Transformational invariance - a primer

David Forsyth, Joe Mundy and Andrew Zisserman

Oxford University Robotics Research Group
Department of Engineering Science
Oxford, UK

The shape of objects seen in images depends on the viewpoint. This effect confounds recognition. We demonstrate a theoretical framework within which it is possible to construct descriptors for both curves and surfaces, which do not vary with viewpoint. These descriptors are known as invariants.
We use this framework to construct invariant shape descriptors for plane curves. These invariant shape descriptors make it possible to recognise plane curves, without explicitly determining the relationship between the curve reference frame and the camera coordinate system, and can be used to index quickly and efficiently into a large model base of curves.
Many of these ideas are demonstrated by experiments on real image data.

The fundamental problem of computer vision is that shape measured in images depends not only on object shape, but also on the position, orientation and intrinsic parameters of the camera. If it is possible to define shape descriptors that are unaffected by perspective transformations, then image measurements of these descriptors can be matched to object properties regardless of camera viewpoint. Shape descriptors with these properties are known as invariants. This paper lays the foundation for systematic application of invariants in vision.
Many properties are invariant to projection: for example, straight lines project to straight lines and colinearities and intersections are preserved. The exploitation of these invariants has been responsible for the success of polyhedral model based vision. However, for smooth curves and surfaces, invariants such as zeroes of curvature, the cross ratio and Gaussian curvature do not provide a sufficiently strong set of constraints for successful model based vision. There has been a correspondingly limited success in representing and recognising curved objects. This paper shows how a rich invariant theory for curves and surfaces may be constructed.
We believe that invariance is the essential property of a shape description.
An invariant is defined in the context of a particular transformation. Area and curvature are invariant under translations and rotations in the plane, but not under perspective projection. As the generality of the transformation increases there is a corresponding increase in the complexity of the invariant. The generality is loosely determined by the number of parameters specifying the transformation.
In vision we are principally concerned with *perspective* projection. The invariants for plane objects subject to rigid motion and perspective projection, and the application of these invariants are the substance of this paper.
In section 1, we give a broad discussion of the mathematics and ideas underlying our use of invariant theory, and demonstrate the construction of invariants for a broad range of transformations. Although this does not cover every case, most of the transformations that naturally occur in vision are covered by this discussion. In particular, given a space and the transformations acting on the space, we show how to count the number of independent invariants of that action, and how to construct all of them.
In section 2, we show briefly the usefulness of this theory in model based vision. We demonstrate a projectively invariant representation for plane curves, using conic curves, and use this representation to build a simple and effective model based vision system.
A number of invariants are exploited in vision at present, with varied success and coherence. [12] demonstrated the value of using invariants of camera rotation in computing optical flow. [18] has also raised the issue of invariant representations. However, wide applications of invariant techniques to date have been limited because invariants are used only where they are easily guessed, or exist already in the literature.

1 INVARIANT THEORY

We have used the term "transformation", which is widely accepted to refer to an effect that is invertible. The idea is usefully generalised 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 $g : \mathcal{M} \rightarrow \mathcal{M}$:

$$id(x) = x \quad (1)$$
$$(g_1 \times g_2)(x) = (g_1(g_2(x)) \quad (2)$$

where $g_1, g_2 \in \mathcal{G}$, id is the identity element of the group, and $\times$ is the group composition function. An *invariant* of a group action is defined as follows:

Definition Consider a group $\mathcal{G}$ acting on a space, $\mathcal{M}$. The action of a group element g takes a point $\mathbf{p} \in \mathcal{M}$ to the point $\mathbf{p}' = g(\mathbf{p})$. An invariant, $I(\mathbf{p})$, is a function of $\mathbf{p}$ alone with the property that $I(\mathbf{p}') =$

BMVC 1990 doi:10.5244/C.4.3

$I(\mathbf{p})h(g)$. Here $h(g)$ is a function[1] of g alone. A *scalar* invariant is an invariant where $h(g) = 1$.

