Invariant Descriptors for 3-D Object Recognition and Pose

David Forsyth, *Member, IEEE,* Joseph L. Mundy, Andrew Zisserman, *Member, IEEE,* Chris Coelho, Aaron Heller, *Member, IEEE,* and Charles Rothwell

Abstract—**Invariant descriptors are shape descriptors that are unaffected by object pose, by perspective projection, and by the intrinsic parameters of the camera. These descriptors can be constructed using the methods of invariant theory, which are briefly surveyed.**

A range of applications of invariant descriptors in three-dimensional model-based vision is demonstrated. First, a model-based vision system that recognizes curved plane objects, irrespective of their pose, is demonstrated. Curves are not reduced to polyhedral approximations but are handled as objects in their own right. Models are generated directly from image data. Once objects have been recognized, their pose can be computed. Invariant descriptors for three-dimensional objects with plane faces are described. All these ideas are demonstrated on images of real scenes.

The stability of a range of invariant descriptors to measurement error is treated in detail.

Index Terms—**Computer vision, invariants, pose computation, recognition.**

I. Introduction

SYSTEMS FOR recognizing objects in images have focused largely on fixed, polyhedral models (there are a few exceptions, e.g., [6], [37]–[39]). In general, recognition proceeds by extracting a collection of features from the image, hypothesizing a correspondence between these features and an appropriate set of model features, and determining the position and orientation (pose) of the model from this hypothesis. The hypothesis is verified by using the computed pose to predict further image features and confirming that some or all of these features are found as predicted. Alternatively, the pose of a hypothesized model can be predicted separately from several different sets of features; verification then proceeds by clustering these predictions in the six-dimensional pose parameter space.

Many quite successful systems have been developed around polyhedral models using these techniques (for example, [21], [27], [28], [45]). However, there remain a number of significant difficulties with this approach:

- Representations of curves as polygonal arcs are not unique and are extremely sensitive to the thresholds used to place knots. Current approaches use local measures of curvature to place knots [3], [28] and produce a wide variety or representations that depend on initial conditions and various parameter settings. Similar problems arise in approximating the surface of a three-dimensional object with a polyhedron. As a result, these techniques can break down when dealing with curved objects or curved surfaces.
- In current systems, perspective is approximated by plane affine transformations. This affine approximation is very good when one observes an object from a distance that is an order of magnitude or more greater than the maximum object diameter along the direction of view [45]. However, the affine approximation is inappropriate in some major application areas. The effects of perspective can be pronounced in aerial photographs, particularly in oblique views of such features as roadways, landing strips, and shorelines.
- Fast recognition with a large model library is hard using this approach because recognition is achieved by trying each model in turn and selecting the model that best explains the observed image features, leading to a recognition cost linear in the number of objects in the library. This can be improved if the models consist of an arrangement of a small number of subparts (see, for example, [12]). Unfortunately, this improvement requires that subparts can be distinguished in the image, which is a condition that is sometimes hard to satisfy, particularly with three-dimensional objects. Aspect graph techniques for recognizing three-dimensional objects (see, for example, [22]) exacerbate the problem by vastly expanding the model base.

Current systems invoke pose in *recognition* because shapes measured in images depend not just on the shape of the object viewed but on its pose and the intrinsic parameters of the camera as well. In this paper, we construct shape *descriptors* that are unaffected by the transformation between the object and the image plane. These descriptors are at the heart of the model-based vision systems for planar objects that we demonstrate in Section III. These systems have important, novel features:

Manuscript received October 15, 1990; revised February 28, 1991. This work was supported by Magdalen College, Oxford, England, the United Kingdom Science and Engineering Research Council, the General Electric Coolidge Fellowship, the Department of Physics of the University of Genoa, ELSAG BAILEY, and the General Electric Corporate Research and Development Laboratory, which is supported by DARPA contract DACA-76-86-C-007 and AFOSR contract F49620-89-C-003.

D. Forsyth, A. Zisserman, and C. Rothwell are with the Robotics Research Group, Department of Engineering Science, Oxford University, Oxford, England.

J. L. Mundy and A. Heller are with the General Electric Research and Development Center, Schenectady, NY 12306.

C. Coelho is with ELSAG BAILEY, Italy.

IEEE Log Number 9102643.

Printed by: Black Bear Press Limited, Cambridge, United Kingdom

- Curves are handled as geometric features with an identity and significance of their own and are not broken into line segments.
- All perspective effects are fully accounted for. No approximations are necessary.
- Image curves are represented by invariant shape descriptors, which allow direct indexing into the model library. As a result, recognition is achieved without reference to object pose. There is thus a potential for recognition in a time that is independent of the number of objects in the model library.

Invariant descriptors appear in a number of guises in contemporary vision work but are at present applied unsystematically. For example, the mainstream practice of transforming a set of image features into a distinguished coordinate frame (normally obtained by hypothesizing pose, although Lamdan *et al.* used a more sophisticated system of affine invariant properties based on a system of distinguished frames) and then making measurements in that frame yields invariant descriptors. As a second example, the descriptors used by Grimson *et al.* [19] to control his tree searching algorithm are invariant to plane Euclidean actions. A more systematic application of invariant theory means that invariant descriptors can be computed without the need for distinguished coordinate frames—furthermore, it is possible to produce all the functionally independent invariants for a given situation. In this paper, we introduce the rudiments of invariant theory and demonstrate a number of applications in vision.

A central theme in this work is the idea of the action of a *group,* which is a collection of transformations, that can be composed and inverted. Groups are very effective models of a range of physical effects, particularly rigid motion, and properties that are invariant to such transformations have been widely studied. In vision, the principal transformation is a *perspectivity,* which consists of Euclidean motion in space composed with perspective projection. The product of two perspectivities is not necessarily a perspectivity, which leads to mathematical difficulties; therefore, we focus on the *plane projective group,* which contains all the perspectivities as a subset.

The paper is organized as follows: in Section II, we discuss the mathematics and ideas underlying our use of invariant theory and show a broad range of examples of invariants. In Section III, we show the usefulness of this theory in model-based vision. We demonstrate an affine invariant representation for plane curves using conic curves. We build a simple and effective model-based vision system that uses the affine invariance of this representation to recognize objects with planar faces. Because the descriptor does not change whatever the pose of the object, this system effectively decouples the problem of identifying objects from that of determining their pose. In Section IV, we show that it is possible to recover pose from our invariant representation by solving the simple problem of backprojecting known conics. In Section V, we use these pose results to obtain Euclidean invariant descriptors for three-dimensional objects. In Section VI, we address the effects of noise on invariant descriptors for plane objects.

II. Invariant Theory

We adopt the notation that corresponding entities in two different coordinate frames are distinguished by large and small letters. Vectors are written in bold font, e.g., $\boldsymbol{x}$ and $\boldsymbol{X}$. Matrices are written in monospaced font, e.g., $\mathtt{c}$ and $\mathtt{C}$.

The idea of a change of coordinates can be usefully generalized to that of a group action. Given a group $\mathcal{G}$ and a space $\mathcal{M}$, an action of $\mathcal{G}$ on the space associates with each group element $g \in \mathcal{G}$ a map g: $\mathcal{M} \rightarrow \mathcal{M}$ such that

$$id(x) = x \tag{1}$$

$$(g_1 \times g_2)(x) = (g_1(g_2(x)) \tag{2}$$

where $g_1, g_2 \in \mathcal{G}$, id is the identity element of the group, and $\times$ is the group composition function. Notice that this definition implies that a group action is reversible because the inverse of any group element also lies in the group. By far, the most commonly encountered group actions in computer vision are the actions of the Euclidean group, which represent rigid motions. There is an identity, which is given by not moving an object at all: two rigid motions compose to give a third rigid motion, and it is possible to undo a rigid motion by moving the object back to where it came from. An *invariant* of a group action is defined as follows:

Definition: An invariant $I(\boldsymbol{p})$ of a function $f(\boldsymbol{x}, \boldsymbol{p})$ subject to a group $\mathcal{G}$ of transformations acting on the coordinates $\boldsymbol{x}$ is transformed according to $I(\boldsymbol{P}) = I(\boldsymbol{p})h(g)$. Here, $g \in \mathcal{G}$ and $h(g)$ is a function only of the parameters of the transformation and does not depend on the coordinates $\boldsymbol{x}$ or on the parameters $\boldsymbol{p}$. $I(\boldsymbol{p})$ is a function only of the parameters $\boldsymbol{p}$.

In the cases we study, the effect of a group element is always given by a matrix $\mathtt{M}_g$, and $h(g)$ is a function of the determinant of that matrix $|\mathtt{M}_g|$. We always have $I(\boldsymbol{P}) = I(\boldsymbol{p})|\mathtt{M}_g|^w$. In this expression, w is referred to as the *weight* of the invariant. Invariants for which $w = 0$ are referred to as *scalar* invariants. In what follows, we concentrate on scalar invariants, and the term invariant should be understood to mean scalar invariant, except where the context clearly indicates otherwise.

An invariant is defined in the context of a particular transformation. We have concentrated much attention on the case of the plane projective group. This models the situation where a plane curve is subject to rigid motion in space and projected using perspective. In fact, there is more to the plane projective group than just rigid motion and perspective projection, and the subtle but important distinctions between the two situations are the subject of active research.

A number of invariants are exploited in vision at present with varied success. Kanantani [23] demonstrated the value of using invariants of camera rotation and invariant decompositions in computing such information as optical flow. Lamdan *et al.* [25] used an affine invariant representation of a net of points for matching (see Weiss [49] for a concise review of projective invariants).

A. Examples of Invariants

Example 1. Plane Translation: Any element ϵ of the one-dimensional translation group acts on the plane by the mapping $\{x', y'\} = \{x + \epsilon, y\}$. The y coordinate of any point is

TABLE I
THE MAXIMUM NUMBER OF DERIVATIVES REQUIRED FOR A SCALAR DIFFERENTIAL INVARIANT CONSISTING OF DERIVATIVES MEASURED AT A SINGLE POINT UNDER THREE GROUPS IMPORTANT IN VISION. (THIS ASSUMES INVARIANCE BOTH TO "GEOMETRIC" ACTIONS AND REPARAMETRIZATION).

Plane Euclidean group (3 d.o.f.)	Plane affine group (6 d.o.f.)	Plane projective group (8 d.o.f.)
2	5	7

invariant under the action of the group. Nothing can be inferred from the x coordinate of a point, however.

Example 2. Plane Rotation: The plane rotation group acts on the plane by the mapping $\boldsymbol{x} = \mathtt{R}\boldsymbol{X}$, where R is a 2-D rotation matrix. Any function of the distance from the origin to a point is a scalar invariant under the action of this group. Under the action of this group combined with the multiplicative group, we have $(x, y) :\rightarrow (\lambda(x\cos\theta + y\sin\theta),\ \lambda(-x\sin\theta + y\cos\theta))$,where the function $x^2 + y^2$ is an invariant of weight 2 because $x^2 + y^2 :\rightarrow \lambda^2(x^2 + y^2)$. In the second case, there is no scalar invariant, however.

Example 3. The Cross Ratio: Given four collinear points A, B, C, D, the expression $\frac{AC}{BC} \div \frac{AD}{BD}$, where AB denotes the linear distance from A to B, is well known to be invariant to projection. This expression is known as the *cross ratio* or *anharmonic ratio* of the four points. The cross ratio depends on the order in which the points are labeled. If the labels of the four points are permuted, a different value of the cross ratio results. Of the 24 different labeling possibilities, only six yield distinct values.

Example 4. Differential Invariants: Differential invariants are invariant functions of the position and derivatives of a curve at a single point. Differential invariants are clearly important in vision. Curvature, torsion, and Gaussian curvature (all differential invariants under Euclidean actions) have been widely applied. For example, a scalar projective differential invariant for plane curves has been known for a long time [49], [26]. However, this invariant is an extremely large and complex polynomial in the derivatives of the curve, and it is not known how useful it will be in practice. Table I shows the number of derivatives required for a differential invariant based on measurements at a single point for a plane curve under a range of group actions. These numbers assume that the invariant is required to be invariant to reparametrization as well as to projection, which is the normal case in vision.

Example 5. Differential Invariants Based at More than One Point: It is possible to construct differential invariants based on derivatives taken at more than one point [4], [48]. This has the advantage that fewer derivatives need be taken at each point and appears to be the most practical way to apply differential invariants. For example, given an open-curve segment, an invariant can be constructed for any point of the curve from that point, the tangent at that point, and the end points and their tangents. We explain this construction in Section II-B. Other projectively invariant differential structures, for example, double points, cusps, and flex points, can be exploited in a similar way to yield reference points.

Example 6. Projective Invariants for Systems of Lines or Points: A system of four coplanar lines that intersect in a single point,[1] is dual to a system of four collinear points. This system has the familiar cross ratio as its invariant.

For five general coplanar lines, there are two projective invariants:

$$I_1 = (I_{431}I_{521})/(I_{421}I_{531})$$

and

$$I_2 = (I_{421}I_{532})/(I_{432}I_{521}).$$

The lines are written in homogenous coordinates as $a_i x + b_i y + c_i z = 0$, and I_{ijk} is the determinant of the matrix $\{\boldsymbol{l}_i, \boldsymbol{l}_j, \boldsymbol{l}_k\}$, where $\boldsymbol{l}_i$ is $\{a_i, b_i, c_i\}^T$ (see [14]). Furthermore, because points and lines are dual in the projective plane, we have immediately that functions are also invariants for a system of five coplanar points, no three of which are collinear. These invariants have found important application in machine vision [4], [32] and photogrammetry [31].

