
- (ii). Every path $\bar{z}(t)$ in (i) will either satisfy
 - (a) $F(\bar{z}(1)) = 0$, i.e., $\bar{z}(1)$ is the desired solution, or
 - (b) $\bar{z}(t)$ is finite for $0 \leq t < 1$ and $\bar{z}(t) \rightarrow \infty$ as $t \rightarrow 1$.
- (iii). If $F(x) = 0$ has a finite number of solutions, every solution to $F(x) = 0$ will be obtained as the endpoint of some path in (i).

This theorem is a corollary to results stated in chapters 1, 2, 3, 18, and 22 of Garcia and Zangwill. Part (iii) indicates one reason why homotopy methods might, for certain purposes, be preferable to some other global optimization methods such as annealing: homotopy methods provide a *systematic* way to find *all* solutions to one’s equations. Note that the theorem says nothing about the infinite solutions case. However in the particular case of RFA motion, thanks to the result of Hoffman and Bennett [1], we know that the theorem in its entirety applies to the system of equations 7 we wish to solve. Moreover, it answers all the questions we asked above, except for the computational one, to which we now turn.

We will express the equations governing the paths in the notation of the replacements (15), i.e., in terms of the expanded system. That is, imagine that we have $n = 2m$ polynomials \bar{H}_l , each in the n variables u_1, \dots, u_n . We put $t = u_{n+1}$, i.e., we treat t as yet another variable. We imagine, moreover, that each path is parametrized by some new real variable p . Define the Jacobian \bar{H}' of $\bar{H} = (\bar{H}_l)_{l=1}^n$ to be the $n \times (n+1)$ matrix

$$\bar{H}'_{jk} = \frac{\partial \bar{H}_j}{\partial u_k}, \quad 1 \leq j \leq n, 1 \leq k \leq n+1. \quad (17)$$

If we differentiate the equation $\bar{H}(\bar{u}) = 0$, where $\bar{u} = (u, t)$, with respect to the parameter p we find, using the chain rule, that every path must satisfy

$$\sum_k \bar{H}'_{j,k}(\bar{u}) \frac{d\bar{u}_k}{dp} = 0$$

or, in matrix notation,

$$\bar{H}'(\bar{u}) \frac{d\bar{u}}{dp} = 0. \quad (18)$$

The following is then proved in Garcia and Zangwill, chapter 2:

Theorem 2. For every path in \mathbb{R}^n of the homotopy \bar{H} , there is a parametrization $p \rightarrow \bar{u}(p)$ of that path such that equation (18) is equivalent to

$$\frac{d\bar{u}_j}{dp} = (-1)^j \det[\bar{H}'_{-j}(\bar{u})], \quad 1 \leq j \leq n+1, \quad (19)$$

where \bar{H}'_{-j} is the matrix \bar{H}' with its j th column deleted.

Equation (19) is called the *Basic Differential Equation* (BDE) and forms the basis for computation. That is, we set, at $p = 0$, $u_{2k-1}(0) + iu_{2k}(0)$ to be some $(d_k + 1)$ th root of unity. We set $u_{n+1}(0) = t(0) = 0$. Then $u(0) = (u_j(0))_{j=1}^n$ is a solution to $\bar{H}(\cdot, 0) = 0$. The path issuing from this point is then computed from the BDE, equation (19), until $t = 1$. Theorem 1 guarantees that this will happen, or that the solution will blow up. We then repeat this process for all possible starting points $u(0)$.

THE RIGIDITY ALGORITHM

In this section we use the homotopy methods just described to develop an algorithm for the solution of equations 7 and thereby for the construction of the appropriate rigid interpretations. Under the assumption that the correspondence between points in the two frames is known, we describe the algorithm step by step.

STEP 1.

Take as inputs the image coordinates, $x'_{i,j}$ and $y'_{i,j}$, $i = 1, 2, 3$, $j = 1, 2, 3$, of the three views of three points. (The goal is to infer the $z_{i,j}$'s.)

