Class-Based Grouping in Perspective Images

Andrew Zisserman, Joe Mundy, David Forsyth, Jane Liu
Nic Pillow, Charlie Rothwell, Sven Utcke
Robotics Research Group, Dept of Engineering Science
University of Oxford, Oxford OX1 3PJ, UK.

Abstract

In any object recognition system a major and primary task is to associate those image features, within an image of a complex scene, that arise from an individual object. The key idea here is that a geometric class defined in 3D induces *relationships in the image which must hold between points on the image outline (the perspective projection of the object). The resulting image constraints enable both identification and grouping of image features belonging to objects of that class.*

The classes include surfaces of revolution, canal surfaces (pipes) and polyhedra. Recognition proceeds by first recognising an object as belonging to one of the classes (for example a surface of revolution) and subsequently identifying the object (for example as a particular vase). This differs from conventional object recognition systems where recognition is generally targetted at particular objects. These classes also support the computation of 3D invariant descriptions including symmetry axes, canonical coordinate frames and projective signatures.

The constraints and grouping methods are viewpoint invariant, and proceed with no information on object pose. We demonstrate the effectiveness of this class-based grouping on real, cluttered scenes using grouping algorithms developed for rotationally symmetric surfaces, canal-surfaces and polyhedra.

1 Introduction

A number of recent model-based recognition papers have established that for certain model *classes* [19], invariants of 3D objects can be obtained from single perspective images. Examples include surfaces of revolution [4], symmetric objects [14], and canal surfaces [11]. The invariants are numbers measured from the image which are unaffected by viewpoint. The invariants are used to identify models in a library, and thereby generate recognition hypotheses for particular objects. In this paper we demonstrate that for such model classes there are exact invariant relationships (transformations) on the object outline in any image. These relationships can be harnessed to *group* outline curves.

As is well known, in building any automatic recognition system working with images of real scenes acquired under ambient conditions, a significant barrier to successful recognition is the extraction of feature groups which correspond to individual object boundaries. Outline curves found using state-of-the-art edge detectors will have 'drop outs', incomplete and incorrect topology and extraneous segments. To overcome such problems, a grouping stage is incorporated that typically involves associating curve segments based on a combination of proximity, connectivity and 'heuristics'. Here it is shown that class-based grouping can provide a principled basis for such heuristics and provide a link between class constraints and feature processing at all representational levels. Three object classes are covered in this paper: two particular types of Generalized Cylinder (GC) [1], and polyhedra.

A number of other authors have utilized image constraints on profiles[1]. Ponce [12] proved that tangent lines constructed at corresponding points on each side of the profile of a Straight Homogeneous Generalized Cylinder (SHGC) intersect on the image of the cylinder axis. This relation holds under perspective viewing. He used this constraint to find the projected axis by constructing a Hough space of potential axis points. This algorithm is of complexity $O(n^2)$ where n is the number of edgel points on the boundary. This complexity was reduced in a second algorithm based only on matching inflection points on the profile. A similar approach was followed by Gross and Boult [5]. Related work on SHGCs [3] has focussed on the problem of pose recovery using a calibrated camera and does not address the grouping problem.

Zerroug and Nevatia [18] significantly advanced

[1] The *profile* is the outline of a surface in the image. It is the image projection of the *contour generator*, which is the set of surface points where rays through the optical centre are tangent to the surface.

grouping for SHGCs by using a single cross-section of the cylinder as a grouping template which is swept tangent to putative profile curve portions. In this manner complete profiles of GCs (and thence their volumetric descriptions) have been recovered from fragments of the outline in the presence of occlusion, clutter and extraneous edges. The computational cost of this approach is reduced to $O(n)$ since the cross-section constraint establishes symmetrical correspondences on each side of the cylinder axis.

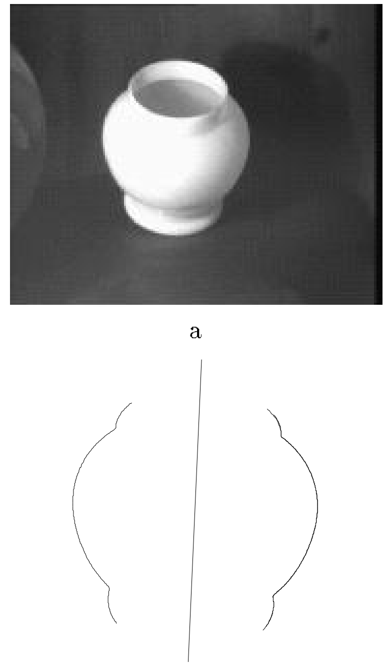
Their algorithm has two limitations which are overcome in the work described below. First, the cylinder cross-section must be visible in the image in order to achieve the grouping. In cluttered scenes with occlusion and due to self-occlusion this limitation can be prohibitive since there are many viewpoints where the cross-section cannot be recovered. Second, the method assumes weak perspective projection, which is a reasonable assumption when the object depth is small compared to the viewing distance, but breaks down when an object is close to the camera and is oriented so that its depth change is significant.

