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

A major issue for object recognition systems is the organization of object structure and separation of individual objects within an image of a complex scene. This paper demonstrates that general geometric object classes can be defined which provide effective constraints for grouping image features into coherent object boundaries. These classes also support the computation of 3D invariant descriptions including symmetry axes, canonical coordinate frames and projective signatures.

The key idea is that a 3D geometric class defines relations which must hold between points on the image outline (the perspective projection of the object's surface). 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 classifying the object (for example as a particular vase). This differs from conventional object recognition systems where recognition is generally targetted at particular objects.

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 canal-surfaces, rotationally symmetric surfaces and polyhedra.

1 Introduction

A number of recent model based recognition papers have established that for certain model *classes* invariants of 3D objects can be obtained from single perspective images. Examples include SOR [5], symmetric objects [18], and canal surfaces [15]. 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 cases of Generalized Cylinders (GC) [1]; and polyhedra.

In the following we use the terms:

- **profile:** the outline of a surface in the image. The profile is the set of image points where rays through the optical centre are tangent to the surface.
- **contour generator:** The curve on the surface which projects to the image profile. The contour generator is the set of points on the surface where the tangent plane contains the optical centre.

A number of other authors have utilised image constraints on profiles. Ponce [16] proved an important relation concerning Straight Homogeneous Generalized Cylinders (SHGC's) which holds under perspective viewing. He showed that tangent lines constructed at corresponding points on each side of the profile intersect on the image of the cylinder axis. Ponce used this fact 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 [6]. Related work on SHGC's [4] has focussed on the problem of pose recovery using a calibrated camera and does not address the grouping problem.

Zerroug and Nevatia [23] made a major advance in grouping for SHGC's. A single cross-section of the cylinder is used as a grouping template by sweeping it 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 include 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.

1.1 Class based recognition

In this paper we explore the grouping power of a number of geometrically defined classes. These classes are:

1. **Surfaces of Revolution (SOR):** a special case of a Straight Homogeneous Right Generalized Cylinder. Examples are bottles and vases.

2. **Canal Surfaces:** a generalised cylinder where the swept curve is a circle of constant radius, with the plane of the circle orthogonal to the axis curve, and with the circle centre on the axis. More formally they are Circular Planar Right Generalized Cylinders with constant cross-section. The

canal surfaces considered here are restricted to have a *planar* axis curve. Examples are pipes.

3. **Polyhedra**

These three classes cover a large number of manufactured objects and provide a good approximation to many natural objects, such as tree branches and crystals.

Recognition is comprised of two major processes. The first is grouping or feature organization which is directed by the constraints provided by a particular geometric class. The second stage is classification which is achieved by using invariant indices derived for specific instances of the class.

For example, if a scene contains several SORs, the grouper first identifies curve pairings that could have arisen from a SOR (they satisfy a conjugate symmetry constraint, see section 2.1), and then groups those arising from the same SOR. Such identification and grouping is possible because the 2D image curve which results from imaging the 3D class is tightly constrained. In turn these constraints can be used during grouping to test the validity of the class assumption. Subsequently, the organized SOR boundaries can be used to derive distinguished points on the axis of symmetry which provide invariant indexing using cross-ratios computed over combinations of the points taken four at time.

Indexing the model library via the invariants generates recognition hypotheses for particular *models* of that class. For example, if an SOR profile had been grouped the SOR invariants might index particular models, e.g. vase #2 or bottle #4, and recognition hypotheses for these models would then be verified by projecting the specific model boundary into the scene. We note that distinguished points on the axis are not generally sufficient fully to determine the 3D shape and pose of the object from its profile alone. It is necessary to have a calibrated camera and the Euclidean geometric coordinates of points on the axis to determine the unique projection of a 3D model of the SOR into the image.

1.2 Camera Assumptions

It is important to distinguish between a number of assumptions about the camera used to image the scene. Various results in grouping and recognition of geometric classes depend on such assumptions. Here we distinguish five classes of image projection.

Projective The camera is represented by a 3×4 projection matrix with no constraints on the elements.

Perspective The elements of the 3×4 matrix are constrained to correspond to central projection.

Calibrated Perspective The internal parameters of the camera such as principal point and aspect ratio are known and built into the image coordinate system. The camera matrix takes on the form:

$$\mathtt{P} = [\mathtt{R}| - \mathtt{R}\ \mathbf{t}]$$

where $\mathtt{R}$ is a rotation matrix and $\mathbf{t}$ is a translation vector relating the camera coordinate frame to the world frame.

Affine The matrix is of the form:

$$\mathtt{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{bmatrix}$$

For this and the following model, parallel lines in 3-space project to parallel image lines. Perspective effects, such as convergent vanishing points, do not occcur.

Weak Perspective The affine matrix is constrained so that the projection is orthographic with a constant scale factor. The camera aspect ratio is assumed to be one and object depth is small compared to the viewing distance.

It is important to clarify what camera assumption is used in claiming a particular invariant constraint or image relation. Many prior results for generalized cylinders depend on the assumption of weak perspective. The grouping methodology of Zerroug and Nevatia assumes weak perspective, and treats perspective effects as perturbations. Their methods are still valid with an affine camera, but are not if significant perspective effects are present - as can arise with a projective camera. Other work [4] assumes that a calibrated camera is available.

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 generator curve of the SOR. It can be shown that the contour generator of a SOR is, in general, a space curve under parallel projection.

2.1 Profile Relationship

For completeness we summarise here a number of properties of this class that have appeared elsewhere. These properties hold for the fully general case of the projective camera.

1. Tangents at corresponding points on the two sides of the profile (i.e. images of points on the same circular cross-section) intersect on the projected symmetry axis [16].

2. 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 projective transformation—a planar harmonic homology [12]. This relationship is exact.

This relationship between the two sides of the image outline of a SOR is the basis for the grouping methods developed here. In the following sections we summarise the properties of this transformation. These properties are used in the implementation to ascertain whether particular pairings of image curves could be parts of profiles of a SOR.

To see how this transformation arises consider a canonical image plane which is perpendicular to the plane, Π_S, containing the axis of the SOR and the optical centre. The plane Π_S is a plane of symmetry for the SOR and camera, and consequently the image outline is symmetric with the projected axis being the line of symmetry. Any other image plane is within a plane projective transformation of this canonical image, and so in any image the sides of the profile are related by a projective transformation, $\mathtt{T}$, which is a conjugate symmetry i.e. $\mathtt{T} = \mathtt{P}^{-1}\mathtt{SP}$, where $\mathtt{S}$ is a reflection. This is the same transformation that results from a planar symmetry under perspective projection. The transformation has the property that $\mathtt{T}^2 = \mathtt{I}$, and is known as an *involution*, a *collineation of period two* and a planar harmonic homology [21]. Two pairs of point correspondences determine $\mathtt{T}$. See figure 1.

$\mathtt{T}$ 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 — see section 2.3; 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.