STEP 2.

Choose one of the three points as the origin, and compute the coordinates $x_{i,j}$ and $y_{i,j}$, $i = 1, 2$, $j = 1, 2, 3$, of the remaining two points relative to this origin.

STEP 3.

Starting at the 729 solutions to the simple equations

$$\begin{aligned} z_{11}^3 - 1 &= 0, \\ z_{12}^3 - 1 &= 0, \\ z_{13}^3 - 1 &= 0, \\ z_{21}^3 - 1 &= 0, \\ z_{22}^3 - 1 &= 0, \\ z_{23}^3 - 1 &= 0, \end{aligned}$$

apply the homotopy method described in the previous section to the homotopy equations

$$\begin{aligned} (1-t)(z_{11}^3 - 1) + t(z_{11}^2 - z_{12}^2 + c_1) &= 0, \\ (1-t)(z_{12}^3 - 1) + t(z_{11}^2 - z_{13}^2 + c_2) &= 0, \\ (1-t)(z_{13}^3 - 1) + t(z_{21}^2 - z_{22}^2 + c_3) &= 0, \\ (1-t)(z_{21}^3 - 1) + t(z_{21}^2 - z_{23}^2 + c_4) &= 0, \\ (1-t)(z_{22}^3 - 1) + t(z_{11}z_{21} - z_{12}z_{22} + c_5) &= 0, \\ (1-t)(z_{23}^3 - 1) + t(z_{11}z_{21} - z_{13}z_{23} + c_6) &= 0. \end{aligned}$$

As described before, these equations must be decomposed into real and imaginary parts for use by the homotopy algorithm. The result of step 3 is a set of 64 solutions, some possibly complex, to the equations 7. (This step could be accomplished more quickly were 729 processors to work in parallel, each following its own path.)

STEP 4.

Eliminate all extra solutions obtained at step 3 which do not satisfy equations 10.

STEP 5.

Return the two interpretations that survive from step 4.

In Step 3 we pick $z_{ij}^3 = 0$ as the starting point of our homotopy because (1) we know the solutions to these equations and (2) these equations have the lowest degree that guarantees that we will find all solutions to the equations 7 (see Theorem 1). Had we

used, say, $z_{ij}^2 = 0$ as our starting point we would not obtain all solutions to equations 7; had we used, say, $z_{ij}^4 = 0$ as our starting point we would obtain all solutions to equations 7, but would have to follow many more paths to do so (4^6 , i.e., 4096, instead of 3^6 , i.e., 729). In this latter case, following more paths need not require more time, since the paths can be followed in parallel, but it would then require more hardware to implement the following of the extra paths.

In the case of dynamic images having more than three points in motion, the above algorithm could be applied to each subset of three points. Those subsets discovered to have identical axes of rotation and identical angular velocities could be identified as moving together rigidly. In this way the "local" analyses could be combined to give a more "global" RFA interpretation.

CONCLUSION

We conclude with a discussion of the abstract structure of the inference carried out by this algorithm. The possible premises for this inference are all possible sets of data $(x_{i,j}, y_{i,j})$, $i = 1, 2, j = 1, 2, 3$. (Here i only goes to 2 since one of the points is taken to be the origin in \mathbb{R}^3 .) Thus the space of all possible premises is \mathbb{R}^{12} . We denote this space by Y . The possible interpretations for this inference are all possible sets of data $(x_{i,j}, y_{i,j}, z_{i,j})$, $i = 1, 2, j = 1, 2, 3$. Thus the space of all possible interpretations is \mathbb{R}^{18} . We denote this space by X . X and Y are related by a map $\pi: X \rightarrow Y$ induced by the orthographic projection $(x, y, z) \mapsto (x, y)$. Only a small subset of the interpretations in X are distinguished in the sense of being RFA interpretations; this subset is defined by equations 7. We denote these distinguished interpretations by E . Similarly, only a small subset of the premises in Y are distinguished in the sense of being compatible with RFA interpretations; this subset is $\pi(E)$. We denote these distinguished premises by S . Finally, there are, generically, just two RFA interpretations compatible with each premise in the subset S . Without information to the contrary, these two interpretations are equally likely. Thus the conclusion of the inference for any premise in S is a probability measure (which, for example, gives weight of one half to each of the two interpretations). We denote the collection of all

such probability measures, one for each point of S , by η .