Example 7. Projective Invariants for Pairs of Plane Conics: A plane conic can be written as $\boldsymbol{x}^t \mathtt{c} \boldsymbol{x} = 0$, for $\boldsymbol{x} = (x, y, 1)$ and a symmetric matrix c, which determines the conic, where we assume$|\mathtt{c}| = 1$. A pair of coplanar conics has two scalar invariants, which we will describe here. Given conics with matrices of coefficients $\mathtt{c}_1$ and $\mathtt{c}_2$, we define

$$I_{\mathtt{c}_1\mathtt{c}_2} = \mathrm{Trace}\left(\mathtt{c}_1^{-1}\mathtt{c}_2\right)$$
$$I_{\mathtt{c}_2\mathtt{c}_1} = \mathrm{Trace}\left(\mathtt{c}_2^{-1}\mathtt{c}_1\right).$$

Under the action $\boldsymbol{x} = \mathtt{T}\boldsymbol{X}$, $\mathtt{c}_1$ and $\mathtt{c}_2$ go to $\mathtt{C}_1 = \mathtt{T}^t\mathtt{c}_1\mathtt{T}$, and $\mathtt{C}_2 = \mathtt{T}^t\mathtt{c}_2\mathtt{T}$. In particular, using the cyclic properties of the trace, we find

$$\begin{aligned} I_{\mathtt{C}_1\mathtt{C}_2} &= \mathrm{Trace}(\mathtt{T}^{-1}\mathtt{c}_1^{-1}(\mathtt{T}^t)^{-1}\mathtt{T}^t\mathtt{c}_2\mathtt{T}) \\ &= \mathrm{Trace}\left(\mathtt{c}_1^{-1}\mathtt{c}_2\right) \\ &= I_{\mathtt{c}_1\mathtt{c}_2}. \end{aligned}$$

A similar derivation holds for $I_{\mathtt{c}_2\mathtt{c}_1}$. Note that $\mathtt{c}_1^{-1}\mathtt{c}_2$ transforms to $\mathtt{T}^{-1}\mathtt{c}_1^{-1}\mathtt{c}_2\mathtt{T}$, which is a similarity transformation, and therefore, its eigenvalues are preserved. This provides an alternative demonstration of invariance. Since scaling a matrix does not affect the conic curve it represents, to evaluate and interpret these invariants, we need to make some assumption to set the relative scale of the conic matrices.

Example 8. Homogenous Polynomials: The previous two examples can be generalized to more general algebraic sets. Consider the space of homogenous polynomials in n variables, $x_0, \cdots, x_{n-1}$. Write $\boldsymbol{x}$ for $\{x_0, \cdots, x_{n-1}\}$. The general linear group (all matrices T of nonzero determinant) acts on this space by taking a polynomial $p(\boldsymbol{x})$ to $P(\boldsymbol{x}) = p(\mathtt{T}\boldsymbol{x})$. Here, the coefficient of a monomial in $P(\boldsymbol{x})$ is determined by computing the coefficient of that monomial in the expansion of $p(\mathtt{T}\boldsymbol{x})$. The degree of a polynomial is preserved under this action; therefore, we can see it as an action on the homogenous polynomials of degree k. Furthermore, if we write $\boldsymbol{p}$ for the coefficients of p, we have for an invariant

[1] A useful example in vision is a set of parallel lines (that intersect in a single point at infinity).

Printed by: Black Bear Press Limited, Cambridge, United Kingdom

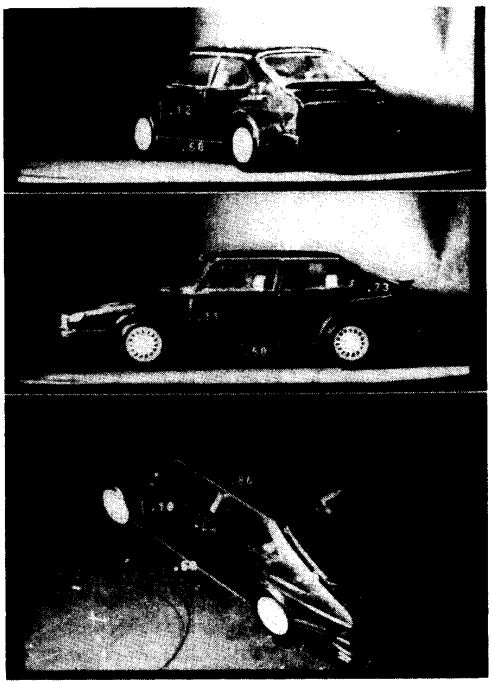


Fig. 1. Projectively invariant measurements of the distance between a number of different reference points on the side of a motorcar taken using the rear wheel of the car as a reference conic. These measurements remain stable over three different views, where the corresponding measurements in the image are significantly foreshortened. Small changes in the values measured may result from the fact that features and wheel are not precisely coplanar.

$I, I(\boldsymbol{P}) = I(\boldsymbol{p})|\mathtt{T}|^w$, where $\mathtt{T}$ is the transformation matrix, and $|\mathtt{T}|$ indicates the determinant of $\mathtt{T}$. Invariants of this action formed a major research topic of nineteenth-century mathematics; an introduction can be found in, for example, [18]. A modern treatment of some of this work is given by [44] or by [10].

Example 9. Projectively Invariant Measurements: If there is a distinguished conic curve in the plane (say, $\mathtt{c}$), then for two points $\boldsymbol{x}_1, \boldsymbol{x}_2$ that do not lie on the conic, the function

$$\frac{(\boldsymbol{x}_1^T \mathtt{c} \boldsymbol{x}_2)^2}{(\boldsymbol{x}_1^T \mathtt{c} \boldsymbol{x}_1)(\boldsymbol{x}_2^T \mathtt{c} \boldsymbol{x}_2)} \tag{3}$$

is independent of the frame in which the points and the conic are measured. In turn, this can be used to define a projectively invariant metric (see [44]). Fig. 1 shows projectively invariant measurements of the distance between a number of different reference points on the side of a motorcar taken using the wheel of the car as a reference conic. These measurements remain stable over three different views where the corresponding measurements in the image are significantly foreshortened.

Example 10. Permutation Invariants: Consider a function depending on several arguments, for example, one of the projective invariants of a pair of conics. Although it is possible to observe a pair of conics, it is not always possible to tell in which order they should appear as arguments to the function. In fact, permuting the arguments represents the action of a finite group. To avoid having to choose which order the conics should appear as arguments, we can form functions that are invariant under this action—these are known as symmetric functions. In this example, there are two such invariants. Write $f(\mathtt{c}_1, \mathtt{c}_2) = I_{\mathtt{c}_1\mathtt{c}_2}$. Then, the symmetric functions of these invariants are

$$S_1(\mathtt{c}_1, \mathtt{c}_2) = f(\mathtt{c}_1, \mathtt{c}_2) + f(\mathtt{c}_2, \mathtt{c}_1)$$
$$S_2(\mathtt{c}_1, \mathtt{c}_2) = f(\mathtt{c}_1, \mathtt{c}_2) f(\mathtt{c}_2, \mathtt{c}_1).$$

It is easy to see that these functions are unaffected by permutations of their arguments. These permutation invariants make it possible to avoid some kinds of correspondence problems. They are discussed further in [50].

B. Interpreting Invariants

It is often hard to interpret projective invariants, particularly when they are constructed using algebraic methods. Invariants can often be related to a cross ratio resulting from a projectively invariant construction. Another method of interpretation is to transform to a distinguished frame and measure metric properties within this frame. These geometric approaches yield insight into the properties of invariants. In this section, we display interpretations for a number of the examples given above.

1) The Invariants of Five Points: Two constructions can be used to demonstrate the projective invariants of five points $\boldsymbol{x}_1 \cdots \boldsymbol{x}_5$ in the plane. For the first construction, consider the lines from $\boldsymbol{x}_1$ to each point. This yields a system of four lines passing through a single point, which is dual to a system of four collinear points, and the cross ratio of these lines is an invariant. The second invariant is obtained by constructing the lines from $\boldsymbol{x}_2$ to each point and taking the cross ratio of these lines. Unfortunately, a wide range of invariants can be obtained in this way, and the construction does not make it obvious that only two of these invariants are functionally independent.

The second construction makes the functional independence more obvious. Since the image of four known points uniquely determines a projective mapping, we can associate a distinguished canonical frame with a system of five points by mapping $\boldsymbol{x}_1 \cdots \boldsymbol{x}_4$ to the unit square. The coordinates of the fifth point *in this frame* are projective invariants because the frame is uniquely determined by the other four points.

2) The Invariants of Two Coplanar Conics: The projective invariants of a pair of coplanar conics can be demonstrated in a number of ways. First, one can observe that for two conics $\mathtt{C}_1$ and $\mathtt{C}_2$, the term $\mathtt{C}_1^{-1}\mathtt{C}_2$ undergoes a similarity transformation, and therefore, its eigenvalues are preserved. A geometric interpretation of this construction appears in [24].

We are indebted to Maybank [30] for a second interpretation. Two plane conics uniquely determine four points of intersection (these points are often complex, but this need not worry us). Choose any point on one of the conics that is not an intersection point, and construct the four lines from the intersection to this point. The cross ratio of these four lines is clearly a projective invariant. Chasles theorem states that this cross ratio is independent of the base point chosen on the conic, and therefore, this cross ratio is a unique invariant. We can now construct a second invariant by choosing an arbitrary base point on the second conic.

3) A Projective Differential Invariant: We can construct a projective differential invariant for any given point on an arc. First, label the end points b_1, b_2, and intersect the tangents at these end points to obtain a third point b_3. The system consisting of these three points and the given point can be mapped uniquely to the unit square. Under the action of this map, the tangent to the given point maps to a line passing through one vertex of the unit square. The orientation of this line is then an invariant. A wide range of similar invariants can be constructed.

C. Constructing New Invariants

In order to apply invariants comprehensively in vision, it is necessary to compute the invariants of the action of various groups on a variety of spaces. For the groups we deal with, it is the case that there are a limited number of "primitive" invariants and that all other invariants can be formed from these primitive invariants. For example, plane curvature is a differential invariant under the action of the plane Euclidean group. We could square the curvature, take its logarithm, or form some complicated polynomial in curvature. These would all be invariant, but we would learn nothing new from these constructions. The key question is then how to obtain a functionally independent set of these primitive invariants.

One method recommends itself immediately. If we write out the effect of a group action, we obtain a system of polynomials in the group parameters and the point on which the group acts. We could then eliminate the group parameters between these expressions to obtain invariants. Our experience has been that in practice, eliminating group parameters is unmanageable except for very simple situations, and therefore, more sophisticated methods are required.

A number of cases are fairly easily dealt with. Weyl [50] provides complete tables of invariants for systems of vectors under the action of the rotation group, the affine group, and the general linear group. Weyl discusses the procedure for using symmetric functions to obtain invariants for the permutation groups. These invariants can be looked up.

The more interesting case is that in *Example 8.* Polynomials represent curves, and therefore, invariants of a system of polynomials give invariants of a set of curves. In this section, we give a brief discussion of two important techniques for constructing these invariants: the infinitesimal method and the symbolic method. Each method has its attractions.

The infinitesimal method gives invariants as the solutions to a system of differential equations. It has the attraction that it works both for algebraic invariants of curves and surfaces and for differential invariants. This approach has the disadvantage that solving these equations cannot, at present, be done wholly by a symbolic algebra package but requires some human assistance. We have found that for some group actions, particularly the action of the 3-D rotation group on algebraic surfaces, solving the equations is difficult.

The symbolic method is, in our experience, easy to use for algebraic invariants of curves and surfaces, but we are not aware of any way of using it to obtain differential invariants. The major attraction of the symbolic method is that it seems possible to write a straightforward program that will generate all primitive invariants for a given situation without requiring human assistance.

TABLE II
THE MINIMUM NUMBER OF FUNCTIONALLY INDEPENDENT SCALAR INVARIANTS FOR PLANE ALGEBRAIC CURVES UNDER A VARIETY OF GROUPS IMPORTANT IN VISION. (BRIEF DEFINITIONS OF EACH GROUP APPEAR IN THE TEXT.)

Curve (no. of d.o.f.)	Plane Euclidean group (3 d.o.f.)	Plane affine group (6 d.o.f.)	Plane projective group (8 d.o.f.)
conic (5)	2	0	0
cubic (9)	6	3	1
quartic (14)	11	8	6
2 coplanar conics (10)	7	4	2
five lines (10)	7	4	2
a conic and two lines (9)	6	3	1

This section focuses mainly on three groups:

- The plane Euclidean group consists of translations, rotations, and reflections of the plane and is sometimes called the group of plane rigid motions.
- The plane affine group consists of plane translations and all two by two matrices with nonzero determinant. This group in homogenous coordinates is given by the group of all three by three matrices of nonzero determinant, whose third row is $(0, 0, 1)$.
- The plane projective group is given in homogenous coordinates by the set of all three by three matrices with nonzero determinant.

The infinitesimal method is, however, applicable to all Lie groups.[2] The symbolic method works for a wide range of Lie groups (see [47]).

1) The Infinitesimal Method:

It is important to know how many functionally independent primitive invariants are available in a given situation. In this section, we use ideas from the theory of Lie groups to show that given a group of dimension m acting on a geometric structure with n degrees of freedom, there are at least $n - m$ functionally independent primitive invariants. This *counting argument* can also be used to predict the number of derivatives required for a differential invariant and provides the results of Tables I and II. This approach yields one way of constructing invariants as well.

We exploit the notion of an *orbit.* The orbit of a group action through a point is given by taking every element of the group and applying it to that point. For example, the orbits of the one-dimensional translation group acting on the plane are lines; the orbits of the plane rotation group are circles centered at the origin. Because the action of the group is invertible, orbits cannot intersect. In this section, we also exploit the idea of the *dimension* of a group. This is meaningful only for continuous groups, for example, the rotation or translation groups, which can be expressed by smooth functions of a

[2] Loosely speaking, a Lie group is a group that has a smooth parametrization, and these parameters determine the group. An important feature is the fact that group elements in a neighborhood of any point can be obtained from tangent vectors at that point by a process called exponentiation.

Printed by: Black Bear Press Limited, Cambridge, United Kingdom

number of parameters. The dimension of the group is then the smallest number of parameters that can be used; for example, the dimension of the three-dimensional rotation group is three.

Two properties of orbits are particularly important:

- The dimension of an orbit cannot be greater than that of the group.
- Invariants are constant on orbits.

