Biol. Cybern. 51, 293–300 (1985)

Biological Cybernetics

The Computation of Structure from Fixed-Axis Motion: Nonrigid Structures

B. M. Bennett[1] and D. D. Hoffman[2]

1 Department of Mathematics, University of California, Irvine, USA
2 School of Social Sciences, University of California, Irvine, USA

Abstract. We show that an assumption of rigidity or quasi-rigidity is not necessary, in principle, for the computation of three-dimensional structure and motion from changing retinal images. In particular, we show that the three-dimensional structure of certain nonrigid objects, namely objects whose texture elements rotate about a common axis but at varying angular velocities, can in principle be computed from three successive retinal images of four texture elements, or from four successive images of two texture elements. We then show that in both cases the computed structure matches the actual structure of the object with probability one.

1 Introduction

Theoretical discussions of the recovery of three-dimensional structure from changing retinal images generally employ an assumption of object rigidity (Koenderink and van Doorn 1975; Ullman, 1979; Longuet-Higgins and Prazdny, 1980; Hoffman and Flinchbaugh, 1982; Bobick, 1983; Hoffman, 1983; Hoffman and Bennett, 1984). When the assumption of rigidity is dropped it is sometimes replaced by an assumption of quasi-rigidity (Ullman, 1983), resulting in the computation of the most rigid three-dimensional structure and motion compatible with the retinal motion. Such analyses recommend themselves because the visual world does, in fact, contain rigid and quasi-rigid structures, and because it is plausible, even necessary, that the visual system should exploit such facts about the visual world to infer its structure from retinal images.

Not all visual objects are rigid, however, and this leads to the following question: Under what circumstances can one compute the three-dimensional structure of a nonrigid object from its changing retinal images? We begin by presenting one answer to this question: One can recover the three-dimensional structure of nonrigid objects in the special case of rotation about a common axis.

> Given four orthographic (parallel) projections of two points moving at independent angular velocities about a generically chosen fixed axis, the axis of rotation and the relative positions of the points in three dimensions are uniquely determined up to a reflection about the image plane.

By "generically chosen fixed axis" we mean any axis of rotation other than one parallel to or orthogonal to the image plane.

2 Proof of First Claim

The proof, in outline, is as follows. We note that the two points sweep out circles in space, which project to two ellipses of identical eccentricity and inclination in the image. We show that this implies that the quadratic coefficients in the equations for the ellipses can be taken to be identical. The four views of two points then give eight linear equations in the eight unknown coefficients of the equations for the two ellipses. Using a "lower semi-continuity" argument, we show that this system of equations is generically nondegenerate, and therefore that in general it has a unique solution. Finally, we use the extra constraint that the minor axes of the two ellipses must be aligned, giving an inconsistent set of nine equations in eight unknowns. If all nine equations are simultaneously satisfied by the image data, we can with confidence make the inference from the two-dimensional images to the nonrigid three-dimensional structure.

We consider a projection, π, of $\mathscr{R}^3$, with coordinates (x, y, z), onto a plane through the origin with unit normal vector $\mathbf{N}$ and orthogonal coordinate system (u, v) induced by the metric on $\mathscr{R}^3$

$$\mathscr{R}^2 \overset{\pi}{\leftarrow} \mathscr{R}^3 ,$$
$$(u, v) \leftarrow (x, y, z) ,$$

where $u(x, y, z)$, $v(x, y, z)$ are (orthogonal) linear forms, say

$$\begin{pmatrix} u \\ v \\ n \end{pmatrix} = \mathbf{M} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

$\mathbf{M}$ is an orthogonal matrix, say

$$\mathbf{M} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & l \end{pmatrix},$$

and n is the coordinate along the N axis. Note that

$$u(x+\alpha, y+\beta, z+\gamma) = u(x, y, z) + u(\alpha, \beta, \gamma)$$

and

$$v(x+\alpha, y+\beta, z+\gamma) = v(x, y, z) + v(\alpha, \beta, \gamma).$$

Thus $\pi \circ$(translation in $\mathscr{R}^3$ by (α, β, γ)) = translation in $\mathscr{R}^2$ by $(u(\alpha, \beta, \gamma), v(\alpha, \beta, \gamma))$. Similarly, for some dilation, δ,

$$u(\delta x, \delta y, \delta z) = \delta u(x, y, z)$$

and

$$v(\delta x, \delta y, \delta z) = \delta v(x, y, z).$$

Therefore $\pi \circ$(dilation of vectors in $\mathscr{R}^3$ by δ) = (δ-dilation in $\mathscr{R}^2$) $\circ \pi$.