In what follows, we concentrate on scalar invariants, the term invariant should be understood to mean scalar invariant, except where the context clearly indicates otherwise.

Example 1: The plane rotation group acts on the plane, by the mapping $\mathbf{x}' = \mathbf{R}\mathbf{x}$, where R is a 2D rotation matrix. Any function of the distance from the origin to a point is invariant under the action of this group. Under the action of this group combined with the multiplicative group, $(x, y) :\longrightarrow \lambda(x, y)$, the function $x^2 + y^2$ is an invariant of weight 2. In the second case there is no scalar invariant, however.

Example 2: Differential invariants are functions of the position and derivatives of a curve at a point. The space on which the group acts in this case is the position of a point on a curve, and the values of various of the derivatives of the curve at that point because, when a group acts on a curve, it acts not just on the points of the curves but on all its derivatives as well. Differential invariants are clearly important in vision. Curvature, torsion and Gaussian curvature, all invariants under Euclidean actions, have been widely applied. Differential invariants for more general actions can be constructed using techniques shown below. For example, a projective differential invariant for plane curves has been known for a long time [18, 13]. However, this invariant is an extremely large and complex polynomial in the derivatives of the curve, and it is not known how useful in practice it will be.

Example 3: A plane conic can be written as $\mathbf{x}^t A\mathbf{x}$, for $\mathbf{x} = (x, y, z)$ and a symmetric matrix A, which determines the conic. A pair of coplanar conics has two scalar invariants, which we will describe here. Given conics with matrices of coefficients $\mathbf{A}$ and $\mathbf{B}$, we define:

$$\begin{aligned} I_{ab1} &= Trace(\mathbf{A}^{-1}\mathbf{B}) \\ I_{ab2} &= Trace(\mathbf{B}^{-1}\mathbf{A}) \end{aligned}$$

Under the action, $\mathbf{x} = \mathbf{T}\mathbf{x}$, $\mathbf{A}$ and $\mathbf{B}$ go to $\mathbf{A}' = \mathbf{T}^t\mathbf{A}\mathbf{T}$ and $\mathbf{B}' = \mathbf{T}^t\mathbf{B}\mathbf{T}$. In particular, using the cyclic properties of the trace, we find:

$$\begin{aligned} I'_{ab1} &= Trace(\mathbf{T}^{-1}\mathbf{A}^{-1}(\mathbf{T}^t)^{-1}\mathbf{T}^t\mathbf{B}\mathbf{T}) \\ &= Trace(\mathbf{A}^{-1}\mathbf{B}) \\ &= I_{ab1} \end{aligned}$$

A similar derivation holds for I_{ab2}. We will use these joint scalar invariants in section 3. Further examples of invariants appear in [9].

1.1 Invariants and orbits

In this section, we demonstrate the rudiments of a theory that allows us to deal with invariants using geometric ideas. This theory makes it possible to count invariants and, as we shall see below, to determine invariants for the action of a large number of different groups. The groups for which we can construct invariants are connected Lie[2] groups. We do not consider finite groups, such as the crystallographic groups. Examples of connected Lie groups include the n dimensional translation group, which is an n dimensional manifold isomorphic to $\Re^n$, and the plane rotation group, which is a one dimensional manifold that looks like a circle (obvious from the angular parametrisation of plane rotations). Conveniently, the Euclidean groups, which are the most important groups for vision applications, happen to be connected Lie groups.

An orbit of the action of a group through a point x is the set $\{g(x) | \forall g \in G\}$, for some x, where $g(x)$ denotes the action of g on x. Thus, to obtain an orbit, we take a point in the space and apply the action of every element of the group to it. Two orbits, one through x and another through y, are either disjoint, or coincide completely - orbits cannot intersect one another. An orbit could be the whole space (for example, take the space to be a line; the natural action of the one dimensional translation group translates points along the line, and so it is possible to reach every point from every other point), or it could be a "nice" submanifold of the space (for example, take the space to be the plane; under the action of the one dimensional translation group, each orbit is a line consisting of all the points that can be reached by translation in a fixed direction from a given start point). In the first case, there are clearly only trivial scalar invariants, because every point in the space can be reached from every other point, and hence the value of the scalar invariant must be the same at every point. The second case is more interesting, because there will be scalar invariants. In the second example, the invariant will be the y-intercept of each line.