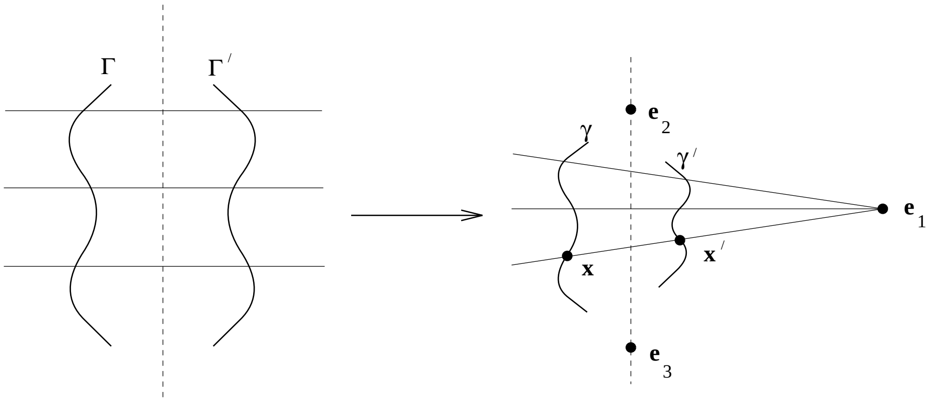

Figure 1: The profile of a surface of revolution is projectively equivalent to two curves with bilateral symmetry. Under a projective transformation parallel correspondences (left) converge to a vanishing point (right). Corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}'$ are related in this case by a particular plane projective transformation, T, called a planar harmonic homology. The transformation has a line of fixed points, the image of the axis of symmetry, which result from two of the eigenvalues of T being equal. There is also a fixed point, $\mathbf{e}_1$, not on the line, called the center of the homology which defines correspondences between symmetrical points on each side of the contour. That is, corresponding point pairs and the center of the homology are collinear. (The line of fixed points is $\mathbf{e}_2 \times \mathbf{e}_3$, where $\mathbf{e}_2$ and $\mathbf{e}_3$ are the eigenvectors with equal eigenvalues. The third eigenvector, $\mathbf{e}_1$, is distinct and non-zero, and is the centre for a pencil of fixed lines).

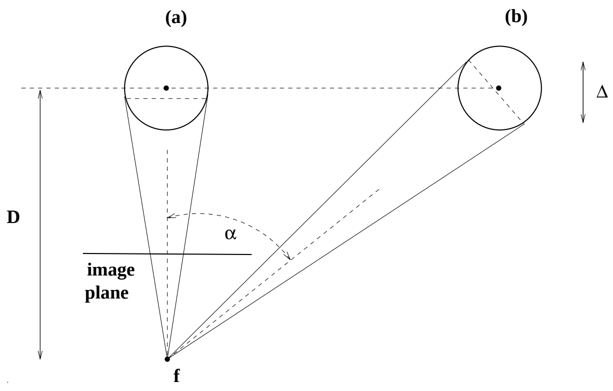

Figure 2: Schematic plan view of a surface of revolution and tangents from the optical centre. α is the angle between the plane containing the axis of the surface and the focal point, and the image plane normal. In (a) lines joining corresponding points on the contour generator are parallel to the image plane. Consequently, their image projections are parallel. This holds even when the axis of the surface is *not* parallel to the image plane. In (b) the object has the same depth but is translated parallel to the image plane. In this case lines joining corresponding points do not project to parallel lines.

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 $\mathtt{T}^2 = \mathtt{I}$. However, under real imaging situations, $\mathtt{T}$ is approximately affine. This affine transformation is implemented for grouping.

2.1.1 Affine approximation

Provided the field of view is not too large, $\mathtt{T}$ is approximately affine [11]. Figure 2 illustrates that the significant angle in terms of the affine approximation to projection is not the inclination of the axis, but the angle between the plane containing the axis and focal point, and the image plane normal, α. It can be shown that Δ/D is proportional to $\tan\alpha$. For a finite image plane, α is bounded by the finite field of view. Consequently, this limits the "extent" of non-parallelism of lines joining corresponding points on the outline.

This is an excellent approximation (as demonstrated empirically below)[1]. See figure 3. It can be shown [12] that the affine transformation has the following properties:

1. Corresponding points $\mathbf{x}'$ and $\mathbf{x}$ on the two profile sides are related by the affine transformation

[1] If the image aspect ratio is correct, then under weak perspective the profile sides are related by a mirror symmetry [14]. 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.

a b c

Figure 3: (a) 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). Vector $\mathbf{a}$ is parallel to the skewed symmetry axis, vector $\mathbf{b}$ is parallel to $\mathbf{x}' - \mathbf{x}$ - i.e. the correspondence direction. A point on the left side is transformed to the right by translating in the correspondence direction twice the distance between the left side and the symmetry axis. (c) A typical example of the accuracy of this transformation on a real image (b). The transformation is computed using an affine snake, as described in the text, and is used to map the left side onto the right. The two curves (original right profile, and transformed left) are indistinguishable. The parameters of the transformation also define the symmetry line shown. In this example the three parameters are $\theta_c = 2.7$, $\theta_s = 87.4$ and $r = 329.8$.

$\mathbf{x}' = \mathtt{A}\mathbf{x} + \mathbf{b}$. (Corresponding points arise from the same circular surface cross-section). The transformation is an involution, so $\mathbf{x} = \mathtt{A}\mathbf{x}' + \mathbf{b}$.

2. The transformation has three degrees of freedom—two to specify the "symmetry" line (the projection of the 3D axis of revolution), and one for the correspondence direction. See figure 3a.

3. The matrix $\mathtt{A}$ satisfies $\mathtt{A}^2 = \mathtt{I}$. It has eigenvectors $\mathbf{a}$ and $\mathbf{b}$ ($\mathbf{b}$ as above) with eigenvalues $+1$ and -1 respectively. Vector $\mathbf{a}$ is parallel to the symmetry axis, vector $\mathbf{b}$ is parallel to $\mathbf{x}' - \mathbf{x}$ - i.e. the correspondence direction.

4. The equation of the symmetry line is: $2a_y x - 2a_x y - a_y b_x + a_x b_y = 0$

This is the transformation that applies in the case of a 'skewed symmetry' [8] - an affine transformation of a planar reflectional symmetry.

2.1.2 Computing the affine transformation

The transformation between the two sides 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 $\{\mathtt{A}, \mathbf{b}\}$. 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. Ten sample points are usually sufficient and the minimization converges within ten iterations generally to an average distance of less than a pixel. A typical example is shown in figure 3c. The transformation is parameterized by: correspondence angle, θ_c; symmetry line angle, θ_s; and distance of the symmetry line from the origin, r_s.

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 partition the profile curves into those arising from different surfaces. Grouping proceeds in three stages:

1. **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.