We begin with a figure in $\mathscr{R}^3$ consisting of two circles C_1, C_2 with radii r_1, r_2, respectively. We assume these circles lie in parallel planes a distance ζ apart, and that the line joining their centers is perpendicular to the two planes of the respective circles. This geometry is illustrated in Fig. 1. Thus we can express the circles as,

$$C_1: x^2 + y^2 = r_1^2, \; z = \zeta,$$
$$C_2: x^2 + y^2 = r_2^2, \; z = 0. \tag{2.1}$$

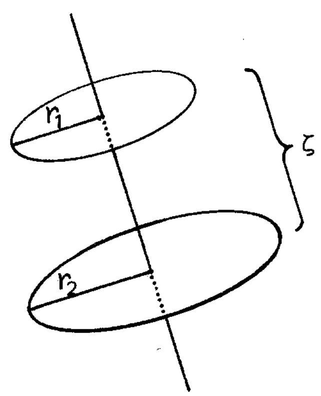

Fig. 1. Geometry underlying the computation of structure from nonrigid motion

However, since $\mathbf{M}$ is orthogonal, $\mathbf{M}^{-1} = \mathbf{M}^T$, and

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{M}^T \begin{pmatrix} u \\ v \\ n \end{pmatrix}. \tag{2.2}$$

Consequently we can rewrite the first Eq. of (2.1) as

$$(au + dv + gn)^2 + (bu + ev + hn)^2 = r_1^2, \tag{2.3a}$$
$$cu + fv + ln = \zeta. \tag{2.3b}$$

We can solve (2.3b) for n, giving

$$n = \frac{\zeta - cu - fv}{l}, \tag{2.4}$$

and use (2.4) to eliminate n from (2.3a),

$$\pi C_1: \left(au + dv + g \frac{\zeta - cu - fv}{l} \right)^2 + \left(bu + ev + h \frac{\zeta - cu - fv}{l} \right)^2 = r_1^2. \tag{2.5}$$

Similarly,

$$\pi C_2: \left(au + dv + g \frac{-cu - fv}{l} \right)^2 + \left(bu + ev + h \frac{-cu - fv}{l} \right)^2 = r_2^2. \tag{2.6}$$

Hence πC_1, πC_2 are quadrics (actually ellipses) in $\mathscr{R}^2$, say,

$$\pi C_1: p_{11}u^2 + p_{12}uv + p_{22}v^2 + p_1 u + p_2 v + p_0 = 0, \tag{2.7}$$
$$\pi C_2: q_{11}u^2 + q_{12}uv + q_{22}v^2 + q_1 u + q_2 v + q_0 = 0. \tag{2.8}$$

Note that in $\mathscr{R}^3$, C_2 = (dilation by δ) $\circ$ (translation by (α, β, γ)) C_1, where $\delta = r_2/r_1$ and (α, β, γ) is some vector. Therefore by the remarks above πC_2 = (dilation by δ) $\circ$ (translation by (s, t)) πC_1, for suitable (s, t). It follows that (2.7) and

$$q_{11}\delta^2(u+s)^2 + q_{12}\delta^2(u+s)(v+t) + q_{22}\delta^2(v+t)^2 + q_1\delta(u+s) + q_2\delta(v+t) + q_0 = 0 \tag{2.9}$$

are both equations for πC_1. Dividing (2.9) by δ^2 and collecting terms according to powers of u and v, we get

$$q_{11}u^2 + q_{12}uv + q_{22}v^2 + f_1 u + f_2 v + f_3 = 0, \tag{2.10}$$

where f_1, f_2, f_3 depend on the q's and s, t, δ. This shows that we may take the first three coefficients of πC_2 to be the same as those for πC_1.

Assuming that q_0 is not zero (i.e., that none of the data points lies on the origin of the retinal coordinate system), we may divide both (2.7) and (2.8) by q_0, i.e., we may make $q_0 = 1$. In consequence we have eight unknown coefficients $p_{11}, p_{12}, p_{22}, p_1, p_2, p_0, q_1, q_2$, which describe the two quadrics, remembering that $q_{11}=p_{11}, q_{12}=p_{12}, q_{22}=p_{22}$, and $q_0=1$.

Now suppose we are given a pair of points $(u(1), v(1))$ and $(u(2), v(2))$ on $\pi(C_1)$, $\pi(C_2)$ respectively. Such data will be called a "view". We get

$$u^2(1)p_{11}+u(1)v(1)p_{12}+v^2(1)p_{22}+u(1)p_1 +v(1)p_2+p_0=0 \tag{2.11}$$

and

$$u^2(2)p_{11}+u(2)v(2)p_{12}+v^2(2)p_{22}+u(2)q_1 +v(2)q_2+1=0\,. \tag{2.12}$$

If we have four views, $\{(u_1(1), v_1(1)), (u_1(2), v_1(2))\}, \ldots, \{(u_4(1), v_4(1)), (u_4(2), v_4(2))\}$, say, we get an 8 by 8 system for the unknowns $p_{11}, p_{12}, p_{22}, p_1, p_2, p_0, q_1, q_2$,

$$\begin{pmatrix} u_1^2(1) & u_1(1)v_1(1) & v_1^2(1) & u_1(1) & v_1(1) & 1 & 0 & 0 \\ u_2^2(1) & u_2(1)v_2(1) & v_2^2(1) & u_2(1) & v_2(1) & 1 & 0 & 0 \\ u_3^2(1) & u_3(1)v_3(1) & v_3^2(1) & u_3(1) & v_3(1) & 1 & 0 & 0 \\ u_4^2(1) & u_4(1)v_4(1) & v_4^2(1) & u_4(1) & v_4(1) & 1 & 0 & 0 \\ u_1^2(2) & u_1(2)v_1(2) & v_1^2(2) & 0 & 0 & 0 & u_1(2) & v_1(2) \\ u_2^2(2) & u_2(2)v_2(2) & v_2^2(2) & 0 & 0 & 0 & u_2(2) & v_2(2) \\ u_3^2(2) & u_3(2)v_3(2) & v_3^2(2) & 0 & 0 & 0 & u_3(2) & v_3(2) \\ u_4^2(2) & u_4(2)v_4(2) & v_4^2(2) & 0 & 0 & 0 & u_4(2) & v_4(2) \end{pmatrix} \begin{pmatrix} p_{11} \\ p_{12} \\ p_{22} \\ p_1 \\ p_2 \\ p_0 \\ q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}. \tag{2.13}$$