The two following points are essential:

- A scalar invariant is a function that is constant on an orbit.
- The maximum dimension of any orbit is the dimension of the group.

The first point follows from the definition. The second point allows us to count the number of invariants. In particular, to specify a closed k dimensional submanifold of p dimensional space, we must "fix" $p - k$ coordinates; for example, to specify a surface in three space, one coordinate must be constant, and to specify a curve, two coordinates must be constant. In turn, this means we have $p - k$ functions that are constant on this submanifold, and, if the submanifold is an orbit, we then have $p - k$ invariants. This means that there can be at most

[1] In fact, a homomorphism of $\mathcal{G}$.

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

curve (no. of d.o.f.)	plane Euclidean group (3 d.o.f.)	orthographic projection (5 d.o.f.)	Affine projection (6 d.o.f.)	projective mappings (8 d.o.f)
conic (5)	2	0	0	0
cubic (9)	6	4	3	1
quartic (14)	11	9	8	6
2 coplanar conics (10)	7	5	4	2

Table 1: The number of functionally independent scalar invariants for plane algebraic curves under a variety of groups important in vision. By "orthographic projection", we mean that the plane on which the curve lies is subject to rigid motions in three space, and then projected onto the image plane using orthography. This case has only 5 degrees of freedom because the image curve is completely unaffected by changes in the distance to the object plane.

$Dim(Space) - Dim(LargestOrbit)$ invariants. Any function of these functions is a scalar invariant, and any other scalar invariant is a function of these functions alone. We shall show how this geometric view of invariants can be employed to derive invariant functions.

This argument works as well for differential invariants of parametrised curves. In this case, however, it is essential to consider reparametrisations as well as geometrical camera actions. This is because it is usually impossible to identify a unique starting point or parametrisation measure from image data giving an arbitrary projection of a curve. If we consider a plane curve and n of its derivatives, we have a space of $2 + 2n$ degrees of freedom. If the final expression involves derivatives of the curve up to the n'th, it will also involve derivatives of any reparametrisation up to the n'th. As a result, we need consider only n derivatives of the reparametrisation, but we obtain a reparametrisation group that uses up n degrees of freedom. Assume that the dimension of the "geometric" group of camera actions is m. To obtain a scalar invariant, we must have $m + n < 2 + 2n$, from which expression we can determine n. Note that we need not consider reparametrisation effects when we construct the invariants of algebraic curves under the action of example 3, because the action is on the *coefficients* of the curve, and does not involve any parametrisation.

This counting argument makes it possible to predict the number of scalar invariants we expect under different group actions. In table 1, we give the number of scalar invariants we expect for a plane polynomial of a given order under a number of different groups important in vision. In table 2, we give the number of derivatives we expect to be required for a differential invariant, under the action of these same groups.

In vision applications the full projective group is uncommon and the main effect is that of the Euclidean group acting on an object, which then undergoes perspective distortion. Although the counting argument suggests that more invariants will be available under these circumstances, because the group is smaller, it is in fact not

plane Euclidean group (3 d.o.f.)	orthographic projection (5 d.o.f.)	Affine projection (6 d.o.f.)	projective mappings (8 d.o.f)
2	4	5	7

Table 2: The number of derivatives required for a scalar differential invariant under a variety of groups important in vision. This assumes invariance both to "geometric" actions and reparametrisation.

in general possible to predict the result of a Euclidean movement from the curves observed on the image plane. This means that, although the Euclidean group acts on the **object**, it does not act on the image curves, with resultant complications.