Put simply, then, the abstract structure of this inference is a six-tuple (X, Y, E, S, π, η) , where the members of the six-tuple are as just defined.

This structure is a special case of a formal entity called an *observer* [22-25]. An observer is a six-tuple, $((X, \mathcal{X}), (Y, \mathcal{Y}), E, S, \pi, \eta)$, satisfying the following conditions:

1. (X, \mathcal{X}) and (Y, \mathcal{Y}) are measurable spaces. $E \in \mathcal{X}$ and $S \in \mathcal{Y}$.
2. $\pi: X \rightarrow Y$ is a measurable surjective function with $\pi(E) = S$.
3. Let (E, \mathcal{E}) and (S, \mathcal{S}) denote the measurable spaces on E and S respectively induced from those of X and Y . Then η is a markovian kernel on $S \times \mathcal{E}$ such that, for each s , $\eta(s, \cdot)$ is a probability measure supported in $\pi^{-1}\{s\} \cap E$.

Suppose μ_X is some unbiased measure (or, more generally, a measure class) on the measurable space (X, \mathcal{X}) , e.g., μ_X might be unbiased in the sense that it is invariant for the principle homogeneous action of

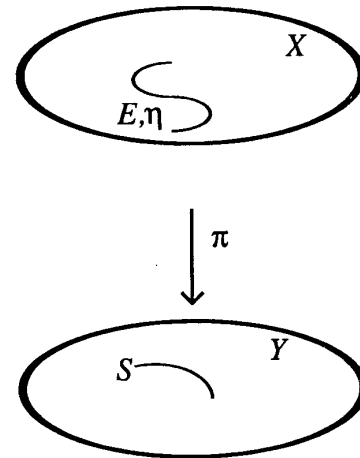


FIGURE 1. Illustration of an observer.

a group on X . Then an observer is called *ideal* if

$$\mu_X(\pi^{-1}(S) - E) = 0.$$

Every point of $\pi^{-1}(S) - E$ is a false target, i.e., a nondistinguished configuration that projects to a dis-

tinguished premise; thus an observer is ideal if its unbiased measure for false targets is zero. A sufficient condition for this to hold is that S have measure zero in Y (with respect to the measure $\pi_*\mu_X$ defined, for all $A \in \mathcal{Y}$, by $\pi_*\mu_X(A) = \mu_X(\pi^{-1}(A))$). This explains the “Minimal false targets” property introduced in section 1.

Bennett et al. propose that the abstract structure of every perceptual inference, whether biological or nonbiological, whether visual or auditory or in any other modality, is without exception an instance of an observer [22-25]. That is, Bennett et al. propose that the definition of an observer provides a formal grounding for the field of perception. (This is in much the same way that the definition of a Turing machine provides a formal grounding for the field of computation: all computations are instances of a Turing machine just as all acts of perception are, according to Bennett et al., instances of an observer.) The perception of structure from motion discussed in this paper provides one example in support of this proposal.

ACKNOWLEDGEMENTS

We thank M. Bodduluri, M. Braunstein, H. Resnikoff, and W. Richards for helpful discussions. This work was supported by National Science Foundation grant IRI-8700924 and by Office of Naval Research contracts N00014-85-K-0529 and N00014-88-K-0354.