We will show by a "lower semi-continuity" argument that the above 8 by 8 system is nonsingular for generic views of generic objects. The precise meaning of this is as follows: We consider the family of all matrices of the type (2.13) parametrized by the 8-dimensional space S of possible $u_i(j)$, $v_i(j)$, $j=1, 2$, and $i=1, \ldots, 4$. That is, we have one matrix for each point of this 8-dimensional space. Note that the determinant of each of these matrices is a polynomial function of the $(u_i(j), v_i(j))$ coordinates of the point of S to which the matrix is associated. We will show below that this function is not uniformly 0 on S, i.e., we will produce a particular point of S for which it is not 0. Therefore the locus of points, D, of S at which the determinant vanishes is an algebraic hypersurface in S, i.e., it has pure dimension one less than S.

We now consider the big family of all sets of four views of 2 points of all objects of the given type in $\mathscr{R}^3$. Assuming orthographic projection, we may consider there to be a 5-dimensional family of such objects parametrized by $(r_1, r_2, \zeta, \alpha_1, \alpha_2)$ where r_1, r_2 are the radii of the circles, ζ their distance apart, and α_1, α_2 give the orientation of the object in space. Moreover, for each object the family of possible sets of four views of 2 points (one point on each circle) is parametrized by a (compact) 8-dimensional space. Thus the big family is parametrized by a 13-dimensional space W.

Actually, the dimension of W is not itself so important; the main point is that the mapping of W to S [which assigns to the real object in $\mathscr{R}^3$ and four views of two points on it the retinal coordinates $(u_i(j), v_i(j))$ of these points] is *algebraic*, i.e., the $u_i(j)$, $v_i(j)$ are *polynomial functions* of r_1, r_2, ζ, α_1, α_2 and the coordinates of the eight points viewed in $\mathscr{R}^3$. Therefore, the inverse image D' of D by this mapping is an algebraic hypersurface of codimension 1 in W (since the computation which follows shows that the image of W by the mapping doesn't lie entirely in D). This means that D' has pure dimension one less than W itself, by the Krull principal ideal theorem (Atiyah-Macdonald, 1969; Hartshorne, 1977), and in particular D' has measure 0 in W. This implies that a random point of W will lie outside of D' with probability one, i.e., a random object and four views of two points on it will result with probability one in retinal image points $(u_i(j), v_i(j))$ for which the determinant of the associated system (2.13) is not 0.

In order to bring to bear the argument above, we must produce a point of W which corresponds to a non-singular matrix (2.13). For this purpose it suffices simply to pick values $u_i(j)$, $v_i(j)$ which make the matrix non-singular. In fact, we can then solve for the coefficients $p_{11}, p_{12}, p_{22}, p_1, p_2, p_0, q_1, q_2$, and the given $u_i(j)$, $v_i(j)$ will automatically lie on the corresponding ellipses given by (2.7) and (2.8), with the special conditions $q_0=0$, $p_{11}=q_{11}$, $p_{12}=q_{12}$, and $p_{22}=q_{22}$. These latter conditions imply that the two ellipses are parallel with the same eccentricity and thus correspond to a point of W.

We choose:

$$\begin{aligned} &u_1(1)=1,\ v_1(1)=2, \\ &u_2(1)=-3,\ v_2(1)=4, \\ &u_3(1)=-2,\ v_3(1)=5, \\ &u_4(1)=4,\ v_4(1)=-1, \\ &u_1(2)=2,\ v_1(2)=9, \\ &u_2(2)=-1,\ v_2(2)=-7, \\ &u_3(2)=11,\ v_3(2)=7, \\ &u_4(2)=3,\ v_4(2)=8\,. \end{aligned} \tag{2.14}$$

This choice gives the eight by eight matrix

$$\begin{pmatrix} 1 & 2 & 4 & 1 & 2 & 1 & 0 & 0 \\ 9 & -12 & 16 & -3 & 4 & 1 & 0 & 0 \\ 4 & -10 & 25 & -2 & 5 & 1 & 0 & 0 \\ 16 & -4 & 1 & 4 & -1 & 1 & 0 & 0 \\ 4 & 18 & 81 & 0 & 0 & 0 & 2 & 9 \\ 1 & 7 & 49 & 0 & 0 & 0 & -1 & -7 \\ 121 & 77 & 49 & 0 & 0 & 0 & 11 & 7 \\ 9 & 24 & 64 & 0 & 0 & 0 & 3 & 8 \end{pmatrix}, \quad (2.15)$$

which has determinant 11060280 (found using Mu-math on an IBM PC), showing that for this choice of $u_i(j)$, $v_i(j)$ the rank of the matrix is eight. By the lower semi-continuity of the rank of matrices this implies that the rank of the matrix is generically eight, so that generically four views of two points allows a unique solution for the two quadrics.