   The (relative) affine invariant implemented here is the area of a *simple concavity.* A concavity is simple if it contains only two inflections. Candidate simple concavities are obtained by: detecting edges using a Canny operator; linking into edgel chains, extrapolating over one pixel gaps; determining bitangents for each edgel chain, using the method described in [19] which determines the contact of the bitangent to the curve to sub-pixel accuracy by locally fitting a quadratic; finally, simple concavities are selected by eliminating bitangents with intersect the edgel chain. See figure 4d. Area is computed using Simpson's method.

   The matching has complexity m^2 in the m concavities, where typically $m = 25$ for a cluttered scene containing two surfaces of revolution.

#'s	areas pixels^2	av. error pixels	parameters $\theta_c/^o$ $\theta_s/^o$ r/pixels
1 6	181.7 190.8	0.23	5.2 96.5 260.1
1 9	181.7 197.9	no match	
2 5	428.6 442.8	0.43	8.5 98.7 271.3
3 8	242.9 263.1	no match	
4 8	275.1 263.1	no match	
4 10	275.1 277.0	no match	
6 9	190.8 197.9	no match	
8 10	263.1 277.0	0.48	6.5 96.1 257.7

Table 1: Concavities are matched on area (which is a relative affine invariant), followed by the affine snake which enforces involution. Only the symmetry related concavities have a low error. The transformation parameters computed from each matched concavity pair are very similar. These parameters are used to group the concavities.

2. **Select profile fragment pairings** A number of these putative matches can then be eliminated because the affine transformation between the curves does not have the additional properties listed above. The transformation is determined using the affine snake. The conclusion of this stage of grouping is a set of paired curve fragments with the three parameters of the transformation $\{\theta_c, \theta_s, r_s\}$ known in each case.

3. **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 $\{\theta_c, \theta_s, r_s\}$ and corresponding pairs of matched concavities are aligned along their common central axis. The associated outline curve fragments are joined using existing local edgel chain topology and smooth curve continuation.

An example is shown in figure 4 and table 1 of the concavities and matching process. The transformation parameters extracted from different concavity pairings on the same profile are very stable, varying by less than 5^o over the three concavities.

As further illustrations of grouping for SORs, Figure 5 shows an image with two SORs, and Figure 6 an image with many SORs of various types and sizes. The matched concavities are partitioned into sets, and the profile curves corresponding to each set grouped. The entire process is automatic and relies only on the affine properties of the homology between symmetrical portions of the outline.

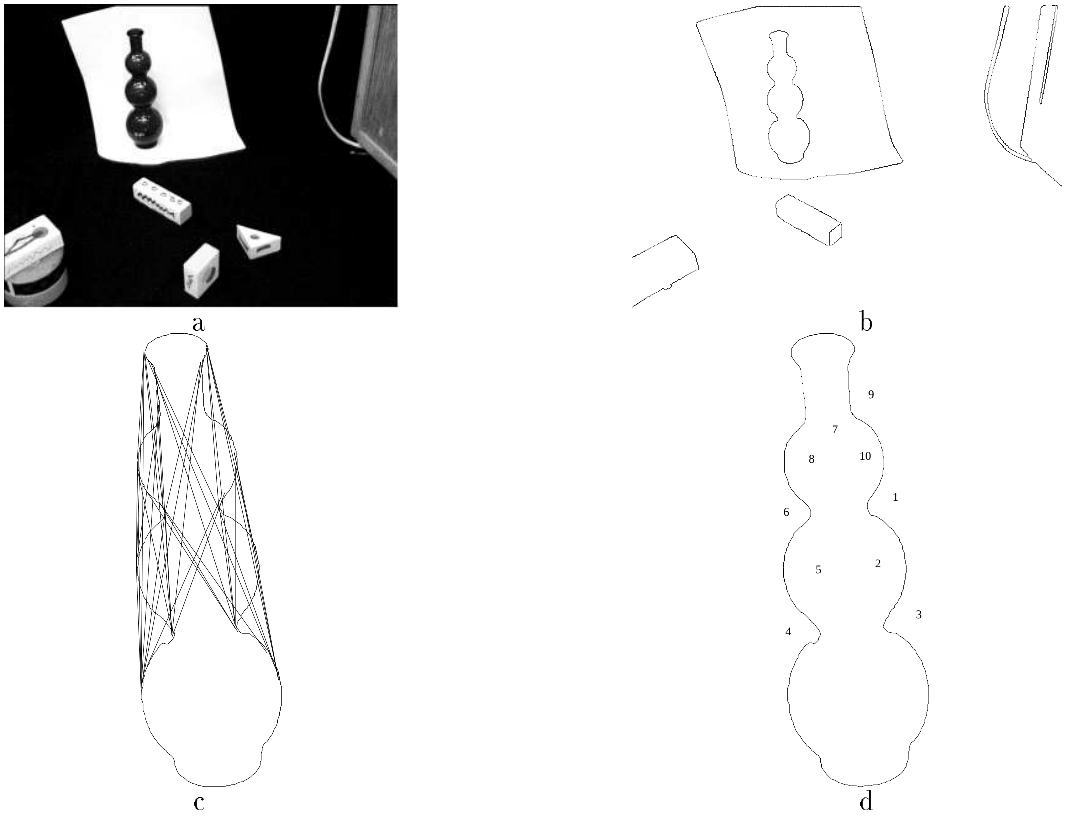

Figure 4: Example of class based grouping for surfaces of revolution. (a) Original image. (b) Edges from (a). (c) Detail showing only bitangents on SOR curves, other bitangents are found on (b). (d) 11 simple concavities are identified. The numbering refers to table1, where the matched concavities and their parameters are given.

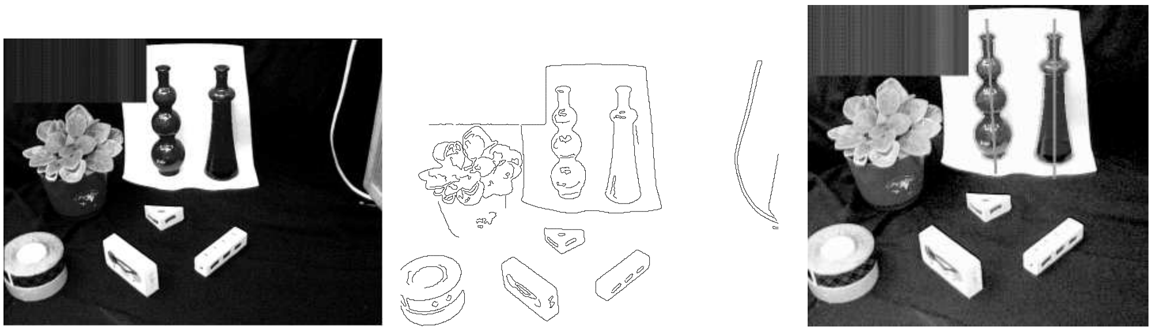

Figure 5: Grouping (as described in the text) is used to automatically extract two profiles of the surfaces of revolution. (a) Original images. (b) Canny edges. (c) Extracted grouped profiles and skewed symmetry axes. 24 concavities are extracted. Matching on area gives 48 possible pairings. Of these only two groups also satisfy the affine involution, one group with two members, the other with three. The axes and outline curves are indicated by a light grey overlay.