These points make it possible to estimate the number of primitive scalar invariants. If the space on which the group acts has dimension n and the group has dimension m, the orbits divide the space up into nonintersecting "slices" of dimension $l \leq m$, where l is the dimension of the highest dimensional orbit. The primitive scalar invariants are, at least locally, coordinates for the slices; therefore, a given slice can be specified by the values of the primitive scalar invariants. To specify an l-dimensional subspace of n-dimensional space, one must generically fix $n - l$ coordinates. Thus, there are $n - l$ primitive scalar invariants. Although we show below how to compute l, we normally assume that the orbits have dimension m and that there are $n - m$ invariants. The counting argument yields the results of Table II. For certain geometries, every orbit will have dimension less than m, and there are more than $n - m$ invariants. This is known as an *isotropy*. The best known example is the case of two lines $\boldsymbol{l}_1, \boldsymbol{l}_2$ and two points $\boldsymbol{x}_1, \boldsymbol{x}_2$ not lying on the lines (both points and lines are expressed in homogenous coordinates). This geometry has eight degrees of freedom and one projective invariant $(\boldsymbol{l}_1^T \boldsymbol{x}_1)(\boldsymbol{l}_2^T \boldsymbol{x}_2)/(\boldsymbol{l}_1^T \boldsymbol{x}_2)(\boldsymbol{l}_2^T \boldsymbol{x}_1)$.

This counting argument extends to differential invariants. In this case, we need invariance to the effects of reparametrization as well. Assume the group of geometric actions (actual movements of the curve rather than reparametrizations) has m degrees of freedom, and the invariant involves d derivatives. Then, since the derivatives of a curve are completely independent, we are considering for a plane curve $2(d + 1)$ degrees of freedom. In the dth derivative of the curve, we encounter the dth derivative of the reparametrization. Thus, we need consider only d reparametrization degrees of freedom, and therefore, to have an invariant, we must have $2(d + 1) > l + d$, or $d > l - 2$. These points are illustrated in the worked example below. Although it is possible to compute l, in general, we assume that the orbits are m dimensional, and therefore, $d > m - 2$. This argument yields the results of Table I and can be applied to the case of differential invariants based at more than one point as well. In general, fewer derivatives are required as more points are considered. In this case, some care is required because orbits occasionally have dimension less than m.

This view of invariants as functions that are constant on orbits provides a conceptually simple procedure for constructing them. The gradient of any function that is constant on an orbit must be normal to that orbit. Thus, for a scalar invariant Φ, if the vector fields $\boldsymbol{V}_i(x)$ span the tangent space to the orbit passing through x for all x in the parameter space, then

$$\boldsymbol{V_i} \cdot \nabla \Phi = 0, \quad \forall\, i$$

and the scalar invariants can be obtained by solving these equations.

Vector fields that span the tangent space to the orbit passing through x for all x in the space on which a group acts are well known in the theory of Lie groups. These fields are known as the infinitesimal generators of the group's action. To find the infinitesimal generators at a point, we compute the effect of an infinitesimally small group action.[3] The rank of the system of generators gives the dimension of the largest orbit and can be used to predict isotropies. We have, therefore, a mechanical process for constructing invariants of a connected Lie group's action:

- Construct the infinitesimal generators of the group's action $\boldsymbol{V}_i$.
- The invariants are the solutions of the set of partial differential equations $\boldsymbol{V}_i \cdot \nabla \Phi = 0, \forall\, i$.

This process constructs a function that is locally invariant; this means that it will be constant on connected components. For a connected group, the function is then a scalar invariant. If the group is not connected, like the general linear group,[4] it is possible for the function to be constant on the connected component but to have different values on distinct connected component. We deal only with connected groups and can largely ignore these difficulties.

Our experience is that the appropriate technique for solving these systems is to solve one equation and express the remaining equations in terms of the solutions of the first. This method for solving such systems is discussed briefly in [36] and in more detail in [2]. This procedure of repeatedly changing variables will always find a solution if the system is integrable[5] [2]. A system of PDE's for functions constant on an orbit found using the procedure detailed is integrable.[6]

D. The Symbolic Method

An elegant process for finding invariants of forms under the action of a range of groups by using simple symbolic manipulations was devised by a number of nineteenth century mathematicians. This section briefly sketches this process, which is not widely known, and is uniquely well adapted to modern techniques of symbolic computation. More detailed discussions can be found in [1], [9], [50], [43], for example.

A binary conic can be written as $a_{00}x_0^2 + a_{01}x_0x_1 + a_{11}x_1^2$. When the relation $a_{00}a_{11} = a_{01}^2$ is satisfied, it is possible to write the conic as the product of two linear factors. One may, however, write any conic as the square of a linear factor *for symbolic purposes only,* as long as one does not exploit this relation. To do this, it is necessary to insist that only quadratic monomials in symbols can be interpreted.

Thus, write this form as $(\alpha_0 x_0 + \alpha_1 x_1)^2$. The terms α_0, α_1 are phantom symbols with no meaning of their own. Conven-

[3] This can be done by taking a set of vectors that span the group's tangent space (also known as its *Lie algebra*) and for each $\boldsymbol{V}_i$ in this set, exponentiating $\epsilon \boldsymbol{V}_i$, and computing the action of the resulting group element. We then differentiate the result of this action by ϵ and set ϵ to zero. The details of this process are explained in [36].

[4] The general linear group is not connected because taking the determinant gives a smooth map from the group to a set that is not connected. Thus, there are two connected components: the matrices with positive determinant, and the matrices with negative determinant.

[5] i.e. the commutator of two generators is a linear combination of generators.

[6] Apply Frobenius' theorem [5].

tionally, these phantom terms are written with Greek letters. The only collections of symbols that can be interpreted are

$$\alpha_0^2 = a_{00}, \qquad \alpha_0\alpha_1 = \frac{a_{01}}{2}, \qquad \alpha_1^2 = a_{11}.$$

Since, for example, the product $\alpha_0^2\alpha_1^2$ is ambiguous (it could be interpreted as either $a_{00}a_{11}$ or as a_{01}^2, which are not the same thing), it cannot be interpreted. The terms α_i are phantoms; they are only meaningful when they appear as the monomials shown. To make it possible to express such terms as a_{01}^2 unambiguously, one must admit different symbolic representations of the form. The form can then be written as $(\alpha_0 x_0 + \alpha_1 x_1)^2 = (\beta_0 x_0 + \beta_1 x_1)^2 = (\gamma_0 x_0 + \gamma_1 x_1)^2 = \cdots$, introducing as many of these symbolic roots as is convenient. $a_{00}a_{11}$ can then be represented as $\alpha_0^2\beta_1^2$.

The power of the symbolic method lies in the theorem that the invariants under the action of the general linear group of a m-ary n-ic form[7] can be expressed as monomials in *brackets.* In the case of a binary form, a typical bracket is $(\alpha\beta)$, which is interpreted as meaning

$$\det\begin{Bmatrix} \alpha_0 & \alpha_1 \\ \beta_0 & \beta_1 \end{Bmatrix}.$$

For an m-ary form, the bracket has m symbols. In this case, the bracket is interpreted as the appropriate m by m determinant. For a ternary form,[8] the bracket $(\alpha\beta\gamma)$ is thus a three by three determinant. At least the primitive invariants can be written as a product of brackets, satisfying the constraint that each symbol appears exactly n times; otherwise, there would be phantom symbols in groups that could not be interpreted. Products of brackets are subject to a system of relations. For example, $(\alpha\alpha) = 0$ and $(\alpha\beta)^n = 0$ for n odd because when the phantom symbols are replaced with meaningful ones, $(\alpha\beta)^n$ has the same meaning as $(\beta\alpha)^n$; yet, $(\alpha\beta)^n = -(\beta\alpha)^n$.

To find the invariants of a m-ary n-ic form, in effect, one writes out the monomials satisfying these constraints and prunes the set using a system of relations to obtain the monomials that are independent. In fact, since the number of independent invariants is known and finite, it is possible to obtain a complete set of invariants rather quickly because one need not examine all possibilities. This process yields nonscalar invariants, whose weight is given by the number of parenthesized terms.

For example, the invariant of a binary conic[9] can be written as $(\alpha\beta)^2$:

$$\begin{aligned}(\alpha\beta)^2 &= \alpha_0^2\beta_1^2 + \alpha_1^2\beta_0^2 - 2\alpha_0\alpha_1\beta_0\beta_1 \\ &= -(1/2)(a_{01}^2 - 4a_{00}a_{11}).\end{aligned}$$

This is $-(1/2)$ times the *discriminant,* which is a well known invariant of quadratic equations. As a second example, a ternary cubic form $a_{000}x_0^3 + a_{001}x_0^2x_1 + a_{002}x_0^2x_2 + a_{011}x_0x_1^2 + a_{012}x_0x_1x_2 + a_{022}x_0x_2^2 + a_{111}x_1^3 + a_{112}x_1^2x_2 + a_{122}x_1x_2^2 + a_{222}x_2^3$ is written as $(\alpha_0x_0 + \alpha_1x_1 + \alpha_2x_2)^3 = (\beta_0x_0 + \beta_1x_1 + \beta_2x_2)^3 = (\gamma_0x_0 + \gamma_1x_1 + \gamma_2x_2)^3 = \cdots,$, where symbols are interpreted as

$$\alpha_0^3 = a_{000}, \qquad \alpha_0^2\alpha_1 = \frac{1}{3}a_{001}, \text{ etc.}$$

The invariants of a ternary cubic form (i.e., a cubic curve) can be written as

$$S = (\alpha\beta\gamma)(\alpha\beta\delta)(\alpha\gamma\delta)(\beta\gamma\delta)$$
$$T = (\alpha\beta\gamma)(\alpha\beta\delta)(\alpha\gamma\epsilon)(\beta\gamma\phi)(\delta\epsilon\phi)^2$$

S and T are familiar from the nineteenth century literature on projective plane curves (e.g., [42], where they appear written out in full[10]). S is an invariant of weight four, and T is an invariant of weight six. There is one scalar invariant S^3/T^2. These invariants are naturally a great deal more complicated when written as polynomials in the coefficients of the form.

The bracket monomials for a single form represent invariants under the action of the general linear group. As Weyl [50] shows, invariants under the action of other groups can be obtained by adjoining further forms and then specializing the values of these forms. In effect, this specialization confines the group action to actions that leave the extra forms unchanged. As an example, we consider the action of the plane affine group on a plane curve. The projective plane is often thought of as the affine plane together with a line at infinity; the affine plane can be seen as the projective plane less a line at infinity. The plane affine group is then the set of plane projections that take the line at infinity to itself. The invariants of this action are obtained by taking the invariants of a system consisting of the original form $p(x_0, x_1, x_2)$ and a linear form $l = l_0x_0 + l_1x_1 + l_2x_2$, and then setting $l_0 = 0$, and $l_1 = 0$, and $l_2 = 1$.

As a second example, the action of the rotation group on a degree n surface in three-dimensional space is given by a system of three forms: a homogenous 4-ary form expressing the surface in projective three space, a linear form whose value is specialized as in the first example because we want an affine subset of that space, and a conic (known as the absolute conic, which is familiar in the vision literature from the work of Maybank and Faugeras), whose value is specialized to ensure that the affine actions we obtain are only the orthogonal affine actions and, hence, represent the rotation group. Further details are beyond the scope of this discussion; the reader is referred to Weyl [50], particularly pp. 255–258 and to Turnbull [47].

The symbolic method largely fell out of use because it was seen as merely a computational trick. However, our experience has been that as a computational trick for generating invariants, it is extremely well adapted to modern symbolic computation with surprisingly little effort required.

III. Invariants in Model-Based Vision: Recognizing Curved Planar Objects Using Algebraic Invariants

This and the following three sections treat different aspects of the application of invariant theory to model-based vision. In this section, we demonstrate recognition of plane objects using

[7] A homogenous polynomial of degree n in m variables.

[8] i.e. a projective plane curve in homogenous coordinates.

[9] i.e. a homogenous quadratic in two variables.

[10] Some confusion can be caused by the fact that different authors use different conventions about constant multipliers.

Printed by : Black Bear Press Limited, Cambridge, United Kingdom

TABLE III
THE JOINT SCALAR INVARIANTS COMPUTED FOR THE INDICATED PAIRS OF CONICS FOR THE FOUR DIFFERENT IMAGES OF A COMPUTER TAPE FROM DIFFERENT POSITIONS AND ANGLES, SHOWN IN FIG. 1. (NOTE THAT THE JOINT SCALAR INVARIANTS FOR THE COPLANAR CONICS a, b FOR THE FOUR IMAGES ARE EFFECTIVELY CONSTANT. FURTHERMORE, THE VALUES OF THE INVARIANTS FOR DIFFERENT PAIRS OF CONICS ARE DIFFERENT.)

Conics	First joint invariant	Second joint invariant
Conics a and b from Fig.1(a)	3.419	3.546
Conics a and b from Fig.1(b)	3.418	3.543
Conics a and b from Fig.1(c)	3.414	3.538
Conics a and b from Fig.1(d)	3.407	3.528
Conics b and c from Fig.1(a)	3.022	3.021
Conics b and c from Fig.1(b)	3.023	3.021

the conic invariants of Example 7 above. In the following three sections, we discuss the range of Euclidean invariant descriptors for three-dimensional objects, which are obtained by computing the pose of a number of planes. We have used the term object to refer to the object modeled, the term model to refer to the object model (which will be a pair of coplanar conics), and the term descriptor to refer to the viewpoint-invariant description of shape that we exploit.

A. Objects Containing Algebraic Curves

First, we consider objects that consist of a pair of coplanar conics. These objects are modeled by their joint projective invariants (Section II, *Example 7*), yielding a representation that is invariant to Euclidean motion and perspective, i.e., the joint invariants calculated for image curves will have the same values as those calculated from object curves, whatever the object pose and the camera parameters. In practice, these descriptors are stable and have sufficient dynamic range to discriminate between distinct pairs of conics (see Table III, and Fig. 2).

We implemented a model-based vision system that recognizes objects of this type using the following signal flow:

- **Edge Detection:** This system uses a local implementation of Canny's [7] edge detector to mark image edges.
- **Curve Extraction:** The edgels found are then chained into a system of curves; sharp corners are broken, straight lines and image curves that are too short for conic fitting are rejected. Single pixel drop-outs in edges are repaired using a simple following algorithm.
- **Conic Fitting:** Edgel chains are passed on to an exact conic fitter, which works by computing the kernel of the scatter matrix for that chain (see, for example, [13]). Curves that are judged not to be exact conics (by a check on the eigenvalues of the scatter matrix) are not considered further.
- **Computing Shape Descriptions:** For all pairs of exact conics, the joint scalar invariants are formed.