3 Inferences

If the visual system, using the method just presented, found that it could compute a three-dimensional structure, would it be justified in asserting that, in fact, an object with that structure was present in the visual field? More formally, what is the value of the conditional probability $P(w/i)$, where w (world) stands for the predicate "is an object with nonrigid structure α rotating about a fixed axis", and i (image) stands for the predicate "has retinal images consistent with being an object with nonrigid structure α rotating about a fixed axis"? We show here that $P(w/i)$ is one, assuming infinite resolution on the retina, by showing that the conditional probability $P(i/\neg w)$ is zero.

If we expand $P(w/i)$ using Bayes theorem we find:

$$P(w/i) = \frac{P(w) \cdot P(i/w)}{P(w) \cdot P(i/w) + P(\neg w) \cdot P(i/\neg w)}. \quad (3.1)$$

Since the numerator and the first term of the denominator are identical and nonzero, it suffices to show that the second term of the denominator is zero in order to show that the entire expression is one. [In particular it is not necessary to make any statement about the prior probability $P(w)$ other than that it is greater than zero.] We show that the second term of the denominator is zero by showing that $P(i/\neg w)$ is zero. For suppose we have eight points (four views of two points) which, rather than chosen to be the projection of a nonrigid structure rotating about a fixed axis, are chosen at random on the image plane. In general, these eight points will have an interpretation as two ellipses of identical eccentricity and inclination (as was established in the lower semi-continuity proof). However, the three-dimensional interpretation will not in general be of two circles whose centers both lie on a line orthogonal to the circles, i.e., the minor axes of the resulting ellipses will not necessarily be collinear. This is because our proof in Sect. 2 nowhere used this independent constraint on the three-dimensional geometry. This constraint can be expressed mathematically as follows. Let $\boldsymbol{P}$ be the quadratic form associated with the two ellipses,

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & \frac{1}{2}p_{12} \\ \frac{1}{2}p_{12} & p_{22} \end{pmatrix}. \quad (3.2)$$

The eigenvectors of $\boldsymbol{P}$ point in the directions of the major and minor axes of the two ellipses (Acton, 1970). The centers of the ellipses, say (x_1, y_1) and (x_2, y_2), are the solutions to the equations

$$\boldsymbol{P}(x_1\, y_1)^T = -\tfrac{1}{2}(p_1\, p_2)^T,$$
$$\boldsymbol{P}(x_2\, y_2)^T = -\tfrac{1}{2}(q_1\, q_2)^T. \quad (3.3)$$

Thus the line defined by the solutions to (3.3) must lie in the direction of the eigenvector of $\boldsymbol{P}$ which is associated with its largest eigenvalue. Eight randomly chosen points will satisfy this extra independent condition only with probability zero (we have an inconsistent set of equations). Thus $P(i/\neg w)$ is zero, and $P(w/i)$ is one. Therefore if the visual system finds that it can compute a three-dimensional structure using the method presented in Sect. 2, it can be quite confident in asserting that, in fact, an object with that structure is present in the visual field. This result makes it possible to design an autonomous processor that computes the three-dimensional structure of some nonrigid (and, as a special case, rigid) objects without recourse to knowledge other than that specified above, i.e., a processor which is "informationally encapsulated" (Fodor, 1983).

4 Three Views of Four Points: Preliminaries

Four views of two points are sufficient, but not necessary, to compute the three-dimensional structure of a nonrigid object from its changing retinal images. A second sufficient condition is the following:

> Given three orthographic projections of four points moving at independent angular velocities about a fixed axis, the axis of rotation and the relative positions of the points in three dimensions are uniquely determined up to a reflection about the image plane.

Our proof, in outline, is a follows. In this section we derive a polynomial equation which holds when the minor axes of two ellipses are collinear. In Sect. 5 we note that fourteen parameters are needed to specify four ellipses of identical eccentricity and inclination. We produce twelve linear equations of the type shown in (2.13), and two quadratic equations which express the condition that the minor axes must be aligned. We show that this system of fourteen equations in 14 unknowns generically has four solutions, of which one is canonically distinguished. To check for false targets, we produce one further sextic equation to test this distinguished solution, a sextic which independently expresses the condition that the minor axes must be aligned.

We begin our proof of this claim by considering two ellipses in the u, v plane with the same eccentricity and inclination, i.e., the ellipses are related by translation and dilation. As we have seen in Sect. 2, this condition is equivalent to the assertion that the three quadratic coefficients of both ellipses are the same (even after one of the equations has been normalized so that its constant term is 1). Thus we may assume the equations are

$$\begin{aligned} E_1 &: p_{11}u^2+p_{12}uv+p_{22}v^2+p_1u+p_2v+p_0=0\,, \\ E_2 &: p_{11}u^2+p_{12}uv+p_{22}v^2+q_1u+q_2v+q_0=0 \end{aligned} \tag{4.1}$$

(and we may take one of p_0 or q_0 to be 1).