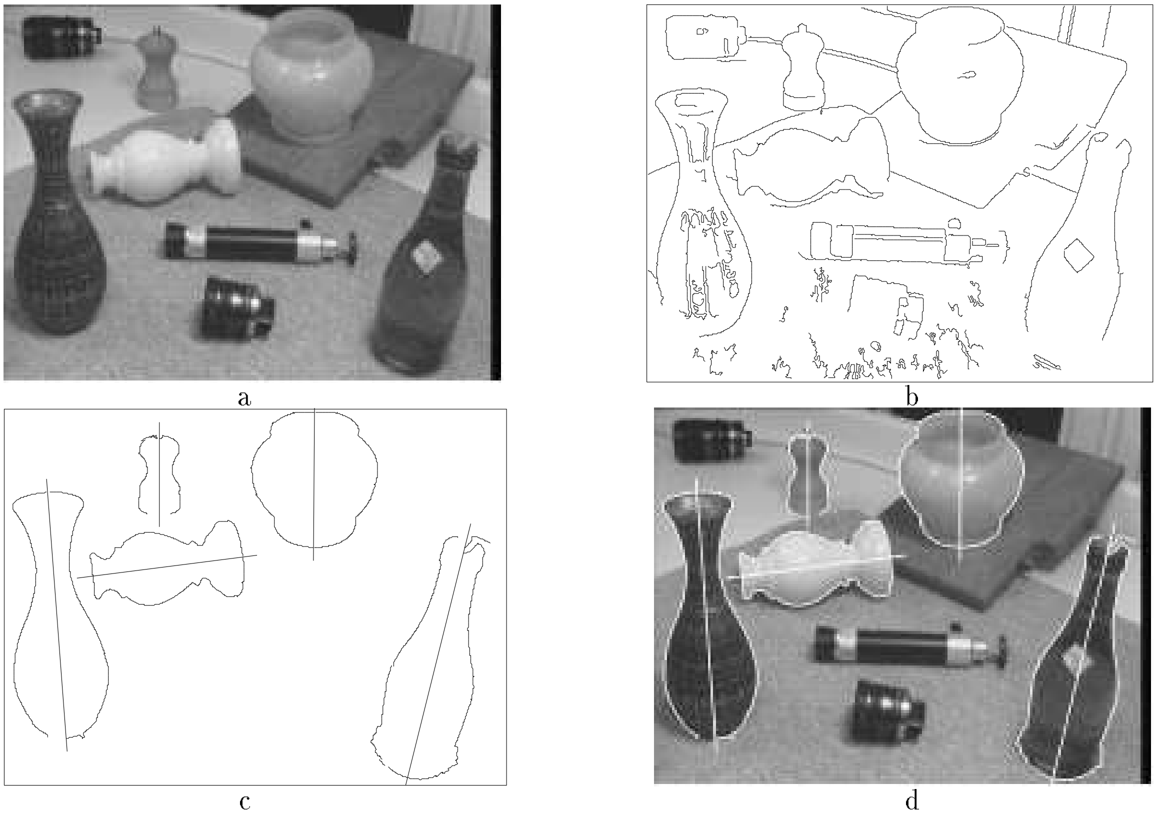

Figure 6: (a) Original image containing several SORs. (b) The linked edges computed from (a). (c) Extracted SOR profiles with axes using the SOR grouper (d) Extracted SORs superimposed on original image.

2.3 Invariants

We briefly review invariant computation for this class. Corresponding profile bitangents intersect at an image point $\mathbf{p}$, which is the image of $\mathbf{P}$, the point of intersection between the object axis and planes bitangent to the surface. $\mathbf{P}$ is viewpoint-independent [5], and $\mathbf{p}$ can be identified in any image. Four such points define a cross-ratio, which is a projective invariant.

Indexing for surfaces of revolution proceeds by hashing based on invariants [11], and then verifying first by alignment on the axis distinguished points (generated by bitangent intersections), then on the profile. The number of distinguished points is restricted to prevent a combinatorial explosion in the number of invariants. The restriction is achieved by only considering bitangents between convex curve portions (convex defined to be facing out from the symmetry axis). This condition is preserved under perspective projections. Restricting to only convex-convex bitangents has an additional advantage - due to self-occlusion concave-concave and concave-convex bitangents may be missing, since concave profile portions are the first to be self-occluded. Consequently distinguished points formed from these bitangents may well be missing. Typically a model has 5-10 distinguished points from convex-convex bitangents out of a total of 25-50 distinguished points of all types. This bitangent classification scheme has been successfully tested on object model outline curves, which are well formed and do not have missing distinguished points. Work is still in progress on applying the new indexing approach to complex scenes, where only a subset of the distinguished points in a full model can be recovered.

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

3.1.1 Weak perspective projection

We make strong use of the fact that the 'envelope of the profile is the profile of the envelope' [20]. 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 [3, 7], the locus of centres of circles bitangent to the profile. Since a single scaling applies for weak perspective imaging, the circles have have equal radius. There are a number of consequences of these results:

1. 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.
2. The curve traced out by the circle centres, σ, is the projection of the sphere centres, i.e. of the canal surface axis, α.
3. The sides of the profile are parallel curves of σ, i.e. curves generated by tracing a fixed distance along the normal to σ. Thus tangents at corresponding points on the profile are parallel.
4. Since the axis is planar, points on $\sigma(s)$ are within a plane affine transformation of points on the canal surface axis, α. This is the only property that requires that the axis be planar, rather than a space curve.

3.1.2 Perspective projection

Under perspective projection a sphere profile is an ellipse and the sphere centre does not project to the ellipse centre. In practice though these differences from weak perspective projection are often not significant - we return to this point in section 3.3. However, the size of the profile ellipse varies in a perspective imaging (according to the distance of the sphere from the optical centre). It follows that the profile — the ellipse envelope — is not a parallel curve. The following properties of the profile can be derived by explicitly solving for the contour generator and projecting, and also by explicitly solving for the circle envelope. Here we give geometric proofs:

1. Inflections in the axis: (a) produce inflection pairs on the profile - one on each side; (b) tangents at corresponding profile inflections (on each side of the profile) intersect on the vanishing line of the axis curve's plane. See figure 7a. Proof: (a) consider an inflection in the axis curve α. In the plane of α, a parallel curve at a distance r to α has curvature $\kappa = \kappa_\alpha/(1 - r\kappa_\alpha)$ where κ_α is the curvature of α. So if $\kappa_\alpha = 0$, there is a corresponding zero of curvature in the parallel curve. In 3D an inflection in α generates a ring of parabolic points on the canal surface, which is the parallel surface to α. Since a parabolic surface curve projects to a profile inflection [9], each

side of the profile contains an inflection. (b) At a ring of inflections, locally the canal surface is cylindrical to second order (i.e. the second fundamental form is identical to that of a cylinder). Consequently, on the parabolic circle the tangents to the contour generator are parallel to the tangent to α, and thus parallel to the plane, Π_α of α. In 3D these three parallel lines (the axis tangent vector at the inflection and corresponding contour generator tangents) intersect on the plane at infinity, Π_∞. The point of intersection lies on a line on Π_∞ which is the axis of the pencil of planes parallel to Π_α. In the image these three projected vectors (the inflection tangent to σ and corresponding profile tangents) intersect on the vanishing line $\mathbf{l}_\infty$ of Π_α. $\square$

Two such intersections determine the $\mathbf{l}_\infty$ in the image. Note: this constraint also applies to line segments on the profile (a line is a 'degenerate' inflection), so can be applied to a piecewise linear axis.