We show that grouping complexity can be substantially reduced over prior approaches without dependence on the image projection of a cross-sectional curve. Complexity does not depend on some power of the number of edgels, n, since the grouping is based on curve matching indexed by invariants derived from GC properties. The current algorithms cover surfaces of revolution and canal surfaces, and in these cases the grouping complexity is reduced to $O(m)$ where m is the number of invariants to be matched (a few tens) for each curve segment.

In the following we are careful to state clearly the camera assumptions made in deriving a particular invariant constraint or image relation. The main distinction between cameras is whether or not parallel projection is assumed. A subsidiary distinction is whether or not the internal camera parameters, such as aspect ratio, are assumed. Many prior results for generalized cylinders depend on the assumption of weak perspective (i.e. parallel projection with known aspect ratio) and treat perspective effects as perturbation.

2 Surfaces of Revolution

This is a surface constructed by rotating a planar generating curve about a symmetry axis. Unfortunately, the profile curve is not related in any straightforward invariant manner to the planar generating curve of the SOR. It can be shown that the contour generator of a SOR is, in general, a space curve even under parallel projection.



a

b

Figure 1: The two sides of the profile of a surface of revolution are related (to an excellent approximation) by a particular type of affine transformation (an involution). This transformation is computed using an affine snake, as described in the text. (a) The profile of this SOR is shown in (b), with the transformed left side superimposed. The two curves (original right profile, and transformed left) are indistinguishable, demonstrating the accuracy of this transformation on real images.

2.1 Profile Relationship

The profile (image outline) of a surface of revolution can be separated into two 'sides' by the projected symmetry axis. The two sides are tightly constrained — they are related by a particular four degree of freedom plane projective transformation — a planar harmonic homology [9]. This relationship is exact. The transformation is represented by a non-singular 3×3 matrix $\mathtt{T}$, where $\mathtt{T}^2 = \mathtt{I}$ [16]. Two pairs of point correspondences determine $\mathtt{T}$. This transformation is fundamental to processing profiles of SORs: the line of fixed points of $\mathtt{T}$ is the imaged symmetry axis; $\mathtt{T}$ provides point to point correspondence between the sides of the profile; this disambiguates the matching of bitangents used to form the invariants; and finally, $\mathtt{T}$ can be used to *repair* missing profile portions, filling in gaps by transforming over points from the other side of the profile.

Based on these properties, grouping for this class could be carried out by associating curves which are projectively equivalent, and then testing if the transformation between two projectively related curves is a harmonic homology. If it is not then the associated curves can be ruled out as members of this class. This is simply tested by checking if $\mathtt{T}^2 = \mathtt{I}$.

Provided the field of view is not too large, $\mathtt{T}$ can be approximated by a three degree of freedom affine transformation (an involution) [7]. This is an excellent approximation[2] (as demonstrated empirically below). A decomposition of the affine transformation is given in [9].

2.2 Grouping Mechanism

Grouping involves determining which curves could have arisen from a surface of revolution, and repairing, where possible, missing curve segments due to occlusion and the usual problems of segmentation. In particular profile curves are often broken into a number of portions. Since the image may well contain several surfaces of revolution the grouping stage also partitions the profile curves into those arising from different surfaces. Grouping proceeds in three stages:

(i) Associate conjugate curve fragments The two sides of the profile are affine related, and consequently corresponding profile curves have the same affine invariants. Associating conjugate (corresponding) curve fragments is then a matter of matching curve segments with the same affine invariants. Here area of profile concavities is used as the (relative) affine invariant. The matching has complexity m^2 in the m concavities, where typically $m = 25$ for a cluttered scene containing two surfaces of revolution.

(ii) Select profile fragment pairings A number of these putative matches can then be eliminated because the affine transformation between the curves is not an involution.