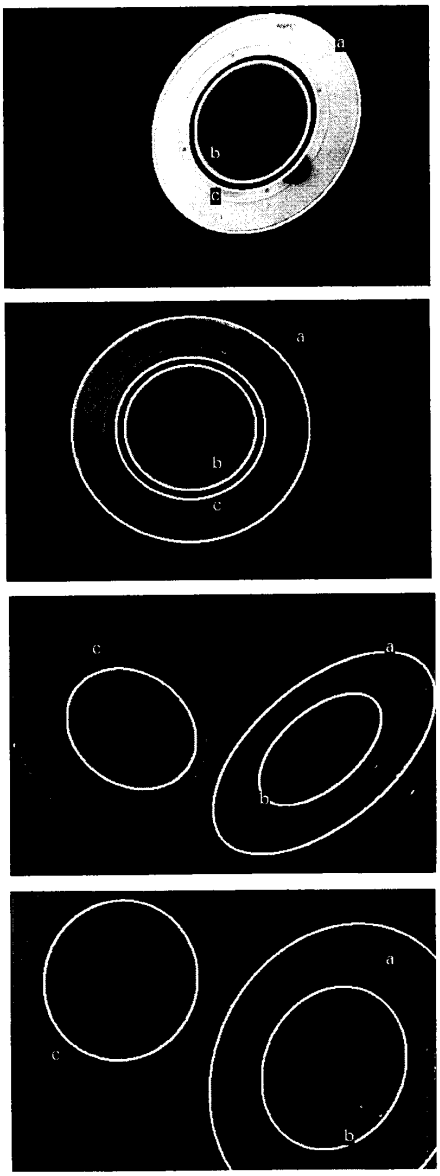


Fig. 2. Four images of a computer tape with fitted conics in overlay. In these images, the conics have been drawn three pixels thick to make them visible. These conics were used to obtain the joint scalar invariants shown in Table III.

- **Model Matching:** A hash table is then used to pair values of joint scalar invariants with the descriptions of the labels.

Our present system works on a model base of 15 objects. Some typical examples of this system in use are demonstrated in Fig. 3.

Models are acquired by imaging an object, on its own, under favorable lighting conditions and from an arbitrary viewpoint. As long as the object is the only object visible in the image

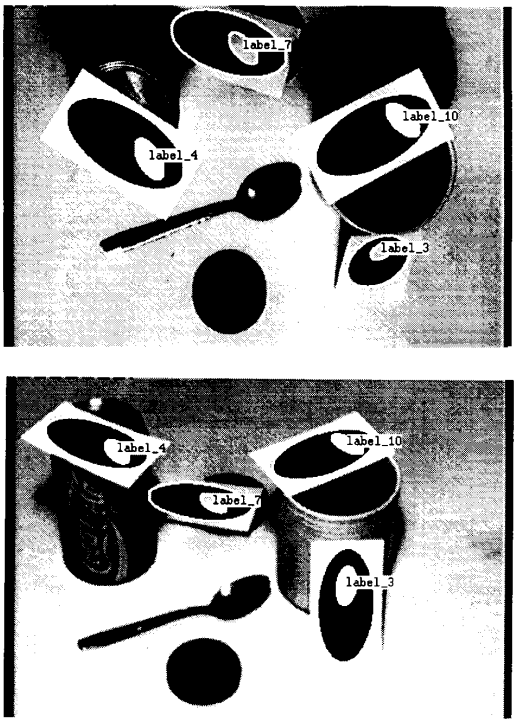


Fig. 3. Two views of labels consisting of pairs of ellipses on card in cluttered scenes. These labels are conceptually similar to Nielsen's [34] labels but use different invariants. The labels come from a model base of 15 labels. These labels are positioned at different angles and distances to the camera and are recognized by their invariant descriptors. Model instances are found by detecting edges, pruning edges that are too short or are obviously not conics, fitting conics to all edges, and computing invariant descriptors for every pair of fitted conics. These descriptors are then used to index into the model base. The labels are correctly identified in each case, and their corresponding number is superimposed on the image.

and its conics are sharply defined, the recognition system can be used to compute the object shape descriptors directly from these images. A library entry consists of a pair of projective invariant values, each with a tolerance and an associated label number. The ease with which models can be acquired is a significant advantage of using projective invariants as shape descriptors.

B. Representing General Curves by Conics

For more general curves, we can still exploit algebraic invariants if the general curve is *represented* by an algebraic curve. However, the representation must have the following crucial frame independence property:

Given an observation of a data set in a transformed frame, the representation computed for this set is exactly the original representation transformed according to the change of frame.

We refer to the process of constructing a representation with this property as invariant fitting.

1) Affine Invariant Fitting: It is possible to construct a representation with this property *for the affine group* if the fitting metric used is *algebraic distance.* The algebraic distance of a set of points $\boldsymbol{x}_i$ from a curve $Q(\boldsymbol{x}) = 0$ is given as

$$\sum_i (Q(\boldsymbol{x}_i))^2.$$

Unfortunately, algebraic distance as it stands does not give a useful metric for *curves*[11] because for any polynomial Q, the curve $Q(\boldsymbol{x}) = 0$ is the same as the curve $kQ(\boldsymbol{x}) = 0$, for $k \neq 0$. We must therefore fix a convention for the scale of the curve, which is a process known as *normalization.* In general, we normalize a curve with coefficients $\boldsymbol{p}$ by $N(\boldsymbol{p}) = 1$, where N is an algebraic invariant of $\boldsymbol{p}$. As the following theorem, proved in [13], shows, this guarantees that the representation has the desired properties:

Theorem 1: Let $I(\boldsymbol{p})$ be an invariant of the polynomial form $Q(\boldsymbol{x}, \boldsymbol{p})$ under the affine group of linear transformations $\mathcal{G}$. Assume I is homogenous of degree n with weight w. Let $\langle \boldsymbol{p} \rangle$ be the parameter vector determined by minimizing $\sum_i Q^2(\boldsymbol{x}_i, \boldsymbol{p})$ over a set of points $\boldsymbol{x}_i$ subject to the constraint $N(\boldsymbol{p}) = I(\boldsymbol{p}) = constant$. If the point set is transformed under the affine group (or some subgroup) $\mathcal{G}$, i.e., $\boldsymbol{x} = \mathtt{T}_g \boldsymbol{X}$, let $\mathcal{T}_g$ be the corresponding transformation matrix *for the coefficients* $\boldsymbol{p}$. The coefficients of the polynomial fitted to the point set in the new frame are given by $\langle \boldsymbol{P} \rangle$. Assume that n is odd or that w is even (or both). Under these conditions, we have

$$\langle \boldsymbol{p} \rangle = k_g \cdot \mathcal{T}_g \langle \boldsymbol{P} \rangle$$

where k_g is a scalar depending on $g \in \mathcal{G}$.

The theorem applies to algebraic curves of higher degree as well as to conics. It is possible to show that a solution to this fitting problem exists. We do not yet know under what circumstances it is unique. The fitting problem presents interesting numerical difficulties; currently, it is solved using a subdivision technique [13]. We are currently investigating using polynomial continuation (see, for example [33]) for solving this problem.

Note that the theorem applies only to *collections of points* that are in affine correspondence. A version that applies to curves can be obtained by noting that the algebraic distance must now be integrated with respect to some affine invariant parameter. In our experience, it is possible to obtain acceptable fits in practice by simply summing the algebraic distance over the points on the curve.

2) Projectively Invariant Fitting: The approach used for affine invariant fitting requires some modification for the projective case; a function on the projective plane is given by the ratio of two homogenous polynomials of the same degree with the result that algebraic distance is not well defined. This is because scaling any point on the projective plane gives an equivalent representation of the same point. If we measure the algebraic distance of a point $(x, y, 1)$ from a curve in one frame, the point in the new frame that has the same algebraic distance will be some scaling of $(X, Y, 1)$. There are three ways to proceed:

- Assume that a distinguished conic, or three distinguished lines, are available, and use these to define a projec-

[11] It is perfectly good for polynomials.

Printed by: Black Bear Press Limited, Cambridge, United Kingdom

tive version of algebraic distance; in the case of the distinguished conic $\boldsymbol{d}$, this would be given by

$$\sum_i \frac{Q^2(\boldsymbol{x}_i, \boldsymbol{p})}{Q^2(\boldsymbol{x}_i, \boldsymbol{d})}$$

The theorem would then go through.

- Notice that because objects and the viewing area of the camera are small and because we are dealing with perspectivities, the errors incurred by this scaling problem tend to be small. Accept these small errors and proceed.
- Transform the curve to a distinguished, canonical frame, which is determined by intrinsic geometrical properties of the curve itself. For example, inflections, cusps, or, more practically, points of contact of bitangent lines determine a distinguished frame in which fitting can proceed. This approach is more general in that a wide range of invariant measurements can be made in the distinguished frame.

C. Recognizing Planar Objects

We can now extend the model-based vision system described in Section III-A to cope with more general objects. The system is extended in the following respects:

- Curves that are not conics are no longer discarded. We discard lines, however, because fitting a conic to a line is unfruitful.
- We interpose a fitting stage between the curve extraction stage and the stage that computes invariants. In this fitting procedure, we compute an affine invariant fit and assume that the small errors are tolerable.

A block diagram appears in Fig. 4. An implementation of this system worked on a library of five gaskets in cluttered scenes. Several examples of this system's output appear in Fig. 5. It is sensitive to occlusion, however, because the relationship between the conic fitted to a curve and that fitted to a piece of the original curve is not well understood ([40] discusses some of the noise issues that appear in fitting conics to small number of data points).

This system is different from earlier model-based vision systems in a number of important ways:

- Curves are not segmented into polygonal approximations.
- It is unnecessary to search the model base. The invariant descriptors index a model directly.
- At this stage, no pose information is involved in recognition. It is therefore possible to *identify* an object without knowing *where* it is. Section IV shows that, given that the model has been identified, pose recovery is simple.
- It is straightforward to acquire models for this system because the descriptors used are projectively invariant. The invariant descriptors measured in any view of the object have the same value so that they can be calculated directly from conics fitted to curves extracted from an *image* of the object. The object can be imaged from any viewpoint, and no correction for aspect ratio or camera parameters is required. The model library then consists of the pair of invariants and error thresholds for each object.

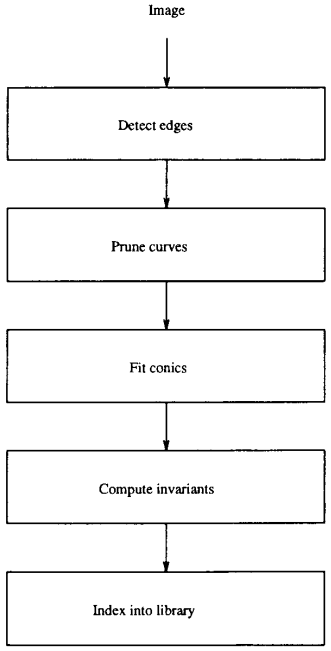


Fig. 4. The Block diagram of a model-based vision system for objects consisting of pairs of general coplanar curves.

IV. Pose Determination: A Single Plane in Space

A single plane in space is the simplest of three-dimensional objects. In this section, we show that once an object has been positively identified, the extra constraints offered by its known identity can be exploited to determine the transformation parameters between the object frame and the camera frame.

Invariant fitting allows a pair of coplanar curves to be modeled by a pair of coplanar conics. By construction, the modeling conics undergo the same projective distortion that the original curves do. Consequently, the problem of pose determination is equivalent to the following:

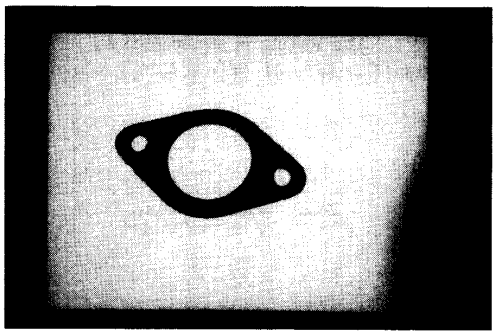
> Given a known pair of conics on the world plane and their corresponding conics in the image, determine the transformation between the two planes.

Previous methods for pose determination have concentrated on sets of points and lines [17], [45], [46]. However, apart from consideration of circles [8], [29], very little attention has been paid to pose recovery from plane curves. The novel aspects of the method described below lie first in the use of equiform invariants to recover pose for algebraic curves and second in recovering pose for arbitrary plane curves using invariant algebraic curve representation.

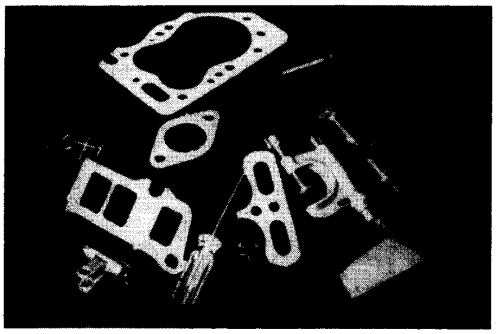
A. Back Projection of a Conic Pair

A perspective projection between the image plane and the object plane is determined by six parameters[12] (three transla-

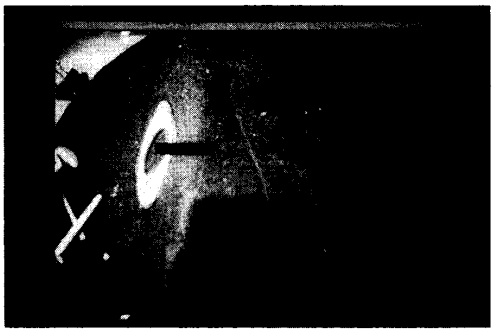
[12] Assuming a fixed camera with known calibration.