2. Corresponding profile tangents (i.e. points whose pre-image is on the same circular cross-section) intersect on $\mathbf{l}_\infty$. See figure 7b. **Sketch proof**: The pre-image of the profile tangents are *not* in general parallel to the corresponding axis tangent, and therefore not parallel to Π_α. Consider consecutive spheres used to generate the canal surface. The contour generator of each sphere is a circle whose plane is perpendicular to the ray from the optical centre to the sphere centre. However, the pair of planes (externally) tangent to both spheres and passing through the optical centre, intersect in a line through the optical centre which is parallel to the vector between the sphere centres. It follows that in the infinitesimal limit of this vector, the tangent planes defined by the profile tangents and optical centre, are parallel to the corresponding axis tangent vector. Consider intersecting these tangent planes with an image plane parallel to Π_α — the lines of intersection are parallel, and are parallel to the tangent vector to σ. Consequently, in any other image the three vectors (profile and imaged axis tangents) intersect at the vanishing point of the tangent line to σ — which lies on $\mathbf{l}_\infty$. $\square$

Both the above results also hold under weak perspective imaging — though in this case the vanishing line $\mathbf{l}_\infty$ is at infinity in the image.

This relationship is not as powerful as the projective transformation between profile sides of a SOR. For a canal surface, the relationship constrains only profile tangents (a two degree of freedom constraint on the tangent dual space), whereas for a SOR there is a four degree of freedom point to point transformation.

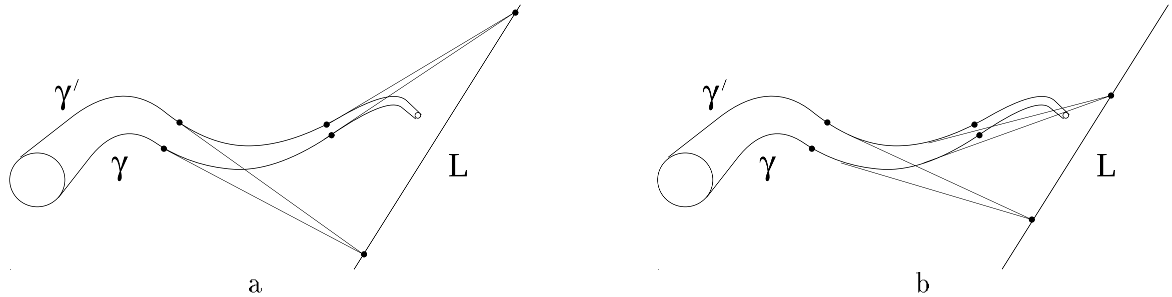

Figure 7: For a canal surface with a planar axis, (a) inflections in the profile occur in pairs for each inflection of the axis. The intersection of a pair of inflection tangents determines the vanishing point of the tangent line at the axis inflection. Two such vanishing points determine the vanishing line, $\mathbf{l}_\infty$, of the plane of the axis; (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.

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 utilised because an inflection is a distinguished point which can be localised 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 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.

3.2.1 Inflection based grouping

The target profiles are assumed to contain linear segments - for example near an inflection. Putative groupings are assembled in four stages which are illustrated in figure 8:

1. Straight lines segments are fitted to chained edgels to a very tight tolerance using orthogonal regression (figure 8c).
2. Line segments are only associated in pairs if they could have arisen from corresponding profiles (e.g. the difference in their orientation is limited).
3. Line segment pairs are associated if one (or more) segment of each pair is connected to one of another by an edgel chain.
4. Two pairs determine a vanishing line. Additional pairs are only associated if the vanishing point of the new pair is sufficiently close to the existing vanishing line, and again one or more segments is connected by a common edgel chain (figure 8d). Only associations of 3 or more pairs are considered further.

3.2.2 Verifying grouping

Consider corresponding profile points $\mathbf{p}$ and $\mathbf{p}'$. The profile tangent at each point $\mathbf{p}$ on one side of the profile, intersects $\mathbf{l}_\infty$ in a point $\mathbf{v}_p$. There is a one parameter family of lines through $\mathbf{v}_p$, and a member of this family must be parallel to the profile at the corresponding point $\mathbf{p}'$ on the other side from $\mathbf{p}$.