We now want to investigate the relations among the coefficients imposed by the additional geometric assumption that (*) *the line joining the centers of the two ellipses is perpendicular to their major axes*, i.e., the minor axes are collinear. As we have remarked in Sect. 3, these conditions are equivalent to saying that the two circles in 3-space, of which the ellipses are the projections, lie in parallel planes, and the line joining their centers is perpendicular to those planes, i.e., that the figure has axial symmetry about this line.

To proceed with our analysis of the condition (*) we first rotate the (u, v) coordinate system so that the new coordinate axes are parallel to the major and minor axes of the ellipses. This amounts to finding a rotation matrix

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{4.2}$$

so that the substitution $u=az+bw$, $v=cz+dw$ in (4.1) eliminates the zw term. (This amounts to saying that $\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A}=\boldsymbol{A}^{-1}\boldsymbol{P}\boldsymbol{A}$ is diagonal, so it is really the same as the eigenvector description of the major and minor axes mentioned in Sect. 3.) Now for an arbitrary matrix of the form (4.2), after making the above substitution and collecting terms in z and w, the system (4.1) becomes:

$$\begin{aligned} E_1 &: (a^2p_{11}+acp_{12}+c^2p_{22})z^2 \\ &\quad +(2abp_{11}+(ad+bc)p_{12}+2cdp_{22})zw \\ &\quad +(b^2p_{11}+bdp_{12}+d^2p_{22})w^2+(ap_1+cp_2)z \\ &\quad +(bp_1+dp_2)w+p_0=0\,, \\ E_2 &: (a^2p_{11}+acp_{12}+c^2p_{22})z^2 \\ &\quad +(2abp_{11}+(ad+bc)p_{12}+2cdp_{22})zw \\ &\quad +(b^2p_{11}+bdp_{12}+d^2p_{22})w^2+(aq_1+cq_2)z \\ &\quad +(bq_1+dq_2)w+q_0=0\,. \end{aligned} \tag{4.3}$$

In our case, since A is a rotation matrix, we have $a=\cos\alpha$ for suitable α, and $b=-\sqrt{1-a^2}$, $c=\sqrt{1-a^2}$, $d=a$. The angle α must therefore be chosen so that

$$\begin{aligned} &-2a\sqrt{1-a^2}\,p_{11}+(a^2-(\sqrt{1-a^2})^2)p_{12} \\ &\quad +2a\sqrt{1-a^2}\,p_{22}=0\,. \end{aligned} \tag{4.4}$$

Note that the coefficients of the p_{ij} in this equation are conveniently expressed in terms of $\sin(2\alpha)$ and $\cos(2\alpha)$. If we let $e=\cos(2\alpha)$, we may write (4.4) in the form

$$-\sqrt{1-e^2}\,p_{11}+ep_{12}+\sqrt{1-e^2}\,p_{22}=0 \tag{4.5}$$

and we can then solve for e:

$$e=\sqrt{\frac{(p_{11}-p_{22})^2}{p_{12}^2+(p_{11}-p_{22})^2}}\,. \tag{4.6}$$

(By assuming that $|a|\leq\pi/4$ we may take $e\geq 0$; this assumption is justified geometrically since we can always find either a positive or negative α with this property which will align the ellipse axes in some way with the coordinate axes.)

We will assume that e is as in (4.6), and using the facts that $a=\sqrt{(1+e)/2}$, $\sqrt{1-a^2}=\sqrt{(1-e)/2}$, henceforth the quantities a, b, c, d will refer to $\sqrt{(1+e)/2}$, $-\sqrt{(1-e)/2}$, $\sqrt{(1-e)/2}$, $\sqrt{(1+e)/2}$ respectively. Equations (4.3) now become

$$\begin{aligned} E_1 &: (a^2p_{11}+acp_{12}+c^2p_{22})z^2 \\ &\quad +(b^2p_{11}+bdp_{12}+d^2p_{22})w^2 \\ &\quad +(ap_1+cp_2)z+(bp_1+dp_2)w+p_0=0\,, \\ E_2 &: (a^2p_{11}+acp_{12}+c^2p_{22})z^2 \\ &\quad +(b^2p_{11}+bdp_{12}+d^2p_{22})w^2 \\ &\quad +(aq_1+cq_2)z+(bq_1+dq_2)w+q_0=0\,. \end{aligned} \tag{4.7}$$

We now use the fact that the center of an ellipse given by an equation of this type corresponds to the translation of the coordinate axes which eliminates the linear terms from the equations, i.e., which "completes the square" in z and w. Thus we find that the centers are

$$\begin{aligned} E_1 &: \left(\frac{ap_1+cp_2}{2(a^2p_{11}+acp_{12}+c^2p_{22})},\; \frac{bp_1+dp_2}{2(b^2p_{11}+bdp_{12}+d^2p_{22})}\right), \\ E_2 &: \left(\frac{aq_1+cq_2}{2(a^2p_{11}+acp_{12}+c^2p_{22})},\; \frac{bq_1+dq_2}{2(b^2p_{11}+bdp_{12}+d^2p_{22})}\right). \end{aligned} \tag{4.8}$$