The transformation is determined using a type of *affine snake*. This is computed in two stages: first, an approximate solution is determined by matching a number of distinguished points (such as bitangent contact points); second, the approximate solution is used to transform a number of sample points from one profile side to the other. The squared distance between the transformed sample points and the other profile side is then minimized numerically over the three parameters of the transformation. Effectively this uses one side of the profile to define a snake, and then determines the affine transformation (constrained to have the above properties) which most closely aligns it with the other side (see figure 1).

(iii) Group profile fragments It then remains to group curve fragments which may have arisen from the same profile curve. Grouping is based on the similarity of the three parameters of the affine transformation (i.e. the symmetry axis and correspondence direction). Using these parameters the paired SOR concavities are partitioned into sets, and the associated profile curve fragments are joined using existing local edgel chain topology and smooth curve continuation.

Further details are given in [20], and examples of the grouping process in figures 3 and 4.

3 Canal Surfaces

A canal surface can be thought of as the surface generated by a circle swept along the axis, and also as the envelope of a *sphere* swept along the axis. The latter model provides a more direct route to profile properties because the projection of a sphere is more simply described than that of a circle. It can be shown that the contour generator of a canal surface is, in general, a space curve even under parallel projection.

3.1 Profile Relationship

Weak perspective projection We make strong use of the fact that 'the envelope of the profile is the profile of the envelope' [15]. Under parallel projection the profile of a sphere is a circle, and the sphere centre projects to the circle centre. It follows that the profile of a canal surface is an envelope of circles. These circles can be recovered in the image from the profile by constructing the symmetry set [2, 6], the locus of centres of circles bitangent to the profile. Since a single scaling applies for weak perspective imaging, the circles have equal radius. There are a number of consequences of these results [11]. For example, the contact of image circles with the profile identifies corresponding points on the surface, i.e. points which lie on the same circular cross-section. The curve traced out by the circle centres is the projection of the sphere centres, i.e. of the canal surface axis.

Perspective projection Under perspective projection a sphere profile is an ellipse and the sphere centre does not project to the ellipse centre. It follows that the two profile 'sides' — the ellipse envelope — are not parallel curves (as they are in the weak perspective case). Figure 2 illustrates the profile relationships which do hold under perspective imaging in the case of a *planar* axis curve. These results also hold under weak perspective imaging — though in this case the vanishing line $\mathbf{l}_\infty$ is at infinity in the image.

[2] If the image aspect ratio is correct, then under weak perspective the profile sides are related by a mirror symmetry [10]. The restricted affine transformation used here (only one more parameter in this case than a mirror symmetry) covers the case of unknown aspect ratio and, more importantly, perspective effects.

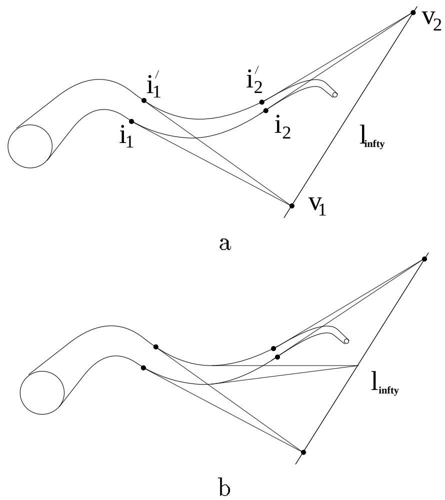


Figure 2: For a canal surface, inflections in the profile occur in pairs for each inflection of the axis. The intersection of a pair of inflection tangents determines the vanishing point of the tangent line at the axis inflection. If the axis curve is planar, then (a) two such vanishing points determine the vanishing line, $\mathbf{l}_\infty$, of the plane of the axis; and (b) corresponding profile tangents (profile points arising from the same surface circular cross-section) also intersect on $\mathbf{l}_\infty$. Their intersection point is the vanishing point of the corresponding axis tangent line.

These relationships are not as powerful as the projective transformation between profile sides of a SOR. For a canal surface, the relationship constrains only profile tangents (a constraint on the tangent dual space), whereas for an SOR there is a point to point transformation.