1.2 Constructing invariants

Viewing 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 a submanifold must be normal to that submanifold. Thus, for a scalar invariant Φ, if the vector fields $\mathbf{V}_i(x)$ span the tangent space to the orbit passing through x for all x in the parameter space, then:

$$\mathbf{V}_i . \nabla\Phi = 0, \forall i$$

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]. As a result, we have a mechanical process for constructing invariants of a connected Lie group's action:

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

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, it is possible for the function to be constant on the connected component, but to have different values on distinct connected component. This means that invariants of groups that are not connected

[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 $\mathbf{v}_i$ in this set, exponentiating $\epsilon\mathbf{v}_i$, 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 [14].

are significantly more difficult to compute. We deal only with connected groups, and can largely ignore these difficulties.

1.2.1 Example: Coplanar lines under projective mappings.

Parallel coplanar lines are easily dealt with. A set of lines is parallel on the projective plane if they intersect in a single point. Such lines and collinear points on the projective plane are dual. As a result, four parallel lines yield four collinear points, for which the cross ratio is a well known projective invariant.

The invariants for a system of five general coplanar lines require slightly more work. A line in homogenous coordinates is given by $ax_0 + bx_1 + cx_2 = 0$. We represent this line as $\mathbf{a} = (a, b, c)$. Clearly, we cannot observe a change in scale factor $\mathbf{a} \rightarrow \lambda \mathbf{a}$ and this must be taken into account in deriving the invariants.

We represent the projective group as the set of matrices with determinant one. The group acting will be the projective group, and five copies of the reals as a multiplicative group (the lines must be scaled separately, because we cannot observe their individual scalings). The total dimension of the group is 13, and five lines have collectively 15 degrees of freedom, so we expect two invariants. Given a line $\mathbf{a}$ and a group element $\mathcal{G}$, the action of $\mathcal{G}$ on $\mathbf{a}$ is given by $\mathcal{G}^t\mathbf{a}$. The infinitesimal generators are:

$$
\begin{gathered}
a_1\partial_{b_1} + a_2\partial_{b_2} + a_3\partial_{b_3} + a_4\partial_{b_4} + a_5\partial_{b_5}, \\
a_1\partial_{c_1} + a_2\partial_{c_2} + a_3\partial_{c_3} + a_4\partial_{c_4} + a_5\partial_{c_5}, \\
b_1\partial_{a_1} b_2\partial_{a_2} b_3\partial_{a_3} b_4\partial_{a_4} b_5\partial_{a_5}, \\
b_1\partial_{c_1} + b_2\partial_{c_2} + b_3\partial_{c_3} + b_4\partial_{c_4} + b_5\partial_{c_5}, \\
c_1\partial_{a_1} c_2\partial_{a_2} c_3\partial_{a_3} c_4\partial_{a_4} c_5\partial_{a_5}, \\
c_1\partial_{b_1} + c_2\partial_{b_2} + c_3\partial_{b_3} + c_4\partial_{b_4} + c_5\partial_{b_5}, \\
a_1\partial_{a_1} - b_1\partial_{b_1} + a_2\partial_{a_2} - b_2\partial_{b_2} + a_3\partial_{a_3} - b_3\partial_{b_3} + \\
a_4\partial_{a_4} - b_4\partial_{b_4} + a_5\partial_{a_5} - b_5\partial_{b_5}, \\
a_1\partial_{a_1} - c_1\partial_{c_1} + a_2\partial_{a_2} - c_2\partial_{c_2} + a_3\partial_{a_3} - c_3\partial_{c_3} + \\
a_4\partial_{a_4} - c_4\partial_{c_4} + a_5\partial_{a_5} - c_5\partial_{c_5}
\end{gathered}
$$