Therefore, since the coordinate axes are now aligned with the major and minor axes of the ellipse (although we do not know in what order), it follows that the condition (*) implies that one of the following equations hold:

$$ap_1+cp_2=aq_1+cq_2, \quad bp_1+dp_2=bq_1+dq_2. \tag{4.9}$$

Recalling the values of a, b, c, d in terms of e, we obtain

$$(1+e)(p_1-q_1)^2=(1-e)(q_2-p_2)^2, \quad (1-e)(p_1-q_1)^2=(1+e)(q_2-p_2)^2, \tag{4.10}$$

or equivalently

$$e[(p_1-q_1)^2+(q_2-p_2)^2]=[(q_2-p_2)^2-(p_1-q_1)^2], \quad -e[(p_1-q_1)^2+(q_2-p_2)^2]=[(q_2-p_2)^2-(p_1-q_1)^2], \tag{4.11}$$

which, upon using (4.6) to eliminate e, becomes

$$(p_{11}-p_{22})[(p_1-q_1)^2+(q_2-p_2)^2] =[p_{12}^2+(p_{11}-p_{22})^2]^{1/2}[(q_2-p_2)^2-(p_1-q_1)^2], \quad -(p_{11}-p_{22})[(p_1-q_1)^2+(q_2-p_2)^2] =[p_{12}^2+(p_{11}-p_{22})^2]^{1/2}[(q_2-p_2)^2-(p_1-q_1)^2]. \tag{4.12}$$

5 Three Views of Four Points: Proof

We now consider four ellipses, E_1, E_2, E_3, E_4 in the (u, v) plane, which satisfy pairwise the condition (*) of Sect. 4. In other words, these ellipses are projections onto the (u, v) plane of four circles in 3-space, whose centers lie on a single line perpendicular to the planes of the circles. We will also assume that the direction of projection is not parallel to this line, and that the centers of the circles are distinct.

The equations for the ellipses are:

$$\begin{aligned} E_1&: p_{11}u^2+p_{12}uv+p_{22}v^2+p_1u+p_2v+p_0=0, \\ E_2&: p_{11}u^2+p_{12}uv+p_{22}v^2+q_1u+q_2v+q_0=0, \\ E_3&: p_{11}u^2+p_{12}uv+p_{22}v^2+r_1u+r_2v+r_0=0, \\ E_4&: p_{11}u^2+p_{12}uv+p_{22}v^2+s_1u+s_2v+1=0. \end{aligned} \tag{5.1}$$

Let

$$\begin{aligned} L_1&=p_1-q_1,\ L_2=p_2-q_2, \\ M_1&=p_1-r_1,\ M_2=p_2-r_2, \\ N_1&=p_1-s_1,\ N_2=p_2-s_2. \end{aligned} \tag{5.2}$$

Thus for the ellipse pairs (E_1, E_2), (E_1, E_3), and (E_1, E_4) the Eq. (4.11) may be written, respectively,

$$\begin{aligned} e(L_1^2+L_2^2)&=\pm(L_1^2-L_2^2), \\ e(M_1^2+M_2^2)&=\pm(M_1^2-M_2^2), \\ e(N_1^2+N_2^2)&=\pm(N_1^2-N_2^2), \end{aligned} \tag{5.3}$$

where the signs are either all + or all −. We will refer to these two cases as forms (i) and (ii) of (5.3) respectively. Either form (i) or form (ii) holds.

Now we observe that none of $(L_1^2+L_2^2)$, $(M_1^2+M_2^2)$, $(N_1^2+N_2^2)$ are 0, otherwise two of the circles are concentric, since the projection is not parallel to the line joining their centers. Moreover $e \neq 0$ (i.e., $p_{11} \neq p_{22}$), otherwise the ellipses are circles, which would again contradict the projection assumption. It follows that the system (5.3) is equivalent to

$$e=\pm\left(\frac{L_1^2-L_2^2}{L_1^2+L_2^2}\right)=\pm\left(\frac{M_1^2-M_2^2}{M_1^2+M_2^2}\right)=\pm\left(\frac{N_1^2-N_2^2}{N_1^2+N_2^2}\right), \tag{5.4}$$

where the signs are all + or − corresponding to forms (i) and (ii) of (5.3).

If we equate, say, the L and M terms in (5.4), clearing denominators and simplifying we obtain: $L_1^2M_2^2-L_2^2M_1^2=0$. Similarly we obtain $M_1^2N_2^2-M_2^2N_1^2=0$ and $L_1^2N_2^2-L_2^2N_1^2=0$. [The distinction between forms (i) and (ii) disappears in these equations.] However it is evident that any one of these relations is deducible from the other two. Therefore (5.4) is equivalent to the following three equations:

$$e\pm\left(\frac{L_1^2-L_2^2}{L_1^2+L_2^2}\right)=0, \tag{5.5a}$$

$$L_1^2M_2^2-L_2^2M_1^2=0, \tag{5.5b}$$

$$M_1^2N_2^2-M_2^2N_1^2=0. \tag{5.5c}$$

Note that (5.5a) is the same as in (4.11) and is algebraically a sextic relation (if we clear denominators and square both sides to eliminate the radical in e). Note also that (5.5b) and (5.5c) may be written $(L_1M_2+L_2M_1)(L_1M_2-L_2M_1)=0$ and $(M_1N_2+M_2N_1)(M_1N_2-M_2N_1)=0$. Thus (5.3) may finally be written in the form

$$e\pm\left(\frac{L_1^2-L_2^2}{L_1^2+L_2^2}\right)=0, \tag{5.6a}$$

$$L_1M_2\pm L_2M_1=0, \tag{5.6b}$$

$$M_1N_2\pm M_2N_1=0, \tag{5.6c}$$

where the signs may be chosen arbitrarily in each equation. This subsumes both forms (i) and (ii).