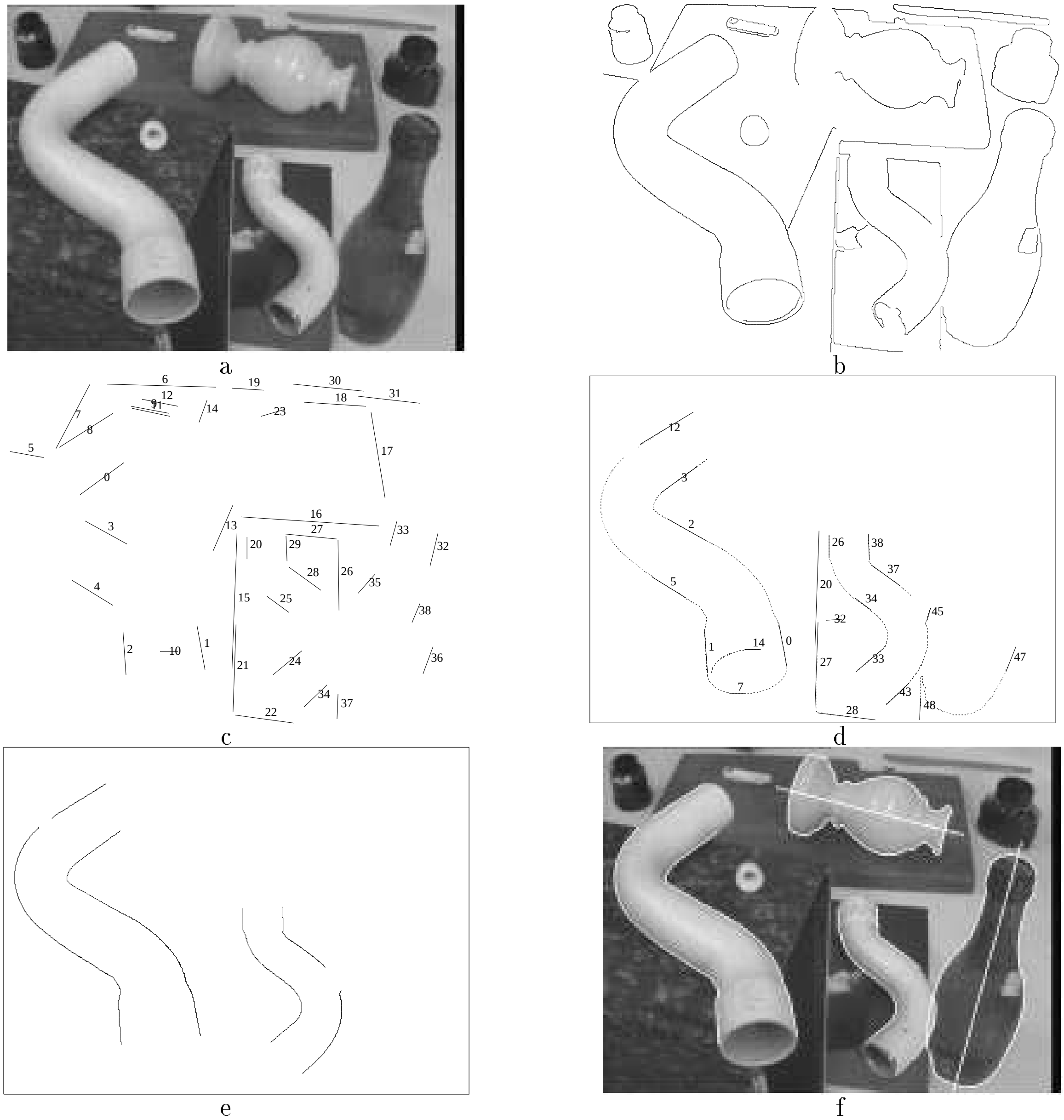

Figure 8: (a) Original image containing two canal surfaces, the leftmost showing significant perspective distortion. (b) Linked edges from a Canny operator, with extrapolation over single pixel gaps. (c) Extracted line segments using a tight threshold. (d) Resulting grouped triplets of paired line segments based on a common vanishing line. 7 grouped triplets are found. Of these, only two pairing are linked by curves whose tangents also consistently intersect on this vanishing line. (e) The two canal surface profiles that are finally extracted. (f) Extracted profiles superimposed on the original image.

	figure 8	figure 12
No. curves	27	87
No. line segments	57	87
No. unpaired segments	18	5
No. line pairs	35	128
No. grouped triplets	7	176
No. verified hypotheses	2	2

Table 2: Statistics for canal surface profile grouping.

Naively, this could lead to a one dimensional search algorithm for corresponding points. However, because profile curves are continuous (other than at self-occlusions, see below), neighbouring points of $\mathbf{p}$ correspond to neighbouring points of $\mathbf{p}'$. So a very restricted local search is involved.

This discussion delivers three criteria that can be used to verify that curve portions belong to a canal surface: (1) For each point $\mathbf{p}$ there is a corresponding point $\mathbf{p}'$ which is determined by profile tangent's intersection; (2) The corresponding point to $\mathbf{p}'$ must be $\mathbf{p}$; (3) Neighbours of $\mathbf{p}$ must match neighbours of $\mathbf{p}'$ with consistent ordering on the respective curves.

Provided a 'sufficient' extent of the curves between the inflections satisfy these criteria, then the hypothesised association of inflections is verified. The criteria will not be satisfied at all points on the curves between inflections because of profile drop out due to: first, missing features due to edge detector limitations; second, occlusion due to other objects; and third, self-occlusion (for opaque surfaces) which occurs when the contour generator is parallel to the ray through the optical centre [9].

The success of this hypothesise and verify strategy is illustrated in figure 8e, where only the canal surface profiles, and no other associations, are extracted. Statistics are given in table 2.

Implementation Accurate tangents are computed at points on edgel-chains by locally fitting a quadratic using least squares regression with central Gaussian weighting on a neighbourhood of up to 33 points. Given edgel-chains that are hypothesized to be part of the profile of a canal surface, these tangents are then intersected with the vanishing line. The distance between intersection points arising from tangents on either side of the profile must be below a threshold. Provided 65% of the edgel-chain is matched in this way, then the verification is accepted.

3.3 Invariants

An affine invariant signature can be constructed by tracing the profile tangent's intersection on $\mathbf{l}_\infty$ as a source point moves along the profile. However, there are two problems with this representation: first, it only measures directions of the tangents, their relative position is lost; second, and more important, although the signature is an affine invariant the parameterisation of the signature is unknown. Consequently, only qualitative features of the signature (such as turning points) can be utilised, and these are largely dominated by inflections of the original axis.

Instead of exact projective invariants, an approximation is used which allows the extraction of affine invariants of the axis α from perspective images. As is demonstrated by the experimental results the approximation is excellent.

The approximation is to continue to use the symmetry set to determine the imaged axis, even under perspective imaging conditions. As described in section 3.1.2 there are three perspective effects that must be accounted for: first, the 'radius' of the sphere's profile is no-longer constant, and in fact can vary considerably; second, the sphere's profile is elliptical, rather than circular; and, third, the centre of the sphere does not project to the centre of the ellipse. However, the latter two effects are extremely small for images with a reasonable field of view (for example the centre is only displaced by about 1% of the circle diameter at the border of a 46^o field of view lens). The reason that the approximation works so well is because the symmetry set is unaffected by the first, and only significant, effect of the three: Ignoring all but the first effect, the profile is then the envelope of circles of varying radius swept along the projected axis σ, and the symmetry set correctly recovers the curve σ.

To this approximation points on σ are within a plane *projective* transformation of the planar axis α. However, the vanishing line $\mathbf{l}_\infty$ of the plane of α is known in the image, and thus α is known up to a planar *affine* transformation.

Implementation The computation of the symmetry set uses the method described in [15]. The symmetry set is affinely normalised using an affine canonical frame [10]. This requires three basis points to be determined from the symmetry set, and these are provided by three inflection tangents. In the pipes used here, there are at least three straight segments of the symmetry set ('extended inflections'). The intersection of these three lines generates the three basis points, which are mapped to vertices of an isosceles triangle in the canonical frame. Ordering is provided simply by the symmetry