REFERENCES AND NOTES

- [1] D. Hoffman and B. Bennett, “The computation of structure from fixed-axis motion: Rigid structures,” *Biol. Cybern.*, **54**, 71–83 (1986).
- [2] J. Webb and J. Aggarwal, “Structure from motion of rigid and jointed objects,” *Artif. Intell.*, **19**, 107–130 (1982).
- [3] D. Hoffman, “The interpretation of visual illusions,” *Sci. Am.*, **249**, 154–162 (1983).
- [4] D. Hoffman and B. Flinchbaugh, “The interpretation of biological motion,” *Biol. Cybern.*, **42**, 195–204 (1982).
- [5] S. Ullman, *The interpretation of visual motion*, Cambridge, Mass.: MIT Press, (1979).
- [6] B. Green, “Figure coherence in the kinetic depth effect,” *J. Exper. Psychol.*, **62**, 272–282 (1961).
- [7] J. J. Gibson and E. J. Gibson, “Continuous perspective transformations and the perception of rigid motion,” *J. Exper. Psychol.*, **54**, 129–138 (1957).
- [8] C. Hay, “Optical motions and space perception—an extension of Gibson’s analysis,” *Psychol. Rev.*, **73**, 550–565 (1966).
- [9] H. Wallach and D. O’Connell, “The kinetic depth effect,” *J. Exper. Psychol.*, **45**, 205–217 (1953).
- [10] W. Richards, “Structure from stereo and motion,” *Artif. Intell. Lab. Memo*, **731**, Cambridge: MIT (1983).
- [11] M. Braunstein, D. Hoffman, L. Shapiro, G. Andersen and B. Bennett, “Minimum points and views for the recovery of three-dimensional structure,” *J. Exper. Psychol.: Human Percept. and Perform.*, **13**, 3, 335–343 (1987).
- [12] A. Waxman and S. Ullman, “Surface structure and 3-D motion from image flow: A kinematic analysis,” CAR-TR-24, Center for Automation Research, University of Maryland, (1983).
- [13] K. Wohn and A. Waxman, “Contour evolution, neighborhood deformation and local image flow: Curved surfaces in motion,” CAR-TR-134, Center for Automation Research, University of Maryland, (1986).
- [14] J. Aloimonos and A. Bandyopadhyay, “Perception of structure from motion: Lower bound results,” TR-158, Department of Computer Science, University of Rochester, (1985).
- [15] N. Grzywacz and E. Hildreth, “Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations,” *J. Opt. Soc. Am. A*, **4**, 3, 503–518 (1987).
- [16] E. Adelson, “Rigid objects that appear highly non-rigid,” *Invest. Ophthalmol. Vision Sci. Suppl.*, **26**, 56 (1985).
- [17] J. Koenderink and A. Van Doorn, “Local structure of movement parallax of the plane,” *J. Opt. Soc. Am.*, **66**, 717–723 (1976).
- [18] W. Zangwill and C. Garcia, *Pathways to solutions, fixed points, and equilibria*, Englewood Cliffs: Prentice-Hall Inc. , (1981).
- [19] W. Forster, *Numerical solutions of highly non-linear problems*, Amsterdam: North-Holland (1980).

- [20] B. Eaves, F. Gould, H. Peitgen, and M. Todd, *Homotopy methods and global convergence*, New York: Plenum Press (1983).
- [21] The reader may wonder at the need to complexify even in instances where the solutions to G and F are both real. Garcia and Zangwill provide examples where the paths, even in such instances, have nonzero imaginary parts in between $\bar{x}(0)$ and $\bar{x}(1)$. In any event, the complex field, being algebraically closed, is the appropriate field over which solutions to polynomials should be investigated.
- [22] B. Bennett, D. Hoffman, and C. Prakash, "Perception and computation," Proc. IEEE First International Conf. Computer Vis., 356–364 (1987).
- [23] B. Bennett, D. Hoffman, and C. Prakash, "Computational vision and observer theory," (under review).
- [24] B. Bennett, D. Hoffman, and C. Prakash, *Observer mechanics*, New York: Academic Press (1989, to appear).
- [25] B. Bennett and D. Hoffman, "Perceptual representations: meaning and truth conditions," in *Cognition and representation* (S. Schiffer and S. Steele, Eds.), Boulder, CO: Westview Press (1988).