Remark. The purpose of transforming (5.3) into (5.6) is, in the first place, computational efficiency. As we will see below, only two of the three equations (any two) would suffice to solve the system (given the views) up to a finite ambiguity; the third equation is then used as a "test" to eliminate this ambiguity, and to rule out "false targets". From this point of view, it is far easier to solve two simultaneous quadrics and use a sextic as a test, than to solve two simultaneous sextics before testing. In the second place it is much easier to observe the "genericity" of the equations when formulated in terms of the quadrics.

We now suppose that we are given three points (i.e., three "views" of a given point) on each ellipse. The i^{th} point on the j^{th} ellipse will be denoted $(u_i(j), v_i(j))$; $i=1, 2, 3, j=1, 2, 3, 4$. For each ellipse we get three linear equations in the unknown coefficients:

$Ap_{11}+Bp_{22}+C$ where the coefficients A, B, C depend on the $\{u_i(j), v_i(j)\}$. Recalling that $p_1, p_2, q_1, q_2, r_1, r_2$, as determined by (5.7a–c) depend on p_{12}, and that p_{12} as determined by (5.7d) depends on the $\{u_i(4), v_i(4)\}$, we find that the coefficients of L_1 and L_2 depend on all but the $\{u_i(3), v_i(3)\}$, the coefficients of M_1 and M_2 depend on all but the $\{u_i(2), v_i(2)\}$, and the coefficients of N_1 and N_2 depend only on the $\{u_i(1), v_i(1)\}$ and $\{u_i(4), v_i(4)\}$.

$$E_1:\begin{pmatrix} u_1^2(1) & u_1(1)v_1(1) & v_1^2(1) & u_1(1) & v_1(1) & 1\\ u_2^2(1) & u_2(1)v_2(1) & v_2^2(1) & u_2(1) & v_2(1) & 1\\ u_3^2(1) & u_3(1)v_3(1) & v_3^2(1) & u_3(1) & v_3(1) & 1 \end{pmatrix}\begin{pmatrix} p_{11}\\ p_{12}\\ p_{22}\\ p_1\\ p_2\\ p_0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}, \tag{5.7a}$$

$$E_2:\begin{pmatrix} u_1^2(2) & u_1(2)v_1(2) & v_1^2(2) & u_1(2) & v_1(2) & 1\\ u_2^2(2) & u_2(2)v_2(2) & v_2^2(2) & u_2(2) & v_2(2) & 1\\ u_3^2(2) & u_3(2)v_3(2) & v_3^2(2) & u_3(2) & v_3(2) & 1 \end{pmatrix}\begin{pmatrix} p_{11}\\ p_{12}\\ p_{22}\\ q_1\\ q_2\\ q_0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}, \tag{5.7b}$$

$$E_3:\begin{pmatrix} u_1^2(3) & u_1(3)v_1(3) & v_1^2(3) & u_1(3) & v_1(3) & 1\\ u_2^2(3) & u_2(3)v_2(3) & v_2^2(3) & u_2(3) & v_2(3) & 1\\ u_3^2(3) & u_3(3)v_3(3) & v_3^2(3) & u_3(3) & v_3(3) & 1 \end{pmatrix}\begin{pmatrix} p_{11}\\ p_{12}\\ p_{22}\\ r_1\\ r_2\\ r_0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}, \tag{5.7c}$$

$$E_4:\begin{pmatrix} u_1^2(4) & u_1(4)v_1(4) & v_1^2(4) & u_1(4) & v_1(4)\\ u_2^2(4) & u_2(4)v_2(4) & v_2^2(4) & u_2(4) & v_2(4)\\ u_3^2(4) & u_3(4)v_3(4) & v_3^2(4) & u_3(4) & v_3(4) \end{pmatrix}\begin{pmatrix} p_{11}\\ p_{12}\\ p_{22}\\ s_1\\ s_2\end{pmatrix}=\begin{pmatrix}-1\\-1\\-1\end{pmatrix}. \tag{5.7d}$$

It is clear by inspection that for generic views $\{(u_i(j), v_i(j)\}$ the last three columns of the first three systems, and the second, fourth and fifth columns of the E_4 system, will be linearly independent. This means that we can use the first system to express p_0, p_1, p_2 in terms of p_{11}, p_{12}, p_{22}, the second to express q_0, q_1, q_2 in terms of p_{11}, p_{12}, p_{22}, the third to express r_0, r_1, r_2 in terms of p_{11}, p_{12}, p_{22}, and the fourth to express s_1, s_2 and p_{12} in terms of p_{11} and p_{22}. The result is that for generic views, the system (5.7) may be used to express p_0, p_1, p_2, q_0, q_1, q_2, r_0, r_1, r_2, s_1, s_2, p_{12} as linear functions in p_{11} and p_{22}. In particular the L_1, L_2, M_1, M_2, N_1, N_2 defined in (5.2) may be expressed as linear functions in p_{11}, p_{22}, i.e., they are each of the form

Let us now consider the quadrics in (5.6). Note that (5.6a) depends on all but the $\{u_i(3), v_i(3)\}$, (5.6b) depends on all the $\{u_i(j), v_i(j)\}$, and (5.6c) depends on all but the $\{u_i(2), v_i(2)\}$. Since any pair of equations from (5.6) depend on different $u_i(j)$, $v_i(j)$ the three equations are mutually independently generic. This means that for generic values of $\{u_i(j), v_i(j)\}$ the three equations will have no common solutions, (since three generic polynomials in two variables have no common solutions). Thus if the $\{u_i(j), v_i(j)\}$ are not plane projections of points on an object rotating about a fixed axis, then with probability one we will not obtain a solution to the system, i.e., there will be no false targets.