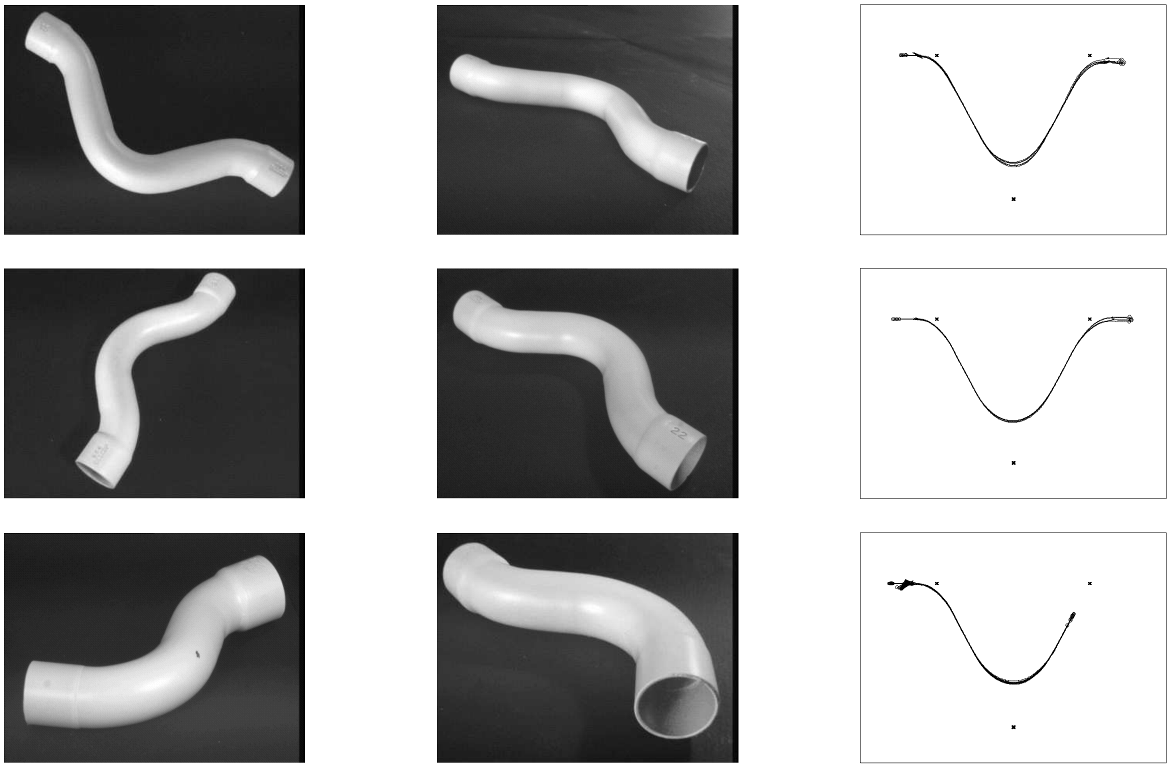

Figure 9: Stability of the canonical frame curves to viewpoint. Example images and canonical frames curves for three canal surfaces (pipes 2, 3 and 4). Note the wide variation in viewing position. From top to bottom, the canonical frames contain symmetry sets generated from six, five and twelve images respectively. At least half of each set show significant perspective distortions—note the variation in pipe width in the middle column due to perspective. The slight instability present towards the ends of the canonical frame curves are due to errors in the extracted symmetry set which occur where the pipe radius changes.

set.

Stability to viewpoint is excellent, as is demonstrated in figure 9. Invariants are computed from measurements on the canonical frame curve in a similar manner to the 'footprints' of Lamdan *et al.* [10], although here lengths are measured rather than areas.

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. Both have trihedral vertices. They are: first, polyhedra whose edges are in one of only three directions (like a cube); and, second, triangular prisms. Examples of both sub-classes are shown in figure 4.

4.1 Grouping Mechanism

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. For example, only the hexagonal outline of a cube may be recovered as shown in figure 10a. This outline then generates a cube-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. For the first class, for every vertex on the silhouette there is only one possible choice of edge direction for the internal boundary; a subset of internal boundaries can then be inferred from the preferred direction property. For boxes, this leads to a complete wireframe. In the case of prisms, a generic silhouette has a pair of line segments that are separated by a single segment, and are almost parallel (assuming that the distance to the prism substantially exceeds its size, a safe assumption). These two segments identify the vanishing point for the three vertical edges of the prism; once they have been identified, one quadrilateral face and one triangular face are known. Assuming an affine view yields the other faces by identifying the edges of the other two quadrilateral faces.

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 for a cuve[2]. As figure 10b shows, the strategy is effective in reasonably complex images.

4.2 Invariants

As shown in [18], the projective invariants of the system of planes bounding many polyhedra can be measured from wireframe outlines in images, by manipulating the (linear) vertex-plane incidence

[2] 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.

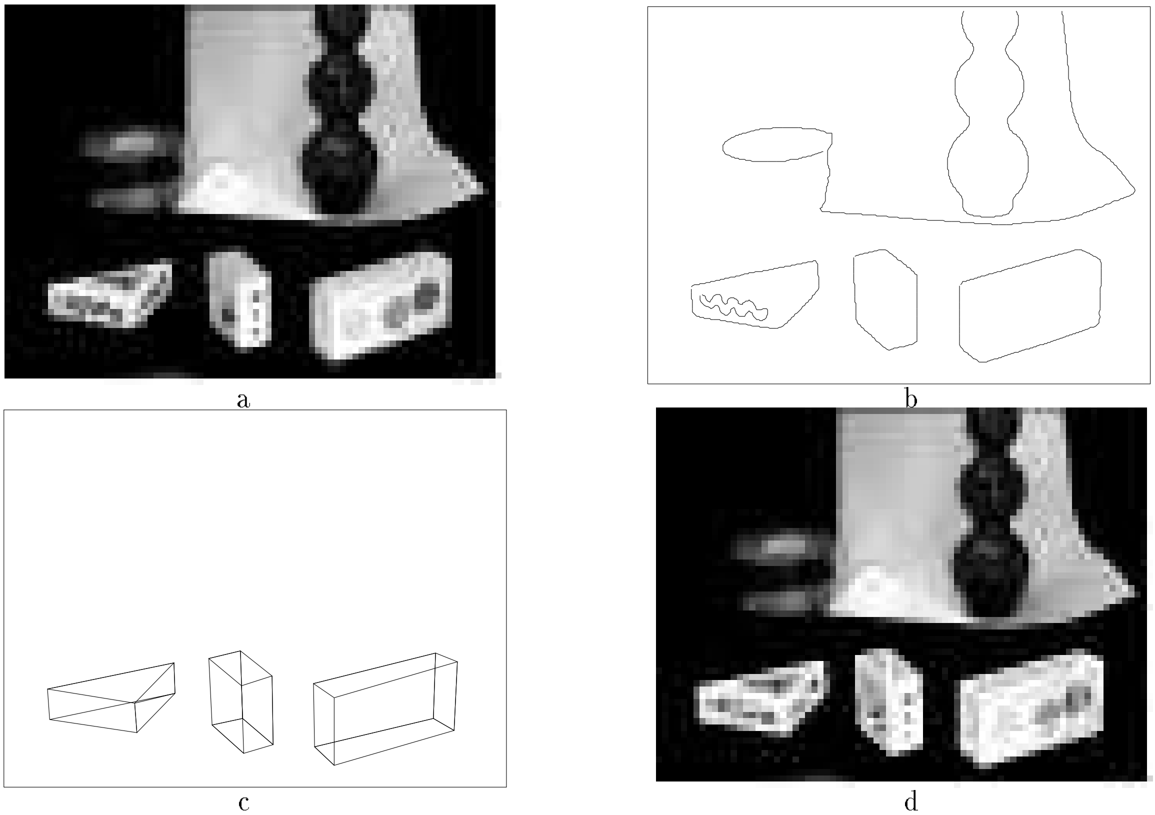

Figure 10: (a) Original image containing several polyhedra. (b) The linked edges computed from (a). These are used as input to the polyhedral snake. (c) Extracted polyhedral wireframes using both sub-class groupers. (d) Outlines superimposed on original image. Note, interior boundaries are correctly recovered.

constraints. These constraints yield a system of homogeneous linear equations, whose solutions are a set of planes projectively equivalent to the original set. At present, we use highly degenerate polyhedra (boxes and prisms), so that the class information obtained in grouping identifies the projective geometry of the polyhedron, and the construction is unnecessary.