3.2 Grouping Mechanism

The results can be harnessed at two stages of grouping: first, at a 'boot-strap' stage, the correspondence between points on either side of the profile is unknown. It is here that the inflection result is utilized because an inflection is a distinguished point which can be localized on a curve. By grouping curve portions with two pairs of inflections, a putative vanishing line for the canal surface projecting to those curves is computed (as the line through the intersection of corresponding profile tangents). This hypothesised vanishing line can be used to group further profile inflection tangents (i.e. the pairs of tangents must intersect on this line). Second, the result that corresponding profile tangents intersect on $\mathbf{l}_\infty$ is used to *verify* putative curve associations. Provided a sufficient extent of the curves between the inflections have tangents which intersect on $\mathbf{l}_\infty$, then the hypothesized association of inflections is verified.

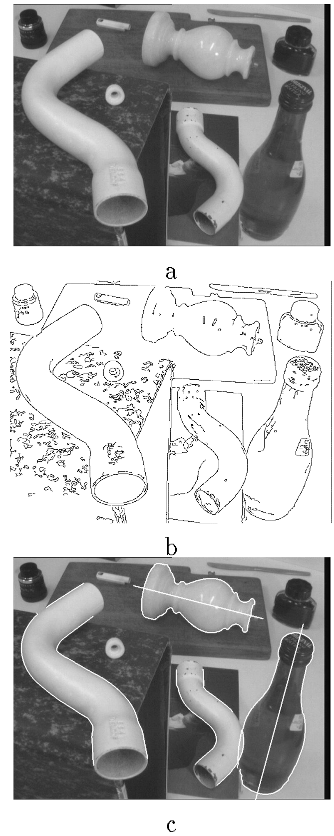


Figure 3: (a) Original image containing two SORs and two canal surfaces. (b) The linked edges computed from (a), note the gaps and incorrect topology. Profiles are extracted and grouped from these linked edges by the class-based groupers. (c) Extracted SOR and canal surface profiles superimposed on the original image. Note, all the correct instances of a class have been grouped, and no false instances grouped. Also, gaps — from edge detector drop out or occlusion — in one side of the profile, shown in (b), have not prevented grouping, and, for the SORs, have been 'repaired'.

The success of this hypothesize and verify strategy is illustrated in figure 3. Details are given in [20].

4 Polyhedra

The polyhedral class can be divided into a number of simple sub-classes, and class specific groupers developed for each sub-class. Two sub-classes are considered here: first, polyhedra whose edges are in one of only three directions (such as a cube); and, second, triangular prisms. Both have trihedral vertices.

4.1 Grouping Mechanism

Lowe [8] describes a grouping constraint for polyhedra, which arises from parallelism of edge segments in

3D. Under affine imaging conditions the corresponding image segments are also parallel. We use this constraint during grouping, but finally impose the correct internal constraints for the polyhedra[3].

The initial stage of grouping aims to extract a complete silhouette curve for a polyhedron. This stage uses typical techniques, including collinear edge extension and vertex definition by intersection of nearly incident edge segments. A polygonal description is then extracted from this complete edgel curve by segmenting into straight sections using a 'worm' corner finder, and line fitting to the sections. Line segments that are near collinear and have endpoints nearby are fused, to give the polygonal silhouette of a polyhedron. This is compared with the possible polyhedral silhouettes for each sub-class, to generate a strategy for filling in internal boundaries. The number of line segments in the silhouette distinguishes between the two existing cases. This outline then generates a sub-class polyhedral snake which incorporates the complete polyhedral structure, including the missing interior boundaries. For the two polyhedral sub-classes implemented, inferring interior boundaries from the silhouette is easy. Finally, the position of the superimposed wireframe snake is polished by minimizing the normal distance of selected points on the snake to peaks in the image gradient map, subject to the required internal constraints of the polyhedron.

Figure 4 includes SORs, canal surfaces and polyhedra. All profiles and outlines are extracted and grouped automatically, and there are no false positives.

5 Conclusions

3D classes based on volumetric primitives lead to powerful internal constraints on the structure of their profiles; these constraints can be used both to generate and to test grouping hypotheses. The grouping methods are either invariant to perspective effects, or make principled approximations at the correct stage, so that perspective is not simply treated as a perturbation. Grouping is completely automatic — the same thresholds are applied to all the images in this paper. Grouping has been demonstrated on images of cluttered scenes where the target object may be partially occluded.

There are a number of ways in which this work can be extended. First, similar constraints can be derived for other geometrically defined object classes. For example, between the sides of the profile of a general SHGC. It is worth noting that under perspective projection imaged cross-sections of an SHGC are related by a planar homology [20]. This enables the generalisation of Zerroug and Nevatia's grouping procedure to perspective. Second, the grouping methods can be extended to cope with even more extensive occlusion by the use of more local invariants. For example, in the case of SORs the local invariant identified by Ponce [13] for skewed ribbons, or the semi-local invariant [17] could be used to initialize matching.