Scaling considerations are dealt with by the following five infinitesimal generators:

$$
a_i\partial_{a_i} + b_i\partial_{b_i} + c_i\partial_{c_i}
$$

Write $\mathbf{l}_i$ for $\{a_i, b_i, c_i\}^T$, and I_{ijk} for the determinant of the matrix $\{\mathbf{l}_i, \mathbf{l}_j, \mathbf{l}_k\}$. The invariants are found to be:

$$
I_{l1} = (I_{431}I_{521})/(I_{421}I_{531})
$$

and

$$
I_{l1} = (I_{421}I_{532})/(I_{432}I_{521})
$$

2 APPLICATION TO OBJECT RECOGNITION

In this section, we show that algebraic invariants can be used to recognise planar, curved objects, or objects containing planar curves. Recognition using algebraic invariants requires that image data can be represented using algebraic curves, with 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.

This frame independence property means that we can associate an algebraic curve with the data set in a projectively invariant manner. The algebraic curve becomes a projectively invariant representation. A representation with this independence property need not be a good *approximation* to the data. It is then possible to compute algebraic invariants of the fitted curve, and use these as projectively invariant descriptors for the curve. For applications in model based vision, it is far more important that a representation be projectively invariant than that it be a good approximation. In [6], we showed how a representation with this frame independence property could be constructed, using a theorem we call "the invariant fitting theorem". We use the techniques of this paper here.

For two sets of coplanar data points, the joint scalar invariants of section 2, example 5, computed from projectively invariant conic approximations are constant whatever the camera viewpoint. These joint scalar invariants yield a projectively invariant descriptor, for any data set. In practice, this descriptor is stable, and has sufficient dynamic range to be useful (see [6]). Furthermore, by computing these joint scalar invariants for a pair of curves from two different views, and comparing the results, we can successfully tell whether the curves are coplanar or not by checking if the invariants have changed significantly [6].

Since we have a projectively invariant fitting system, any pair of plane curves can be represented using these joint scalar invariants. Thus, joint scalar invariants can be used to distinguish between model instances, *even if the model does not contain plane conics.* Given an object that has a pair of coplanar curves that will both be visible at the same time, one may compute the joint scalar invariants for the conics fitted to these curves. These invariants then form a projectively invariant description of this set of plane curves. It is then possible to find instances of a model in an image by fitting conics to every available curve, computing the joint scalar invariants for each pair of conics, and then extracting those pairs of curves that have the appropriate set of joint scalar invariants. In fact, this program prunes candidate curves in the same way that the label recognition program does.

This model based vision system is intrinsically fast, and capable of indexing into large model bases quickly (several examples appear in figure 1: in figure 2, we use invariants to index into a parametrised model base). It is sensitive to occlusion, however ([15] discusses noise issues that appear in fitting conics to small numbers of data points).
These techniques for recognising planar objects can be applied to 3D objects which happen to have a collection of plane curves. We have used the invariant for four parallel lines, discussed briefly in section 1.2.1, to recognise pallets in images. Figures 3 and 4 demonstrate this approach. Further techniques for recognising 3D objects by using simple concepts from algebraic geometry to exploit the strong relationship that exists between a polynomial surface and its outline, are described in [9].

3 DISCUSSION

We have shown a selection of applications of techniques from invariant theory to vision. In particular, these techniques make it possible to index into large model bases quickly.
Many extensions and ramifications on this work are possible. In particular, we have shown a framework within which it is possible to compute the invariants of a wide variety of group actions. Our invariant fitting theorem makes it possible to exploit invariants, such as the joint scalar invariants of a pair of conics, that depend on the global properties of a data set. This theorem is important, because it ensures that the fitted curve is, in a sense, unaffected by the frame in which the curve is fitted. This property is essential in applying invariants that have non-local support. Algebraic curves of higher order possess a richer invariant theory that has yet to be exploited.
The representation we have been using is extremely sparse; one number for a pallet, two numbers for a gear. As a result, ambiguities are likely. One way of increasing the representation's robustness to ambiguities involves computing joint invariants for systems of curves - for example, using four or five conics to describe an object.