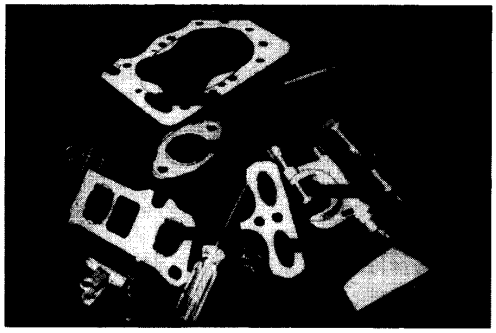
(a)



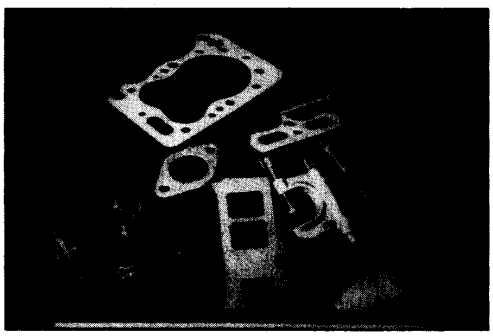
(e)



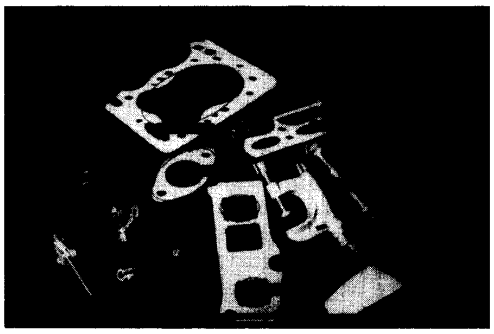
(b)



(f)



(c)



(d)

Fig. 5. Six examples of the model-based vision system working automatically on real objects. Model instances are found by detecting edges, fitting conics to all edges, and computing invariant descriptors for every pair of fitted conics. These descriptors are then used to index into the model base. (a) shows a gasket viewed approximately frontally. Models of the gaskets were made using images like this. (b) shows the gasket shown in (a), seen in a different view. The gasket was recognized and labeled correctly, despite the large change in viewing angle. (c) shows a cluttered scene containing four gaskets, and (d) shows the gaskets correctly recognized and labeled. (e) shows another cluttered scene also containing four gaskets, and (f) shows the gaskets labeled in that scene. Note that the micrometer incorrectly recognized as gasket 3 can easily be dealt with by verifying the model using backprojection of the original *image* curves used to construct the invariant description.

tion and three rotation) that give the pose of the object relative to the camera. Each conic has five independent parameters; therefore, the solution is overdetermined (ten constraints on six unknowns). We have developed and assessed a number of methods for pose recovery, including

1) Solving for the T matrix ($\boldsymbol{x} = T \cdot \boldsymbol{X}$) directly and then decomposing the matrix to give the six pose parameters. A projective mapping can be defined by four world points and their images [20]. Taking four world and image points sets up a system of eight linear simultaneous equations in the elements of the T matrix, which are solved by Gaussian elimination. We obtain four points from the conic pair by extracting intersections between the curves or bitangent lines. These features are both preserved under projection and relatively stable to numerical errors and noise. In general, the points are complex but produce the correct, real T matrix.

Printed by ... Limited, Cambridge, United Kingdom

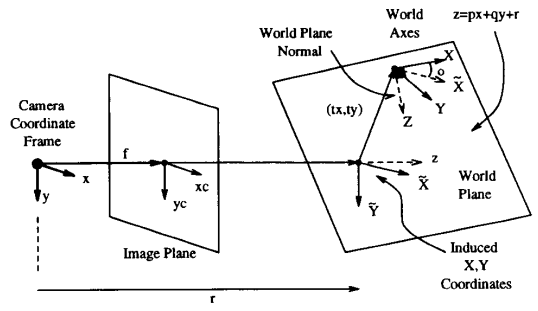


Fig. 6. Camera and world plane geometry. The camera and induced x axes lie in the same plane. The induced $\tilde{X}\tilde{Y}$ plane and the world XY plane both lie in the plane $z = px + qy + r$. $\tilde{X}\tilde{Y}$ centered at $(0, 0, r)$. Hence, a rotation θ and translation (t_x, t_y) in the world plane maps the induced coordinate axes onto the world axes.

2) Decomposing the projective transformation into two transformations, each dependent on different pose parameters. The first transformation is a central projection (perspectivity) from the image plane to a plane parallel to the world plane. The second is an equiform transformation between this plane and the world plane. This is described in more detail below.

A full discussion of the different methods and their ambiguities is given in [41], including experimental evaluations based on real images. We found that the equiform-based method produces the most *stable* (to camera calibration and conic fitting errors) and *unambiguous* results.

B. Pose Recovery via Equiform Invariants

The transformation between camera and world coordinate systems can be accomplished by a Euclidean 3D rotation and a translation. Here, we parameterize the transformation matrix to make explicit the dependence on plane orientation (a similar decomposition was used in [45]). The world plane (see Fig. 6) is expressed in camera coordinates as

$$z = px + qy + r. \tag{4}$$

We define a rectangular world coordinate system on the world plane by X and Y axes. The position and orientation of these axes can be given relative to the camera coordinate frame using an induced rectangular coordinate frame also lying on the world plane. The induced frame has its $\tilde{X}$ and $\tilde{Y}$ axes lying in the world plane centered at $(0, 0, r)$ and has the $\tilde{X}$ axis lying in the same plane as the camera x axis. Hence, the induced frame position and orientation depends only on p, q and r. A rotation and translation within the plane maps the induced coordinate frame onto the world frame. The rotation is represented by θ and the translations by t_x and t_y (see Fig. 6). Hence, the position and orientation of the world axes depends on θ, t_x, t_y and on p, q and r.

The perspective mapping T may be decomposed into two transformation matrices, an equiform matrix $H(r, \theta, t_x, t_y)$, and a central projection matrix $M(p, q)$. An equiform mapping is a combination of a plane Euclidean transformation and an isotropic scaling. It preserves angles and ratio of lengths. The Euclidean mapping is the rotation and translation represented by θ, t_x, and t_y. The scaling is represented by r. Hence, T may be written $T = M(p, q) \cdot H(r, \theta, t_x, t_y)$, where

$$M = \begin{bmatrix} \frac{1}{\sqrt{1+p^2}} & \frac{-pq}{\sqrt{1+p^2+q^2}\sqrt{1+p^2}} & 0 \\ 0 & \frac{\sqrt{1+p^2}}{\sqrt{1+p^2+q^2}} & 0 \\ \frac{-p}{\sqrt{1+p^2}} & \frac{-q}{\sqrt{1+p^2+q^2}\sqrt{1+p^2}} & 1 \end{bmatrix} \tag{5}$$

$$H = \begin{bmatrix} \alpha & -\beta & \beta t_y & -\alpha t_x \\ \beta & \alpha & -\alpha t_y & -\beta t_x \\ 0 & 0 & r & \end{bmatrix} \tag{6}$$

where $\alpha = \cos\theta$ and $\beta = \sin\theta$.

The key idea is to use properties of the curves that are equiform invariants. These are unaffected by r, θ, t_x, t_y (i.e., H) and thus generate equations only involving p and q. Thus, pose recovery proceeds in two stages: The equations generated by the equiform invariants are solved for p and q, and the remaining parameters are recovered using this information. The equiform invariant we use is a ratio of the trace and determinant of the upper 2×2 submatrix of the conic matrix

$$I(\mathsf{c}) = \frac{\left(\text{trace}^{2\times 2}[\mathsf{c}]\right)^2}{\det^{2\times 2}[\mathsf{c}]} \tag{7}$$

where $\text{trace}^{2\times 2}[\mathsf{c}] = A + \mathsf{c}$ and $\det^{2\times 2}[\mathsf{c}] = A \cdot \mathsf{c} - \frac{B^2}{4}$. Geometrically, this is related to the aspect ratio and area of the conic.

We have $\mathsf{c} = \kappa \cdot (T^T)\mathsf{c}T$ and $T = MH$. Thus

$$I(\mathsf{c}) = \frac{\left(\text{trace}^{2\times 2}[M^T \cdot \mathsf{c} \cdot M]\right)^2}{\det^{2\times 2}[M^T \cdot \mathsf{c} \cdot M]} \tag{8}$$

since the equiform invariant eliminates the effect of H. When multiplied out, this gives a quartic in p and q for each conic $\mathsf{c} = +\mathsf{c}_1, \mathsf{c}_2$. In general, there should be 16 solutions that satisfy the intersections of these quartics, but we have never observed more than four distinct real roots in any system. This is a remarkable feature of the polynomials, which reduces the ambiguity of the solution to reasonable size. Using these four values for p and q to determine the other pose parameters gives four distinct solutions. The quartics cannot be solved in closed form. We find a solution by continuation [33]. Continuation is a robust method of tracking the roots of a polynomial.

To determine the remaining pose parameters, we first obtain r (the scaling parameter) by minimizing the difference between the areas of the conics projected onto the plane $z = px + qy + r$ and the model conics. Then, we perform an in-plane rotation and translation to match the induced and model conics.

We now have four distinct pose solutions. We eliminate three of the solutions by choosing the one that produces the best match of the model conics to the image conics back projected onto the world plane. Details of this appear in [41].

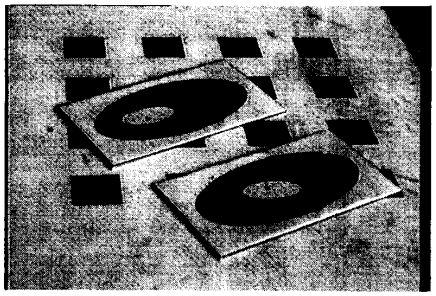


Fig. 7. There are two objects in the image, and we report pose results only for the nearer object. Under the objects, the calibration grid used to find the pose by Tsai's method is visible.

C. Implications for Model Acquisition

Recognition requires knowing only the invariant descriptors for the model. Computing pose, however, requires that the coefficients of the model conic *in the object plane* be known. If the object curves are known conics (this occurs in the case of the labels), then there is no difficulty. One needs to choose a coordinate system on the object plane within which to express these conics, and a sensible choice is the natural frame of one of the conics (e.g., for an ellipse, the center as origin and coordinate axes aligned with the principal axes).

However, if the object curves are not conics, it is necessary to fit conics to them, using the invariant fitting method. If the object curves are known in the world plane coordinate system, then the fitting process produces the model conics directly. If the curves are not known in the world frame, then they can be obtained by fitting conics in an image (using the invariant fitting method) and projecting these fitted conics back to the world plane. Back projection requires that the transformation between the object plane and the image plane be known. This can easily be recovered by imaging the object together with a known *calibration* pair of conics (or a single circle) that is coplanar with the object curves. The calibration conics then determine the transformation between the image plane and the world plane.

D. Pose Results

We describe two examples. The first compares the accuracy of determining absolute pose for a label with pose by Tsai's [46] calibration method for the same plane. The second compares the measured relative motion for a nonconic object (a gasket) rotated on a calibration table.

The pose parameters extracted from a real image (that of Fig. 7) are shown in Table IV. The object was placed on a calibration grid so that plane parameters could be compared with those found by Tsai's method. Instead of showing p and q, we use the plane parameters slant (σ) and tilt (τ) because these give a better indication of the size of the measurement errors ($\sigma = \cos^{-1}(\{1+p^2+q^2\}^{-1/2})$ and $\tau = \tan^{-1}(q/p)$). There is excellent agreement between conic pair based pose and Tsai's method.

TABLE IV
THIS TABLE COMPARES THE PLANE PARAMETERS FOUND USING THE CONIC PAIR WITH THOSE OBTAINED FROM TSAI'S CALIBRATION METHOD FOR THE IMAGE SHOWN IN FIG. 7. (ANGLES ARE IN DEGREES AND DISTANCES IN MILLIMETERS.)

method	σ	τ	r
conic pair	54.6°	85.0°	353.7
Tsai	55.6°	86.2°	365.0

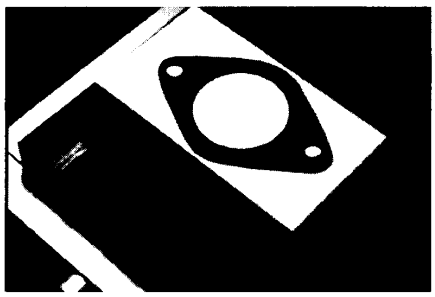


Fig. 8. We are able to represent objects that do not have conic features using the frame-invariant fitting theorem. This representation allows pose recovery as well as recognition. The gasket curves represented by conics are the outer gasket curve and the large circular hole. The gasket was rotated on a calibration table, the estimated relative motion was found by T matrix decomposition and the actual motion was compared.

TABLE V
AS THE GASKET IN FIG. 8 IS ROTATED IN ITS PLANE BY 20°, THE ESTIMATED POSE CHANGES BY A SIMILAR AMOUNT. (THE PLANE POSITION, GIVEN BY σ, τ AND r REMAINS FAIRLY CONSTANT).

position	σ	τ	r (mm)	θ
1	35°	94°	280	182°
2	35°	90°	277	163°
3	36°	88°	277	144°

As an example of relative motion, we can recover the angle that an object has been rotated between consecutive images (with fixed camera). The object used was a gasket, which we represent by two coplanar conics using the invariant fitting technique. The curves used are the outer boundary and the large circular hole (see Fig. 8). The gasket was rotated through 20° in its plane each time. Therefore, the plane parameters σ, τ, and r should remain constant. Table V shows the good agreement between actual and experimental estimation of relative motion.

V. RECOGNIZING THREE-DIMENSIONAL OBJECTS

The motivating principle of our recognition techniques is that descriptors should be unaffected by viewpoint. So far, we have only considered invariants of planar objects. An important factor in extending these recognition techniques to 3-D objects is that the secure framework of the plane projective group is no longer available. In the systems described above, we have modeled perspectivities (Euclidean motions of a plane in three space followed by perspective projection) by

Printed by : Black Bear Press Limited, Cambridge, United Kingdom

TABLE VI
ANGLE BETWEEN PLANES FOUND BY T MATRIX DECOMPOSITION FOR A NUMBER OF OBJECTS CONTAINING TWO LABELS ON DIFFERENT PLANES. (THE CAMERA'S POSITION AND ORIENTATION IS CHANGED BETWEEN VIEW 1 AND VIEW 2.)

object	a	b	c	d	e	f	g
view 1	0.6°	16.2°	33.2°	47.1°	60.3°	78.8°	90.6°
view 2	3.4°	12.8°	28.4°	45.1°	61.8°	74.9°	92.1°
actual	0.0°	15.0°	30.0°	45.0°	60.0°	75.0°	90.0°

actions of the plane projective group. This model is valid in that perspectivities are a proper subset of projectivities and has the advantage that there are standard methods available for generating invariants. Thus, unfortunately, 3-D objects force us to study perspectivities, which are harder to manage than projectivities because they do not form a group (the composition of two perspectivities is not guaranteed to be a perspectivity). For three-dimensional objects, invariant descriptors will consist of Euclidean invariants, rather than projective invariants. Euclidean invariants are object properties unaffected by Euclidean motions. From this point of view, the perspective mapping simply makes Euclidean invariants harder to observe. To summarize, perspective projection ensures that the image curve is a complicated function of the object's Euclidean invariants, its position, and its orientation. In the examples below, we show that image curves provide enough constraints to determine at least some of the Euclidean invariants from a single view.