[3]For example, a cube has three major directions which define a triple of vanishing points in the image. All edges aligned with a major direction must pass through the same vanishing point.

Acknowledgements

We are grateful for a number of discussions with Peter Giblin. Financial support was provided by EU Esprit Project 6448 'VIVA', a UK EPSRC Grant and General Electric, USA.

References

[1] Binford T.O., 'Inferring Surfaces from Images'. *Artificial Intelligence*, vol. 17, pp. 205–244, 1981.

[2] Bruce J.W., Giblin P.J. and Gibson C.G., 'Symmetry sets', *Proc. of Royal Soc. of Edinburgh*, 101A, 163–186, 1985.

[3] Dhome, M., Lapreste, J.T., Rives, G. and Richetin, M. 'Spatial Localization of Modelled Objects of Revolution in Monocular Perspective Vision,' *Proc. ECCV*, LNCS 427, Springer-Verlag, p.475–485, 1990.

[4] Forsyth D.A., Mundy J.L., Zisserman, A. and Rothwell C.A., 'Recognising rotationally symmetric surfaces from their outlines'. *Proc. ECCV*, LNCS 588, Springer-Verlag, 1992.

[5] Gross A.D. and Boult T.E., 'Recovery of Generalized Cylinders from a Single Intensity View', *Proc. of Image Understanding Workshop*, 1990.

[6] Giblin P.J. and Brassett S.A., 'Local Symmetry of Plane Curves'. *American Mathematical Monthly*, vol. 92, no. 10, pp. 689–707, 1985.

[7] Liu J.S., Mundy J.L., Forsyth D.A., Zisserman A. and Rothwell C.A., 'Efficient Recognition of Rotationally Symmetric Surfaces and Straight Homogeneous Generalized Cylinders', *Proc. CVPR*, 1993.

[8] Lowe, D.G. *Perceptual Organization and Visual Recognition,* Kluwer Academic Publishers, 1985.

[9] Mukherjee D.P., Zisserman A. and Brady J.M., 'Shape from symmetry—detecting and exploiting symmetry in affine images', Univ. of Oxford, Dept. Eng. Sci. Report No. OUEL 1988/93, 1993. To appear, *Proc. Royal Soc.*, 1995.

[10] Nalwa V., 'Line drawing interpretation: Bilateral symmetry' *IEEE PAMI*, vol 11, 10, 1989.

[11] Pillow N., Utcke S. and Zisserman A. 'Viewpoint-Invariant Representation of Generalized Cylinders Using the Symmetry Set', *Proc. BMVC*, 1994. To appear *Image and Vision Computing*, 1995.

[12] Ponce J., 'Invariant properties of straight homogeneous generalized cylinders'. *IEEE PAMI*, vol. 11, no. 9, pp. 951–965, 1989.

[13] Ponce J., 'On characterizing ribbons and finding skewed symmetries'. *Proc. Robotics and Automation*, vol. 1, 49–54, 1989.

[14] Rothwell, C.A., Forsyth, D.A., Zisserman, A. and Mundy, J.L. "Extracting Projective Structure from Single Perspective Views of 3D Point Sets", *ICCV*, 573–582, 1993.

[15] Rycroft J.E. *A Geometrical Investigation into the Projections of Surfaces and Space Curves*, Ph. D. Thesis, University of Liverpool, 1992.