Acknowledgements

Our thanks to Michael Brady and the Robotics Research Group, Oxford University, for providing a pleasant environment in which to think and work. We thank Dr. Robert J. Baston of the Oxford University Mathematical Institute, who spent time discussing aspects of this work with DAF. Charles Rothwell skilfully and efficiently implemented the label and model recognition systems described in section 2. Our thanks to Andrew Blake and Margaret Fleck for a number of stimulating discussions. We thank GE for providing telephone conferencing facilities during the development of this work.

References

[1] Arnold, V.I. *Geometrical methods in the study of ordinary differential equations*, Grundlehren der Mathematische Wissenschaft 125, Springer-Verlag, 1982.

[2] Asada, H. and Brady, M. "The Curvature Primal Sketch", Proc. IEEE Workshop on Computer Vision: Representation and Control, 1984.

[3] Boothby, W., *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, 1986.

[4] Canny, J.F. "Finding edges and lines in images," TR 720, MIT AI Lab, 1983.

[5] Dickson, L.E., *Algebraic invariants*, John Wiley and sons, New York, 1913.

[6] Forsyth, D.A, Mundy, J.L., Zisserman, A.P. and Brown, C.M. "Projectively invariant representations using implicit algebraic curves, " *Proc. 1st European Conference on Computer Vision*, Springer Verlag Lecture Notes in Computer Science, 1990.

[7] Forsyth, D.A., Mundy, J.L. Zisserman, A.P. and Brown, C.M. "Invariance- a new framework for vision," *submitted to 3rd International Conference on Computer Vision, 1990.*

[8] Forsyth, D.A., Mundy, J.L., Zisserman, A.P. and Brown, C.M. "Projectively invariant data representation by implicit algebraic curves: a detailed discussion," Internal report, Department of Engineering Science, Oxford University, 1990 (In preparation).

[9] Forsyth, D.A., Mundy, J.L., Zisserman, A.P. and Brown, C.M. "Invariance - A New Framework for Vision," Oxford University Department of Engineering Science Internal Report OUEL 1830/90, 1990

[10] Grace, J.H. and Young, A. *The algebra of invariants*, Cambridge University Press, Cambridge, 1903.

[11] Kung, J.P.S. and Rota, G.-C. "The invariant theory of binary forms," *Bull. Amer. Math Soc.*, **10**, 27-85, 1984

[12] Kanatani, K., *Group Theoretical Methods in Image Understanding*, Springer, 1990.

[13] Lane, E.P. *A treatise on projective differential geometry*, University of Chicago press, 1941.

[14] Olver, P.J. *The application of Lie groups to differential equations*, Springer Graduate texts in mathematics 107, Springer Verlag, 1986.

[15] Porrill, J. "Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter," Proc. 5'th Alvey vision conference, 1989.

[16] Rabiner,L.R., and Levinson,S.E. "Isolated and connected word recognition - theory and selected applications," *IEEE Transactions on Communications*, Vol. 29 Num. 5 (May 1981) pp.621-659. referenced in "Integrating Stereo And Motion Curve Matching," A.T. Brint, J.M. Brady

[17] Springer, T.A. *Invariant theory*, Springer-Verlag lecture notes in Mathematics, **585**, 1977.

[18] Weiss, I. "Projective invariants of shapes," *Proc. DARPA IU workshop*, 1125-1134, 1989.

List of figures

1. Output of model based vision system. Note that outline curves used to model the gear, the spanner (wrench) and the scissors were not conics; thus, this approach works for non-conic data in cluttered scenes.
2. Invariant models can be parametrised, and the model parameter recovered directly from the image. Here a model of a pair of scissors is parametrised by the angle between the blades. The value of the joint scalar invariant of the two conics representing the handles for different different values of the model parameter is easily calculated. The model and its parameter is then recovered by testing the joint scalar invariants of pairs of curves against the predicted range of invariants for the parameter, and solving for the angle. We show here two examples for a pair of scissors: because errors in fitting propagate to make the solution for the angle from the invariant imprecise, we have chosen to specify the angle qualitatively.
3. The model used for a pallet consists of the four parallel lines shown in the figure above. These lines yield a single invariant.
4. Two pallets found in a real image, by finding sets of four parallel lines and marking those sets that have the correct value of the invariant.