A. Rigidly Coupled Planes

To illustrate the extraction of Euclidean invariants, we consider a constrained class of objects. These are objects that have pairs of curves lying on different, rigidly connected plane faces. An important Euclidean invariant in this case is the angle between the plane faces (though there are many others, which are discussed below). If this angle can be extracted from perspective views of the object, then it is an invariant shape descriptor and can be used to distinguish between objects with different angles.

The method of extracting this angle is based on pose, but it may be thought of as a "black box" into which are fed the image curves and from which emerges the angle. We assume that the curves are algebraic. If they are not, we can fit algebraic curves using affine invariant fitting. Then, it is a simple matter to recover the orientation of the planes using the method described in Section IV and, hence, the angle between them. Although some of the elegance obtained by using the plane projective group has been lost, this approach manages to retain the important ability of invariants to index directly into a model base at the cost of assuming a constrained class of models. The feasibility of accurately extracting this angle is illustrated by the results in Table VI. This compares the image measured and actual angle for a series of objects similar to the one shown in Fig. 9.

B. Circle Examples

As a further example of extracting Euclidean invariants from a constrained class of models. We consider three examples of objects that contain circles of equal radius. The advantage of having world conics that are circles is that plane orientation can be recovered directly from the single image curve, rather than requiring two curves, as is the case for noncircular conics. Furthermore, if the circle's radius is known, the plane is completely defined (i.e., its distance from the camera is recovered as well). The details of recovering the orientation and center of the circle are given in the appendix. The disadvantage of using circles is that there is a two-fold ambiguity in the recovered plane.

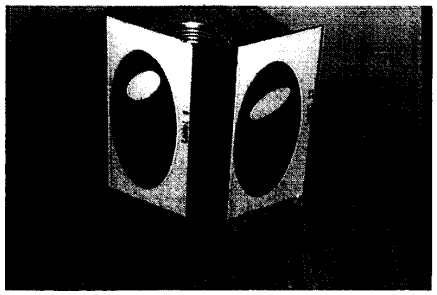


Fig. 9. This is a typical object used to measure the accuracy of recovering the angle between rigidly coupled planes. The object is composed of two planes set at 60° to each other. Each plane contains a pair of ellipses, each of which define a label in the model base.

For two circles in three dimensions, there are a number of invariants:

- **The angle between the planes.** This invariant may be recovered even if the radius of the circle is not known. It depends only on the orientations of the planes, which are recoverable without requiring knowledge of circle radius.
- **The ratio of circle radius to distance between circle centers.** Again, this may be recovered from the image without knowledge of the radius value.
- **The vector joining circle centers.** This vector has three degrees of freedom that may be expressed as:
  - — its length (this requires that the object scale is determined), written d_c
  - — the angle between the vector and normal of first plane, written θ_{n_1}
  - — the angle between the vector and normal of second plane, written θ_{n_2}

  Note that angles and ratios of lengths do not in general require object scale to be known. Another useful measurement of this vector (which is functionally dependent on the first three) is the angle between the vector and the vector product of the normals, written θ_u.

The accuracy and usefulness of these invariants is discussed in the following example.

1) Calibration Objects: A typical object is shown in Fig. 10. We tested three such objects where the angles between the faces were 0, 45 and 90°. Results are given in Tables VII and VIII. Table VII gives the ambiguous solutions. Clearly, some of the measures (e.g., distance) are insensitive to ambiguity, whereas others (e.g., angle) are not. The results in Table VIII exemplify the different accuracies of pose-determined invari-

TABLE VII
DETERMINING EUCLIDEAN INVARIANTS USING CIRCLES HAS A FOUR-FOLD AMBIGUITY. (THIS TABLE INDICATES THE RANGE OF THE AMBIGUOUS SOLUTIONS. IN THIS CASE, THE FOURTH SOLUTION IS THE CORRECT ONE. θ_{n_i} IS THE ANGLE BETWEEN THE VECTOR CONNECTING THE CIRCLE CENTERS AND THE NORMAL TO THE iTH PLANE, θ_u IS THE ANGLE BETWEEN THE VECTOR CONNECTING THE CENTERS AND THE CROSS PRODUCT OF THE PLANE NORMALS, AND d_c IS THE DISTANCE BETWEEN THE CENTERS OF THE CONICS.) (*Note*: Angles are in degrees and distances are in millimeters.)

invariant	actual	solution 1	solution 2	solution 3	solution 4
angle	90.0	80.9	67.9	66.6	89.3
d_c	49.5	47.9	48.6	48.5	49.3
θ_{n_1}	45.0	35.8	36.4	45.7	46.2
θ_{n_2}	45.0	45.9	44.1	46.3	44.6
θ_u	90.0	83.7	83.7	83.9	83.9

TABLE VIII
STABILITY OF RECOVERED EUCLIDEAN INVARIANTS GIVEN BY CIRCLES. (THIS SHOWS THE VARIATION OVER A RANGE OF OBJECTS AND VIEWS. ONLY THE CORRECT SOLUTION IS GIVEN. A TYPICAL OBJECT IS SHOWN IN FIG. 10).

object		angle	distance	Θn1	Θn2	Θu
1	actual	90.0	49.5	45.0	45.0	90.0
1	view 1	89.3	49.3	46.2	44.6	83.9
1	view 2	88.2	49.7	46.1	45.8	73.3
2	actual	45.0	64.7	67.5	67.5	90.0
2	view 1	44.1	64.1	70.0	66.0	79.9
2	view 2	44.5	63.8	67.4	68.1	83.4
3	actual	0.0	70.0	90.0	90.0	–
3	view 1	4.6	72.6	81.4	82.0	–
3	view 2	0.9	64.9	87.0	87.8	–

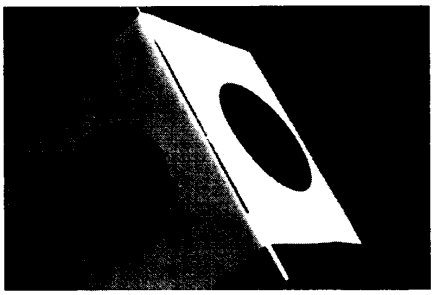


Fig. 10. Each plane contains a circle. This is a typical object used to measure the accuracy of recovering the angle between rigidly coupled planes. In this case, the planes are at 90° to each other.

ants. The angle between the planes is consistently accurate; however, the other angles are far less consistent. The explanation for this is that the angle between planes does not depend on their distance from the camera. However, the other angles are functions of the center separation. Thus, any error that may result from a slight fitting error or incorrect object measurements directly degrades these angles. This is particularly clear in the third object (the coplanar case), where less than a 4% error in the center separation causes a 10% error in angle. This illustrates another criteria for shape descriptors—that they should be first-order insensitive to calibration errors.

2) Glasses: As a final example, we use a constrained class of models consisting of circles on parallel planes. In this case, the ratio of plane separation to circle radius is an Euclidean invariant. If the circle radius is known, the plane separation is recoverable. This is illustrated in Table IX and Fig. 11. The results show that the liquid height can be monitored very accurately. Discrepancies can be attributed to the glass having sides that are not exactly parallel. In fact, for a class of models consisting of a pair of parallel planes, the pose may be recovered from a single known *conic* on each plane (the conics do not have to be circles).

VI. EVALUATING THE STABILITY OF INVARIANTS

In order for invariants to be useful in vision applications, they must be accurate and stable. Errors are introduced by the sensor, the feature extraction scheme, and the numerical precision of the computation. In this section, we derive expressions that characterize the propagation of error, and we report experiments performed to test four plane projective invariants: the cross ratio, the projective invariants of five coplanar points, the projective invariants of two coplanar conics, and a nonEuclidean metric. We emphasize that in the case of the conics, we investigate only the stability of the invariants of exactly fitted conics.

A. Experimental Setup

The experiments consist of measurements of invariants in images of patterns for which the value of the invariant is known for various viewing positions. To make the low-level processing and feature extraction as straightforward as possible, the test patterns consisted of simple convex polygonal shapes, which produce well-defined, sharp edges. The

Printed by The Black Boar Press Limited, Cambridge, United Kingdom

TABLE IX
RESULTS OF USING CIRCLES FOR MONITORING LIQUID HEIGHT. (THE IMAGES ARE SHOWN IN FIG. 11. THE UPPER RIM HAD A 32-MM RADIUS AND THE INNER GLASS A RADIUS OF 30 MM. NOTE, FOR PARALLEL PLANES, THERE IS NO AMBIGUITY IN CIRCLE SOLUTIONS BECAUSE, IN GENERAL, ONLY ONE OF THE FOUR SOLUTIONS WILL BE PARALLEL.)

Fig.11	angle between planes	actual separation	recovered separation
upper	2.5°	38 mm	38.6 mm
lower	2.3°	84 mm	87.7 mm

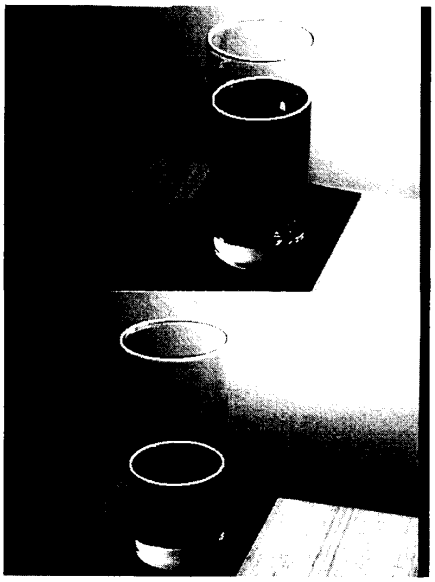


Fig. 11. For a pair of circles on parallel planes, the ratio of circle radius to plane separation can be recovered from any view except the degenerate view. Curves extracted are superimposed in white. Results are given in Table IX.

patterns were generated directly via the Postscript language and rendered on a laser printer.

To collect data, the test patterns were mounted on a platform on an optical bench that could be precisely rotated and translated with respect to the camera: a Sony model DXC-3000 CCD broadcast-quality video camera. The source of illumination was fixed with respect to the camera and positioned to provided reasonably uniform illumination for all the positions of the test patterns. The intensity was chosen so that a small lens aperture could be used to assure that the test pattern was within the depth of field of the camera for the entire experiment. The focal length of the lens was kept constant for the entire experiment, and the aperture was adjusted so that the video level for the white areas of the test pattern was approximately 90 IRE. The video output of the camera was digitized with a spatial resolution of 640 by 484 pixels and intensity resolution of 256 grey levels by a Symbolics frame grabber installed in a Symbolics 3670 Lisp Machine.

For each of the patterns, two sequences of images were taken. For the first, the pattern was positioned at a distance of 26 cm from the focal plane of the camera and rotated about a vertical axis in increments of 5°, where the angle is measured between the optical axis of the camera and the normal to the plane of the pattern. For the second sequence, the pattern was translated along the normal to the image plane in various increments from 26 to 210 cm. In addition, to test the reliability of each invariant with respect to spatial quantization, the images from the rotation sequence were resampled to 75% of their original size by linear interpolation of the pixel values.

The resulting images were segmented into networks of edges and vertices. Our current approach to segmentation uses the zero crossings in the second derivative of image intensity to define the location of feature boundaries [7] and then approximates these boundary pixel chains by straight line segments [3]. Feature points are localized as curvature extrema in the chains.

B. Results

1) The Cross Ratio: In what follows, we avoid correspondence issues by assuming that points are uniquely labeled. To determine the numerical sensitivity of the cross ratio, we apply the *law of propagation of error:*

$$\mathrm{var}^2[cr(x_1,x_2,x_3,x_4)] = \sum_{i=1}^{4}\left(\frac{\partial cr(x_1,x_2,x_3,x_4)}{\partial x_i}\right)^2 \cdot \mathrm{var}^2(x_i)$$

from which it follows that

$$\mathrm{var}[cr(x_1,x_2,x_3,x_4)] \propto |cr(x_1,x_2,x_3,x_4)|. \quad (9)$$

Thus, the variance of the cross ratio is proportional to the magnitude of the cross ratio itself. A large cross ratio occurs when two or more points are near each other on the line. It is also reasonable to expect that the variance of the cross ratio will be significant if the separation of two points is of approximately the same magnitude as the variance in point position.

The cross ratio was tested using a pattern consisting of three sets of four concentric squares alternately filled black and white designed to have cross ratio values $cr_A = 2\frac{2}{3}$, $cr_B = 12$, and $cr_C = 4$, respectively. The value of the cross ratio is computed twice for each set of concentric squares, first using the four upper corners and then the lower four corners of the diagonal. The two values should be equal. To guarantee that the points are collinear, we projected them on the straight line computed as the best eigenvector fit [11] of all eight points of the diagonal of the concentric squares.

Note the following points:

- For the rotation sequence of this test pattern, we found that when the pattern was tilted beyond 60°, the feature detector was not able to recover all four corner vertices of the smallest squares of the test pattern, and consequently, the cross ratio could not be computed.
- The third and fourth rows of Table X contain the values of the cross ratio for the nonreduced and reduced image sizes

TABLE X
EXPERIMENTAL RESULTS OF THE CROSS RATIO INVARIANT. ("REDUCTION" REFERS TO IMAGES THAT WERE ROTATED AND THEN RESAMPLED USING LINEAR INTERPOLATION.)

motion	range	cr_A	cr_B	cr_C
	design values	$2\frac{2}{3}$	12	4
Rotation	$0° \leftrightarrow 60°$	2.69 ±0.02	11.9 ±0.1	4.02 ±0.06
	$0° \leftrightarrow 50°$	2.69 ±0.02	12.00 ±0.06	4.00 ±0.06
Reduction	$0° \leftrightarrow 50°$	2.69 ±0.03	12.0 ±0.2	4.00 ±0.05
Translation	$26 \leftrightarrow 92$ *cm*	2.74 ±0.03	12.3 ±0.2	4.08 ±0.05
All		2.72 ±0.04	12.2 ±0.3	4.05 ±0.07
	σ/cr	0.02	0.02	0.02

TABLE XI
EXPERIMENTAL RESULTS OF THE COPLANAR FIVE-POINT INVARIANT. "REDUCTION" REFERS TO IMAGES THAT WERE ROTATED AND THEN RESAMPLED USING LINEAR INTERPOLATION

motion	range	IA_1	IA_2	IB_1	IB_2	IC_1	IC_2
	design values	0.618	2.618	0.401	4.41	0.858	1.197
Rotation	$0° \leftrightarrow 80°$	0.60 ±0.02	2.7 ±0.1	0.401 ±0.003	4.4 ±0.1	0.853 ±0.003	1.199 ±0.005
	$0° \leftrightarrow 70°$	0.613 ±0.006	2.65 ±0.03	0.401 ±0.003	4.40 ±0.05	0.859 ±0.002	1.196 ±0.003
Reduction	$0° \leftrightarrow 70°$	0.61 ±0.01	2.71 ±0.07	0.402 ±0.008	4.40 ±0.05	0.858 ±0.003	1.199 ±0.005
Translation	$26 \leftrightarrow 205$ *cm*	0.610 ±0.007	2.68 ±0.06	0.412 ±0.007	4.27 ±0.09	0.852 ±0.004	1.211 ±0.009
All		0.61 ±0.01	2.64 ±0.09	0.404 ±0.008	4.38 ±0.09	0.856 ±0.005	1.201 ±0.009
	σ/I	0.02	0.03	0.02	0.02	0.006	0.007

over the same range of angles. These results demonstrate that the cross ratio is stable with respect to spatial quantization noise.