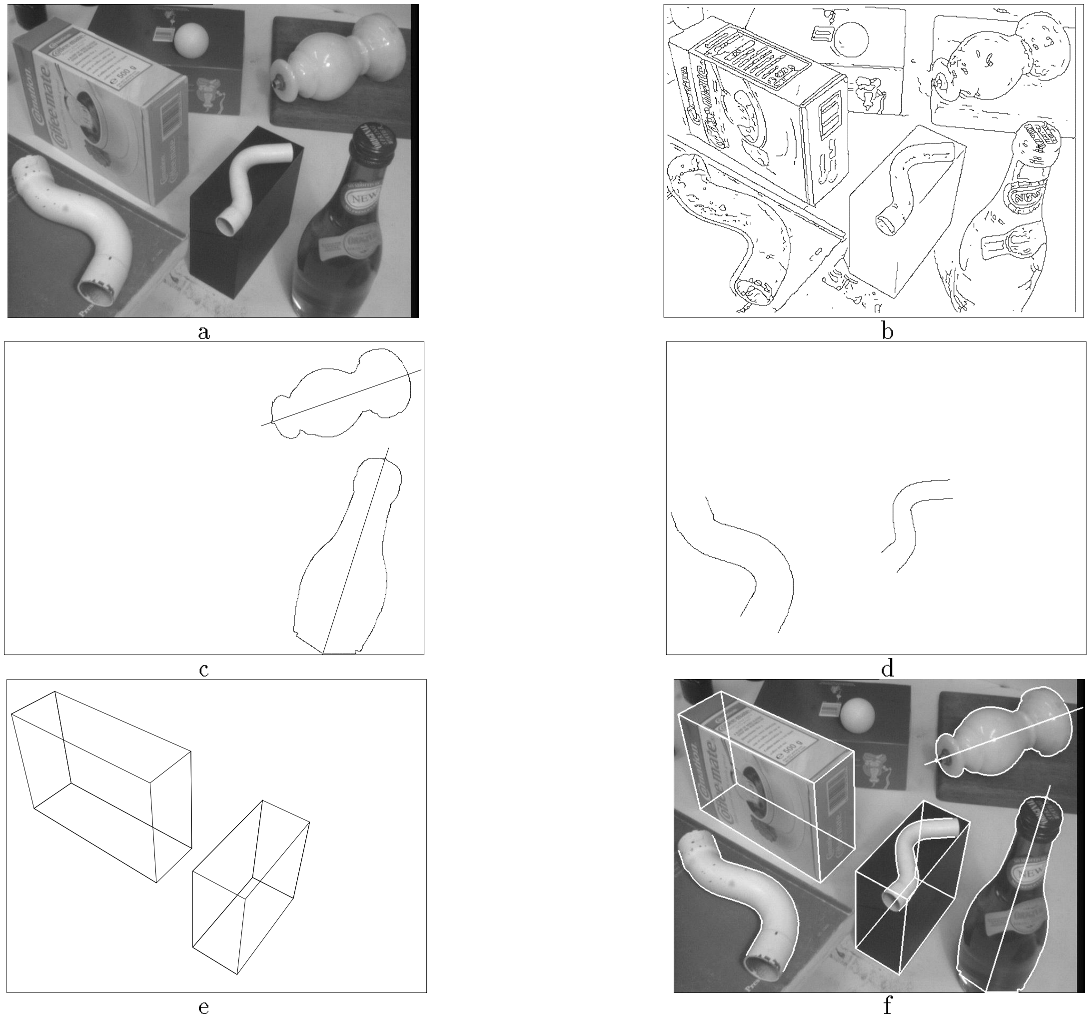


Figure 4: (a) Original image containing two SORs, two canal surfaces, and two polyhedra. (b) The linked edges computed from (a). Profiles are extracted and grouped from these edges by the class-based groupers. (c) Extracted SOR profiles with axes. (d) Extracted canal profiles. (e) Extracted polyhedra outlines. (f) Extracted profiles superimposed on original image. Again, all the correct instances of a class have been grouped, and no false instances grouped.

[16] Springer C.E., *Geometry and Analysis of Projective Spaces.* Freeman, 1964.

[17] Van Gool, L., Moons, T., Van Diest, M. and Pauwels, E. 'Matching perspective views of parallel plane structures', in *Applications of Invariance in Computer Vision*, Mundy, J.L., Zisserman, A. and Forsyth, D.A. (eds), LNCS 825, Springer-Verlag, 1994.

[18] Zerroug M. and Nevatia R., 'Using invariance and quasi-invariance for the segmentation and recovery of curved objects'. In Mundy J.L., Zisserman A., and Forsyth D.A., editors, *Applications of Invariance in Computer Vision*, LNCS 825, Springer-Verlag, 1994.

[19] Zisserman A., Forsyth D., Mundy J., Rothwell C., Liu J. and Pillow N., '3D Object Recognition using Invariance', *Oxford University Engineering Report*, OUEL 2027, 1994. To appear *AI Journal.*

[20] Zisserman A., Mundy J., Forsyth D., Liu J., Pillow N., Rothwell C. and Utcke S., 'Class-Based Grouping in Perspective Images', *Oxford University Engineering Report*, 1995.