If the $\{u_i(j), v_i(j)\}$ are plane projections of points on an object rotating about a fixed axis, then the three equations can have at most four common solutions. Restricting attention for the moment to (5.6b) and (5.6c), it is clear that one solution occurs when $M_1 = M_2 = 0$, since M_1 or M_2 appears in each term of these equations. A second solution occurs when $L_1 = L_2 = 0$, for if we multiply (5.6b) by N_1 and (5.6c) by L_1 and add the two resulting equations we obtain $L_1 N_2 \pm L_2 N_1 = 0$, with the result that L_1 or L_2 now appears in each term of (5.6b) and the newly obtained equation. Similarly, one can show that a third solution occurs when $N_1 = N_2 = 0$. We seek the unique solution of (5.6b) and (5.6c) other than these three. That is, we can identify and discard the three spurious solutions, leaving us with the unique asnwer. The correctness of the answer can be guaranteed with high probability because false targets can be eliminated by testing this unique solution with (5.6a).

6 Conclusion

We have shown that an assumption of rigidity or quasi-rigidity is not necessary, in principle, for the computation of three-dimensional structure and motion from changing retinal images. Two specific results demonstrate this. First, given four orthographic views of two points moving at independent angular velocities about a common fixed axis, the axis of rotation and the relative positions of the points in three dimensions are uniquely determined up to a reflection. Second, given three orthographic views of four points moving at independent angular velocities about a fixed axis, the axis of rotation and the relative positions of the points in three dimensions are uniquely determined up to a reflection about the image plane.

The results presented here suggest further psychophysical experiments to test quantitatively the human visual ability to recover three-dimensional structure from nonrigid motion. Although our ability to perceive nonrigid motion has been explored experimentally several times (see Todd, 1982 for a brief review), some preliminary results by Braunstein (personal communication) seem most relevant to the present analysis. Braunstein has found that if points are made to move at independent angular velocities in circles about a cylinder then one can perceive the cylinder even though none of the visible points move together rigidly. The points appear as ants crawling at different speeds in circles about the cylinder.

We should note that the results proved here assume that the objects under observation are not translating parallel to the image plane. Any such translations must first be nullified by tracking.

We conclude that it is possible to compute the three-dimensional structure of certain nonrigid objects, and that sufficiency conditions for such computations can be obtained with the same degree of rigor as for rigid objects.

Acknowledgements. We thank G.J. Andersen, A. Bobick, M. Braunstein, W. Richards, and R. Yoshino for useful discussions. We thank the Fairchild Laboratory for Artificial Intelligence Research (FLAIR) and the Massachusetts Institute of Technology Laboratory for Artificial Intelligence for support. This material is based upon work supported by NSF Grant No. IST-8413560.

References

Acton, F.S.: Numerical methods that work. London: Harper&Row 1970

Atiyah-Macdonald, M.: Introduction to commutative algebra. Reading, M: Addison-Wesley 1969

Bobick, A.: A hybrid approach to structure-from-motion. Proc. ACM Workshop on Motion: Representation and perception (1983)

Fodor, J.A.: The modularity of mind. Cambridge, MA: MIT Press 1983

Hartshorne, R.: Algebraic geometry. Berlin, Heidelberg, New York: Springer 1977

Hoffman, D.D., Flinchbaugh, B.E.: The interpretation of biological motion. Biol. Cybern. **42**, 195–204 (1982)

Hoffman, D.D.: The interpretation of visual illusions. Sci. Am. **249**, 154–162 (1983)

Hoffman, D.D., Bennett, B.M.: The computation of structure from fixed-axis motion: rigid structures (submitted, 1984)

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

Longuet-Higgins, H.C., Prazdny, K.: The interpretation of a moving retinal image. Proc. R. Soc. London, Ser. B **208**, 385–397 (1980)

Richards, W.A., Rubin, J.M., Hoffman, D.D.: Equation counting and the interpretation of sensory data. Perception **11**, 557–576 (1982)

Todd, J.T.: Visual information about rigid and nonrigid motion: a geometric analysis. J. Exp. Psychol.: Human Percept. Perform. **8**, 238–252 (1982)

Ullman, S.: The interpretation of structure from motion. Proc. R. Soc. London, Ser. B **203**, 405–426 (1979)

Ullman, S.: Maximizing rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion. Massachusetts Institute of Technology, Artificial Intelligence Laboratory Memo 721 (1983)

Received: October 6, 1984

Dr. Don Hoffman
University of California
School of Social Sciences
Irvine, CA 92717
USA