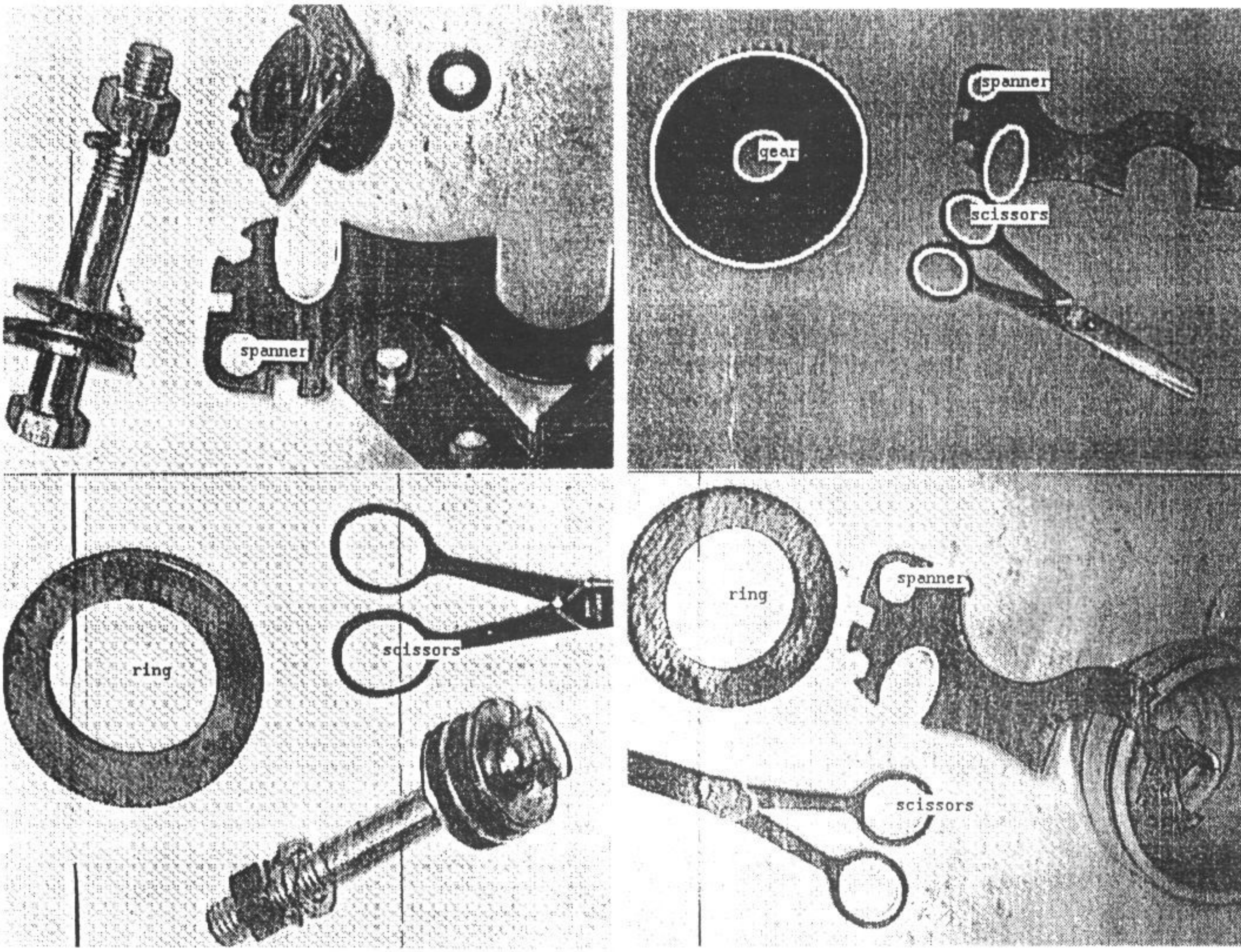

Figure 1

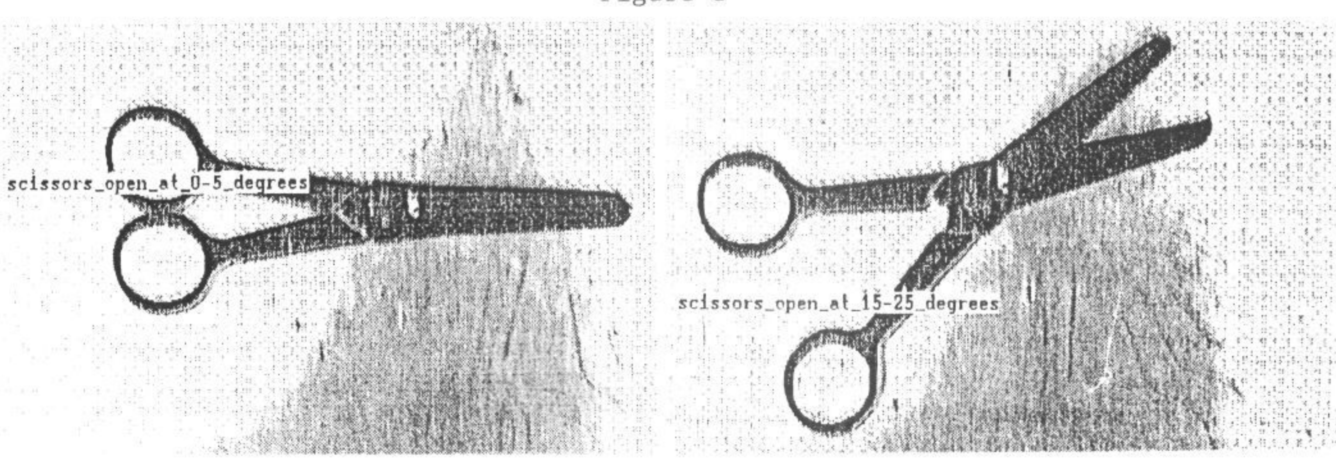

Figure 2