- In the translation sequence, the patterns become too small for the feature detector to recover the four vertices necessary to evaluate the cross ratio at distances greater that 92 cm.
- In the last row of Table X, we have computed the ratio of the standard deviation to the mean of the cross ratio of the previous row. The fact that these values remain constant for different values of the cross ratio agrees with the relation of (14), which predicts that the variance of the value will be proportional to the value of the cross ratio itself.

In conclusion, we can observe from these results that the cross ratio is not only stable but also accurate since the design value of each test pattern lies within one standard deviation of the experimentally measured values.

C. Five Coplanar Points

As *Example 6* shows, five coplanar points give rise to two invariants, which can be written as

$$I_1 = \frac{I_{431} I_{521}}{I_{421} I_{531}}$$
$$I_2 = \frac{I_{421} I_{532}}{I_{432} I_{521}}$$

where I_{ijk} is the determinant of the matrix $\{\boldsymbol{p}_i \boldsymbol{p}_j \boldsymbol{p}_k\}$, and $\boldsymbol{p}_k$ are the homogenous coordinates of the kth point. If a triple of points is collinear, the point matrix $\mathtt{M}_{ijk}$ becomes singular, and the corresponding invariant is undefined. Note that reordering the points is an action of the permutation group on five symbols and that by taking symmetric functions of these two basic invariants, we can construct functions that are invariant to projection and relabeling. For simplicity, we assume that each point is uniquely labeled so that the problem does not arise. The law of propagation of error yields

$$\mathrm{var}^2[I_1(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5)] = \sum_{i=1}^{5} \left(\frac{\partial I_1}{\partial x_i}\right)^2 \mathrm{var}^2(x_i) + \left(\frac{\partial I_1}{\partial y_i}\right)^2 \mathrm{var}^2(y_i).$$

If $\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5$ are five coplanar points, it follows that

$$\mathrm{var}[I_1(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5)] \propto |I_1| \tag{10}$$

and that

$$\mathrm{var}[I_2(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4, \boldsymbol{p}_5)] \propto |I_2|. \tag{11}$$

This implies that given estimates of the value of the invariant and of the error introduced by the sensor and feature extraction scheme, we can estimate the accuracy of the invariant.

The pattern devised for measuring the five-point invariant consists of three black pentagons of varying shape on a white background. The five points used to compute the invariants are the vertices of the pentagon. The values of the invariant for each pentagon are given in Table XI.

Note the following points:

- For this pattern, we found that the maximum usable angle in the rotation sequence was 80°. For angles greater than this, the feature detector does not find the corners reliably. Consequently, since all five points are not available, the invariants cannot be calculated. For angles greater than 70°, the feature detector is able to recover all the vertices but with poor localization accuracy.

Printed by the Black Bear Press Limited, Cambridge, United Kingdom

TABLE XII
THE EXPERIMENTAL RESULTS OF THE TWO COPLANAR CONIC INVARIANT. "REDUCTION" REFERS TO IMAGES THAT WERE ROTATED AND THEN RESAMPLED USING LINEAR INTERPOLATION

motion	range	IAB_1	IAB_2	IAD_1	IAD_2	ICD_1	ICD_2
	design values	4.47	5.46	−30.4	−70.1	9.02	6.46
Rotation	0° ↔ 70°	4.46 ±0.06	5.44 ±0.02	−30.1 ±0.6	−69.7 ±0.4	9.02 ±0.06	6.46 ±0.03
	0° ↔ 65°	4.47 ±0.01	5.43 ±0.01	−30.3 ±0.3	−69.5 ±0.3	9.04 ±0.03	6.45 ±0.02
Reduction	0° ↔ 65°	4.47 ±0.01	5.46 ±0.02	−30.1 ±0.6	−70.3 ±0.1	8.98 ±0.09	6.47 ±0.02
Translation	26 ↔ 210 cm	4.47 ±0.01	5.46 ±0.02	−30.1 ±0.2	−68.5 ±0.6	9.02 ±0.03	6.46 ±0.01
All		4.46 ±0.01	5.44 ±0.02	−30.1 ±0.5	−69.7 ±0.8	9.01 ±0.06	6.45 ±0.03
$\sigma/I^2(\times 10^{-4})$		5	7	6	2	7	7

TABLE XIII
THE RESULTS OF THE NONEUCLIDEAN DISTANCE TESTS. ("REDUCTION" REFERS TO IMAGES THAT WERE ROTATED AND THEN RESAMPLED USING LINEAR INTERPOLATION.)

motion	range	$\delta(\boldsymbol{x}_3, \boldsymbol{x}_4)$	$\delta(\boldsymbol{x}_1, \boldsymbol{x}_3)$	$\delta(\boldsymbol{x}_1, \boldsymbol{x}_2)$
	design values	0.381	1.096	0.0222
Rotation	0° ↔ 70°	0.382 ±0.005	1.096 ±0.009	0.0221 ±0.0007
	0° ↔ 65°	0.385 ±0.001	1.091 ±0.006	0.0222 ±0.0005
Reduction	0° ↔ 65°	0.378 ±0.005	1.097 ±0.006	0.0218 ±0.0007
Translation	26 ↔ 210 cm	0.379 ±0.007	1.102 ±0.004	0.0227 ±0.0004
All		0.381 ±0.009	1.096 ±0.008	0.0221 ±0.0007

- For angles less than 70°, the five-point invariant is stable under spatial quantization noise.
- For the translation sequence, the maximum usable distance is 210 cm. Beyond this distance, the pattern becomes too small for the feature detector to recover the five vertices that are necessary to evaluate the invariants.
- In the last row of Table XI, we have computed the ratio of the deviation to the mean of the invariants of the previous row. These results agree with the relation of (15). Note that the ratio for pattern C is slightly lower than the others. This is because the feature detector is able to localize vertices with greater precision when their incident edges form an angle close to 90°, which is the case in this pattern.

These results demonstrate that the five-point invariant is not only stable but also accurate.

D. Two Coplanar Conics

Example 5 gives projective invariants for a pair of coplanar conic curves.

The pattern for measuring the invariants of two coplanar conics consists of four pentagons. The vertices of each pentagon define a unique ellipse. The design values of the invariants of these pairs of conics are given in Table XII. Out of the six possible pairings of the conics, we have chosen the AB, AC, and CD, whose invariants are given in Table XII.

The limits for recovering all the necessary features under rotation, reduced rotation, and translations sequences were 70°, 65°, and 205 cm, respectively. The results are shown in Table XII. These were produced in the same way as those of the previous sections.

We see that this invariant is both stable and accurate since the design value falls within one standard deviation of the measured values. However, unlike the previous cases where the propagation of error is proportional to the values of the invariants, here, the relationship is quadratic due to the quadratic dependence of the coefficients of the conic on the coordinates of the vertices. The last row of Table XII illustrates this relationship.

E. NonEuclidean Distance

The cross ratio can be used to introduce a distance in the image plane that is invariant under projective transformation.

Definition: The projective distance δ between two points x_1 and x_2 on a line with respect to two points m_1 and m_2 on the line is given by

$$\delta(x_1, x_2) = k \log[cr(x_1, x_2, m_1, m_2)] \tag{12}$$

where k is an arbitrary real or complex constant.

A 2-D generalization is possible if there is a distinguished conic in the plane (*Example 7*). Then, for two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ that lie on the plane but not on the conic c, the function

$$\frac{(\boldsymbol{x}_1^T \mathsf{c} \boldsymbol{x}_2)^2}{(\boldsymbol{x}_1^T \mathsf{c} \boldsymbol{x}_1)(\boldsymbol{x}_2^T \mathsf{c} \boldsymbol{x}_2)} = f(x_1, x_2) = \begin{cases} \cos h^2(\delta(x_1, x_2)) & \text{if } |f(x_1, x_2)| > 1 \\ \cos^2(\delta(x_1, x_2)) & \text{otherwise} \end{cases} \tag{13}$$

defines a distance δ that is independent of the frame in which the points and the conic are measured [44].

The same pattern used for testing a pair of coplanar conics was used for testing nonEuclidean distance, and therefore, the orientation limits that held for that case hold for the tests of nonEuclidean distance. The reference conic c (18) is determined by one of the pentagons. We measured three distances on the figure. The results of these experiments are shown in Table XIII.

Although the Euclidean distance between the selected points changed by over a factor of 3 when the pattern was rotated, the nonEuclidean distance remains essentially constant.

VII. Discussion

We have demonstrated a simple, efficient model-based vision system that uses the principle of invariance to recognize objects, without regard to pose. We have shown that once an object has been identified, its pose can easily and accurately be recovered. Extensive generalization of this work is possible:

- The invariant fitting theorem works for algebraic curves of any degree. Higher degree curves have richer systems of invariants in general, but the complexity of the numerical problem in fitting the curves is massively increased. It may be possible to solve these problems to create richer invariant descriptors for plane curves.
- The invariant fitting theorem applies only to *point sets* that are within projection. To obtain an invariant representation for *curves* requires integrating algebraic distance with respect to a projectively invariant parameter. However, simply summing algebraic distance at all the points on the sampled curve leads to useful invariants in practice.
- The projectively invariant function associated with a single conic (2.1) can be turned into a metric as in Section VI-E. It should be possible to use this function to avoid having to back project models to verify the hypothesis that an instance has been found by checking that pairs taken from a range of nearby features have the right function values. Using this scheme allows richer models, yet the precise camera calibration needed to compute pose and backproject in conventional hypothesis verification is unnecessary.
- The system for working with nonalgebraic curves using algebraic representations is sensitive to occlusion. It may be possible to overcome this difficulty by using the projective differential invariants described in Section II (*Examples 4 and 5*) or by exploiting the projectively invariant function associated with one of the conics (2.1). This would proceed by fitting a conic to an image curve, assuming that it had not been occluded. The projectively invariant distances between other image features could then be used for recognition—a correct system of values for the distances both recognizes a model and verifies that the original curve was not occluded. A system of this kind would be robust to occlusion if each model consisted of the projectively invariant distances between a range of features for several different object curves. Such a model could be recognized as long as one object curve and a number of object features were not occluded.
- Differential invariants based at more than one point yield another approach to occlusion sensitivity for general curves. For example, given a concavity in a smooth curve, we can construct the bitangent that closes off the concavity and two lines tangent to the concavity and passing through the points of tangency of the bitangent. This construction is projectively invariant; therefore, we can consider the system that consists of the four points of tangency and a variable point that runs along the curve. This system will yield at least two invariants for each position of the variable point without taking any further derivatives. By plotting one invariant against the other, we obtain a signature for the curve that is projectively invariant and invariant to reparametrization.
- A unifying theme of this paper has been recovering properties that are invariant to a group action—unfortunately, plane perspectivities do not form a group. However, the really important group in vision is the three-dimensional Euclidean group. We have shown an example where Euclidean invariants could be recovered despite the distorting effects of perspective. We believe that a better understanding of perspectivities is important to progress in vision. The course most likely to succeed in avoiding the effects of perspective involves computing Euclidean invariants from data observed under perspective.
- Some of the polynomial systems encountered, for example, the systems in p and q in pose recovery or the systems in the rotation invariants of a surface, are overdetermined—there are more equations than there are unknowns. These extra constraints may usefully be exploited, for example, in reducing ambiguities and are the subject of active ongoing research.
- Recognizing curved surfaces from a single perspective view is an important problem that has not been addressed here. Recent work suggests invariant theory can make an important contribution to this problem [16].

Appendix A
Pose from Circles

The problem here is, given a known circle on the world plane and its corresponding conic in the image, determine the pose of the world plane.

The solution is in two stages:

1) Determine the orientation of the plane.
2) Determine the distance of the plane from the camera origin.

The method[13] exploits the property, unique to a conic that is a circle, that the back-projected curve must have equal coefficients for X^2 and Y^2 and no term in XY.

1) Plane Orientation: The conic in the image is $ax^2 + bxy + cy^2 + dx + ey + f = 0$. We assume that the origin is at the principal point, and the distances are measured in units of focal length (i.e., the focal length can be set to unity). Then, the image curve defines a cone $ax^2 + bxy + cy^2 + dxz + eyz + fz^2 = 0$ in 3D. The matrix in the quadratic form representation of this cone ($\boldsymbol{x}^T \mathtt{c} \boldsymbol{x}$ where $\boldsymbol{x} = \{x, y, z\}$) may be diagonalized in the standard manner by a 3D rotation of the coordinate system to the eigen-vector frame. We have $\mathtt{c}' = \mathtt{R}_1^T \mathtt{c} \mathtt{R}_1$ and $x' = \mathtt{R}_1^T \boldsymbol{x}$, where $\mathtt{R}_1 = (e_1, e_2, e_3)$ is the matrix of orthonormal eigen vectors and

$$\mathtt{c}' = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

[13] We are grateful to C. Longuet-Higgins for this solution. An independently developed solution, similar to this one, appeared in [8].

where $\{\lambda_1, \lambda_2, \lambda_3\}$ are the eigenvalues in ascending order. Equality of the x^2 and y^2 coefficients is achieved by a second rotation about the y' axis by an angle $\theta = \pm \tan^{-1} \sqrt{\frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_2}}$, which sets both of the coefficients to λ_2. There is therefore a two-fold ambiguity in the recovered orientation. We have $\mathtt{c}'' = \mathtt{R}_2^T \mathtt{c}' \mathtt{R}_2$ and $\boldsymbol{x}'' = \mathtt{R}_2^T \boldsymbol{x}'$ where

$$\mathtt{R}_2 = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}.$$

The composite rotation from image plane to the plane that intersects the cone in a circle is thus $\boldsymbol{x}'' = (\mathtt{R}_c)^T \boldsymbol{x}$, where $\mathtt{R}_c = \mathtt{R}_1 \mathtt{R}_2$, and consequently, the normal to the plane in the camera coordinate system is $n = \mathtt{R}_c (0, 0, -1)^T$ (the -1 allows for the right-handed coordinate system). Thus, the normal is $n = -(\mathtt{R}_{c13}, \mathtt{R}_{c23}, \mathtt{R}_{c33})$.