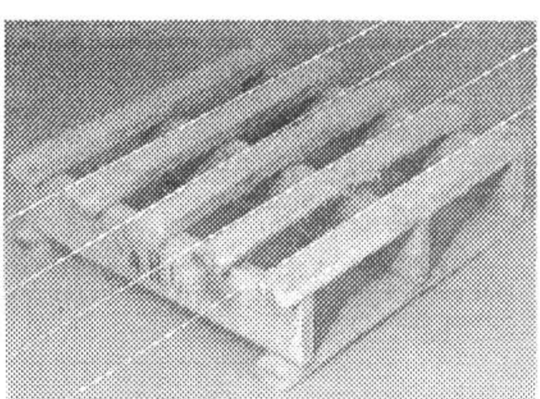

Figure 3

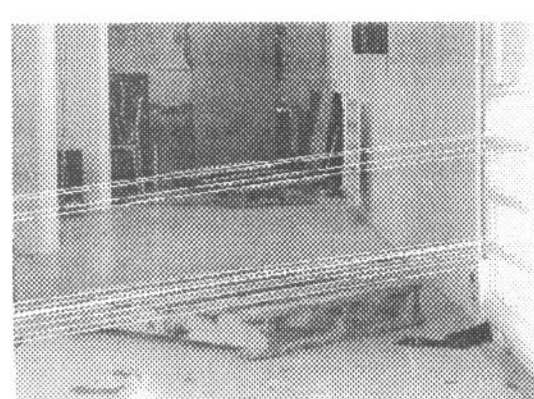

Figure 4