2) Plane Distance: In the x'' coordinate system, the circle has the equation $(x'' - \alpha)^2 + y''^2 = \frac{\lambda_1 \lambda_3 z''^2}{\lambda_2^2}$, where $\alpha^2 = \frac{(\lambda_3 - \lambda_2)(\lambda_2 - \lambda_1) z''^2}{\lambda_2^2}$. If the circle has radius ρ in this plane, then $z'' = d_\perp = \sqrt{\frac{-\lambda_2^2}{\lambda_1 \lambda_3}} \rho$ is the perpendicular distance of the plane from the origin. The center of the circle is at $\mathtt{R}_c (\alpha, 0, d_\perp)^T$ in the camera frame.

ACKNOWLEDGMENT

The authors wish to thank M. Brady for a range of helpful contributions. C. Brown was an active participant in early Oxford discussions about the applications of invariant theory in computer vision. C. Longuet-Higgins provided invaluable help in solving the problem of backprojecting a circle. Thanks also to A. Blake, M. Fleck, D. Kapur, and C. Marinos for a number of stimulating discussions. GE provided a grant that make this paper possible.

REFERENCES

[1] S. S. Abhyankar, "Invariant theory and the enumerative combinatorics of Young tableaux," in *Proc. First DARPA-ESPRIT Joint Workshop Applications Invariance Comput. Vision,* 1991.
[2] V. I. Arnold, "Geometrical methods in the study of ordinary differential equations," in *Grundlehren der Mathematische Wissenschaft.* Berlin: Springer-Verlag, 1990.
[3] H. Asada and M. Brady, "The curvature primal sketch," in *Proc. IEEE Workshop Comput. Vision: Representation Contr.,* 1984.
[4] E. B. Barrett, P. M. Payton, N. N. Haag, and M. H. Brill, "General methods for determining projective invariants in imagery," *Comput. Vision Graphics Image Proc.,* Jan. 1991.
[5] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry.* New York: Academic, 1986.
[6] R. A. Brooks, "Model-based three-dimensional interpretations of two dimensional images," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. 5, no. 2, p. 140, 1983.
[7] J. F. Canny, "Finding edges and lines in images," TR 720, MIT AI Lab, 1983.
[8] M. Dhome, J. T. Lapreste, G. Rives, and M. Richetin, "Spatial localization of modeled objects of revolution in monocular perspective vision," in *Proc. ECCV1,* 1989.
[9] L. E. Dickson, *Algebraic invariants.* New York: Wiley, 1913.
[10] J. Dieudonné and J. B. Carrell, *Invariant Theory Old and New.* New York: Academic, 1971.
[11] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* New York: Wiley, 1973.
[12] G. Ettinger, "Large hierarchical object recognition using libraries of parameterized model sub-parts," in *Proc. CVPR,* 1988, p. 32.
[13] D. A. Forsyth, J. L. Mundy, A. Zisserman, and C. M. Brown, "Projectively invariant representations using implicit algebraic curves," in *Proc. 1st Europ. Conf. Comput. Vision,* Springer Verlag Lecture Notes in Computer Science, 1990.
[14] ——, "Applications of invariant theory in computer vision," in *Proc. Int. Workshop Integrating Symbolic Numerical Comput.,* (Saratoga Springs), 1990.
[15] D. A. Forsyth, J. L. Mundy, and A. Zisserman, "Transformational invariance—a primer," in *British Machine Vision Assoc. Conf. Proc.,* 1990.
[16] D. A. Forsyth, J. L. Mundy, A. Zisserman, and C. A. Rothwell, "Invariant descriptors for 3D object recognition and pose," in *Proc. First DARPA-ESPRIT Joint Workshop Applications Invariance Comput. Vision,* 1991.
[17] S. Ganapathy, "Decomposition of transformation matrices for robot vision," in *Proc. Int. Conf. Robotics Automat.,* 1984.
[18] J. H. Grace, and A. Young, *The Algebra of Invariants.* Cambridge, UK: Cambridge University Press, 1903.
[19] W. E. L. Grimson and T. Lozano-Perez, "Search and sensing strategies for recognition and localization of two and three dimensional objects," in *Proc. 3rd Int. Symp. Robotics Research,* MIT Press, 1985.
[20] R. Hartshorne, *Foundations of Projective Geometry.* Benjamin, 1967.
[21] D. P. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment," in *Proc. DARPA Image Understanding Workshop,* 1988, p. 1114.
[22] I. Ikeuchi and T. Kanade, "Modeling sensors and applying sensor model to automatic generation of object recognition program," in *Proc. DARPA Image Understanding Workshop,* 1988, p. 697.
[23] K. Kanatani, *Group Theoretical Methods in Image Understanding,* New York: Springer, 1990.
[24] D. Kapur and J. L. Mundy, "Fitting affine invariant conics to curves," in *Proc. First DARPA-ESPRIT Joint Workshop Applications Invariance Comput. Vision,* 1991.
[25] Y. Lamdan, J. T. Schwartz, and H. J. Wolfson, "Object recognition by affine invariant matching," in *Proc. CVPR 88,* 1988.
[26] E. P. Lane, *A treatise on Projective Differential Geometry.* Chicago: University of Chicago Press, 1941.
[27] S. Linnainmaa, D. Harwood, and L. S. Davis, "Pose determination for a three dimensional object using triangle pairs," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. 10, no. 5, p. 634, 1988.
[28] D. Lowe, *Perceptual Organization and Visual Recognition.* Boston: Kluwer, 1985.
[29] S. Masciangelo, "3-D cues from a single view: Detection of elliptical arcs and model-based perspective backprojection," in *Proc. BMVC 90,* (Oxford), 1990.
[30] S. J. Maybank, "The projection of two non-coplanar conics," in *Proc. First DARPA-ESPRIT Joint Workshop Applications of Invariance Comput. Vision,* 1991.
[31] F. H. Moffit, and E. Mikhail, *Photogrammetry.* New York: Harper and Row, 1990.
[32] R. Mohr and L. Morin, "Geometric solutions to some 3D vision problems," in *Proc. First DARPA-ESPRIT Joint Workshop Applications Invariance Comput. Vision,* 1991.
[33] A. P. Morgan, *Solving Polynomial Systems Using Continuation for Scientific and Engineering Problems.* Englewood Cliffs, NJ: Prentice-Hall, 1987.
[34] L. Nielsen, "Automated guidance of vehicles using vision and projectively invariant marking," *Automatica,* vol. 24, no. 2, pp. 135–148.
[35] L. Nielsen and G. Sparr, "Projective area-invariants as an extension of the cross-ratio," in *Proc. 6th Scandinavian Conf. Image Analy.,* (Oulu), 1989, pp. 969–986.
[36] P. J. Olver, *The Application of Lie Groups to Differential Equations.* New York: Springer Verlag, 1986.
[37] A. P. Pentland, "Automatic extraction of deformable part models," *Int. J. Comput. Vision,* vol. 4, no. 2, p. 107, 1990.
[38] J. Ponce and D. J. Kriegman, "On recognizing and positioning curved 3 dimensional objects from image contours," in *Proc. DARPA IU Workshop,* 1989, pp. 461–470.
[39] J. Ponce, A. Hoogs, and D. J. Kreigman, "On using CAD models to compute the pose of curved 3D objects," in *Proc. First DARPA-ESPRIT Joint Workshop Applications Invariance Comput. Vision,* 1991.
[40] J. Porrill, "Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter," in *Proc. 5th Alvey Vision Conf.,* 1989.
[41] C. A. Rothwell, A. Zisserman, C. I. Marinos, D. A. Forsyth, and J. L. Mundy, "Relative motion and pose from arbitrary plane curves," *Image Vision Computing,* 1991.
[42] G. Salmon, *A treatise on the Higher Plane Curves* (3rd ed.), 1879.
[43] ——, *Higher Algebra.* Chelsea, 1885.
[44] C. E. Springer, *Geometry and Analysis of Projective Spaces.* San Francisco: W. H. Freeman, 1964.

[45] D. W. Thompson and J. L. Mundy, "3D model matching from an unconstrained viewpoint," in *Proc. IEEE Conf. Robotics Automat.*, 1987.
[46] R. Y. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," *IEEE J. Robotics Automat.*, vol. RA-3, no. 4, 1987.
[47] H. W. Turnbull, *Determinants, Matrices and Invariants.* Glasgow: Blackie and Son, 1928.
[48] Van Gool, L. Wagemans, J. Vandeneende, and A. Osterlinck, "Similarity extraction and modeling," in *Proc. ICCV 3,* 1990.
[49] I. Weiss, "Projective invariants of shapes," in *Proc. DARPA IU Workshop,* 1989, pp. 1125–1134.
[50] H. Weyl, *The Classical Groups and Their Invariants* 2nd ed. Princeton, NJ: Princeton University Press, 1946.

David Forsyth (M'88) received the B.Sc. degree in 1984 and the M.Sc. degree in 1986, both in electrical engineering, from the University of the Witwatersrand, Johannesburg, S. Africa. He received the D.Phil. degree from Oxford University in 1989. He holds a Prize Fellowship at Magdalen College, Oxford.

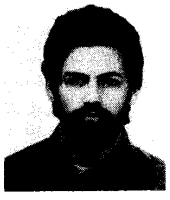
His interests include computer vision, invariant theory, computer algebra, and algebraic geometry.

Andrew Zisserman (M'86) received the B.A. degree in theoretical physics and part III mathematics from the University of Cambridge and, in 1984, the Ph.D. degree in applied mathematics.

From 1984 to 1987, he was an Alvey funded Research Fellow at the University of Edinburgh. He is currently a SERC Advanced Research Fellow in the Robotics Research Group, University of Oxford. His research interests are in robot vision, especially qualitative vision and applications of invariance. He has published a number of papers in this area, and a book with A. Blake (*Visual Reconstruction,* MIT Press).

Charles A. Rothwell was born in Bromsgrove, England, on April 30, 1968. He received the B.A. degree in engineering and computing science from Oxford University in 1990 and is currently pursuing the D.Phil. degree in Computer Vision with the Robotics Research Group, Oxford University.

His research interests are on model-based vision and the applications of invariance in vision.



Christopher Coelho was born in New York, NY, on April 26, 1962. He moved to Italy, where, in 1983, he received the diploma in electronics and, in 1989, the degree in physics from the University of Genoa, Italy.

Since 1989, he has worked on image processing and computer vision at the Department of Physics of Genoa. In 1990, he was a Visiting Scientist at General Electric Research and Development, working on invariant theory in computer vision. His research interests include low-level vision, image understanding, geometric reasoning, and invariant theory. Since February 1991, he has been employed with ELSAG BAILEY, where he is developing parallel algorithms for satellite image processing.

Aaron J. Heller (M'90) received the B.S. degree in computer science and applied mathematics in 1982 and the M.S. degree in computer science, both from the State University of New York, Albany.

He has been with the Artificial Intelligence Laboratory of the General Electric Research and Development Center since 1986. While there, he has conducted research in image understanding with applications in automated photo interpretation, robotics, autonomous navigation, medical imaging, and industrial inspection. His projects have included the development of a 3-D model-based object-recognition system based on the vertex-pair perceptual grouping and extending the system to include facilities for automated model-feature selection, transform search-space reduction, viewpoint constraint, hypothesis verification, and loosely coupled distributed processing. He has recently completed extensive benchmark testing of the object-recognition system on a variety of platforms. He was also a co-developer of a software system to represent and manipulate the geometry and topology of 3-D object models. His current work involves exploring representations and recognition algorithms for multiresolution and parametric models and the use of projective invariants for object recognition. Before joining GE, he was an Instructor at Rensselaer Polytechnic Institute in the Center for Interactive Computer Graphics, where he taught a course in Interactive Computer-Aided Design. He also conducted research in Graphical User Interfaces and graphics algorithms and headed a group to study applications and make recommendations on computer-aided design and manufacturing technology for the Center's corporate associates.

Mr. Heller is a member of the IEEE, the ACM, and the Audio Engineering Society.

Joseph L. Mundy received the B.S. degree, the M. Eng. degree, and the Ph.D. degree in 1963, 1966, and 1969, respectively, all in electrical engineering, from Rensselaer Polytechnic Institute, Troy, NY.

Since 1963, he has been a technical staff member of the General Electric Research and Development Center. His early projects at the Center include high-power microwave tube design, a superconductive computer memory system, the design of high-density associative memory compression and the development of an automatic inspection system for lamp filaments. From 1977 until 1982, he was Manager, Visual Information Processing Program. In this capacity, he led a number of extensive efforts in the application of machine vision techniques to automatic visual inspection for quality control. He is currently leading a project to apply methods in formal geometric analysis to construct object models for recognition and to develop a practical model-based recognition system for photo interpretation. He is also collaborating with Oxford University on a new approach to object recognition based on perspective invariants.

Dr. Mundy's professional activities involve active participation in the areas of pattern recognition and robotics. He was co-chairman of the workshop on industrial application of machine vision that resulted in a special issue of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (PAMI) in 1982. He is currently on the editorial boards of PAMI and the *International Vision Journal.* He also serves on the NSF advisory board for Artificial Intelligence and Robotics (IRIS). In 1988, he was named a Coolidge Fellow, which is the highest technical award at GE. He has recently completed a year's sabbatical as a visiting fellow at Oxford University.