5 Other GCs

For SHGCs we can generalise the procedure of Zerroug and Nevatia [23] to perspective: suppose we have two cross-sections (or fragments of cross-sections), then we can construct a planar homology from these, which determines the projection of all other cross-sections of the surface. The planar homology, corresponds to the plane projective transformation between the imaged cross-sections. The homology has a line of fixed points and a fixed point. The line of fixed points is the vanishing line of planes perpendicular to the SHGC axis (i.e. planes parallel to the cross-section). Note that with this fixed line the geometry of the cross-section can be recovered up to an affine ambiguity from perspective images. Recall that an affine transformation is defined by a fixed line. The fixed point lies on the imaged axis, and is also given by the profile bitangent intersections at the two cross-sections.

To determine the imaged cross-section at another point $\mathbf{p}$ of the profile compute the planar homology which maps one of the reference cross-sections onto this point. To construct the homology: Method 1, Determine the five dof of the homology - Construct the bitangent between the profile point $\mathbf{p}$ and the profile point $\mathbf{q}$ at the reference cross section, this intersects the imaged axis at the fixed point $\mathbf{a}$ (2dof) of the homology, and also determines the characteristic invariant (1dof) of the homology as the cross ratio of four points on the bitangent - the fixed point $\mathbf{a}$, the two bitangent contact points, and the intersection of the bitangent with the vanishing line. The vanishing line is the fixed line (2dof) of the homology. Method 2, compute the projective transformation $\mathtt{T}$ (8dof) with the vanishing line as a line of fixed points, $\mathbf{a}$ as a fixed point, and which transforms $\mathbf{q}$ to $\mathbf{p}$.

6 Examples

Figure 11 shows an example scene in which all the SORs and canal surfaces have been grouped. Figure 12 includes SORs, canal surfaces and polyhedra. In both cases all profiles and outlines are extracted and grouped automatically, and there are no false positives.

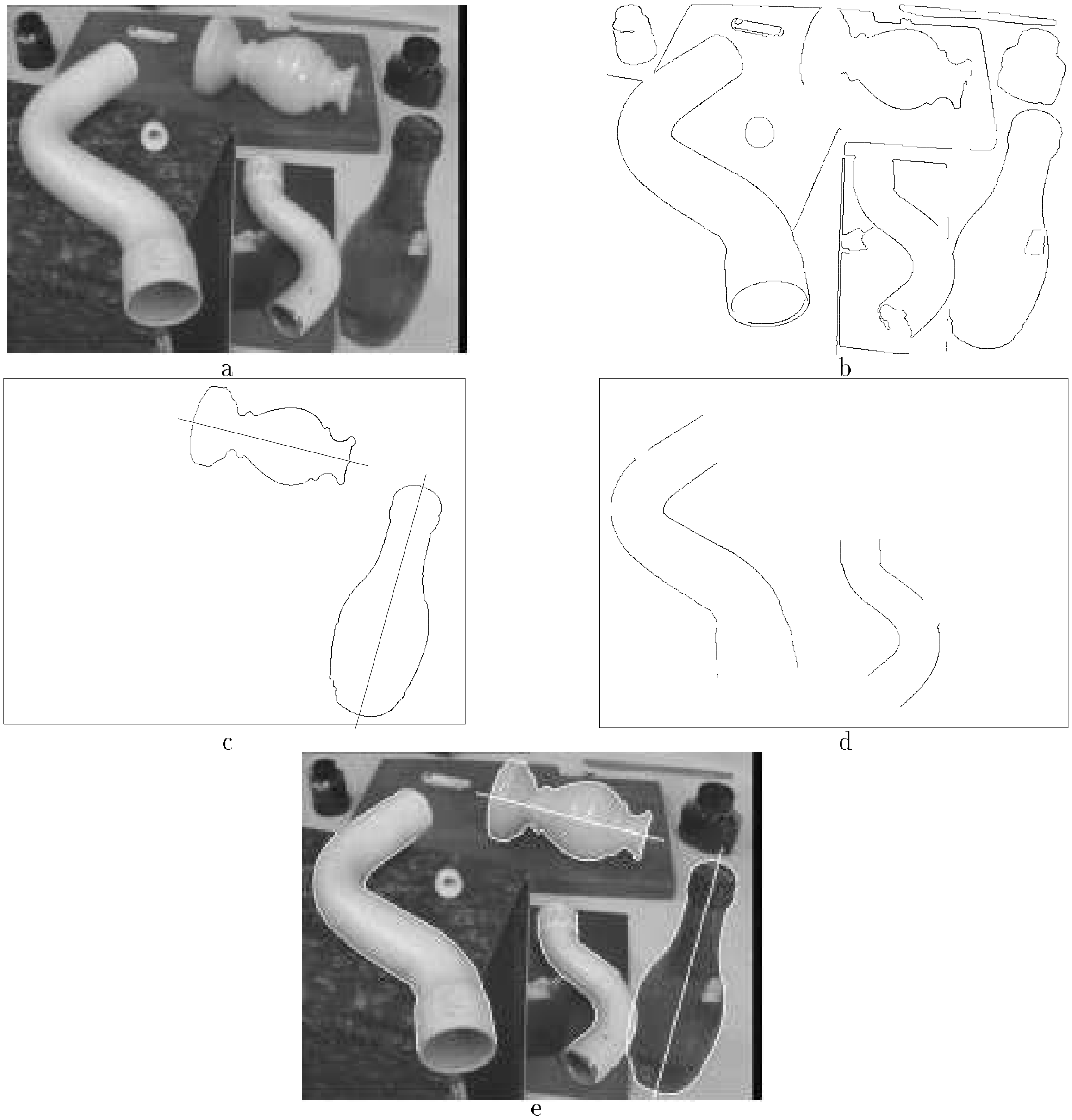

Figure 11: (a) Original image containing two SORs and two canal surfaces. (b) The linked edges computed from (a), note the gaps and incorrect topology. Profiles are extracted and grouped from these linked edges by the class based groupers. (c) Extracted SOR profiles with axes. (d) Extracted canal profiles. (e) Extracted profiles superimposed on original image. Note, all the correct instances of a class have been grouped, and no false instances grouped. Note that gaps — from edge detector drop out or occlusion — in one side of the profile, shown in (b), have not prevented grouping, and have been 'repaired'.

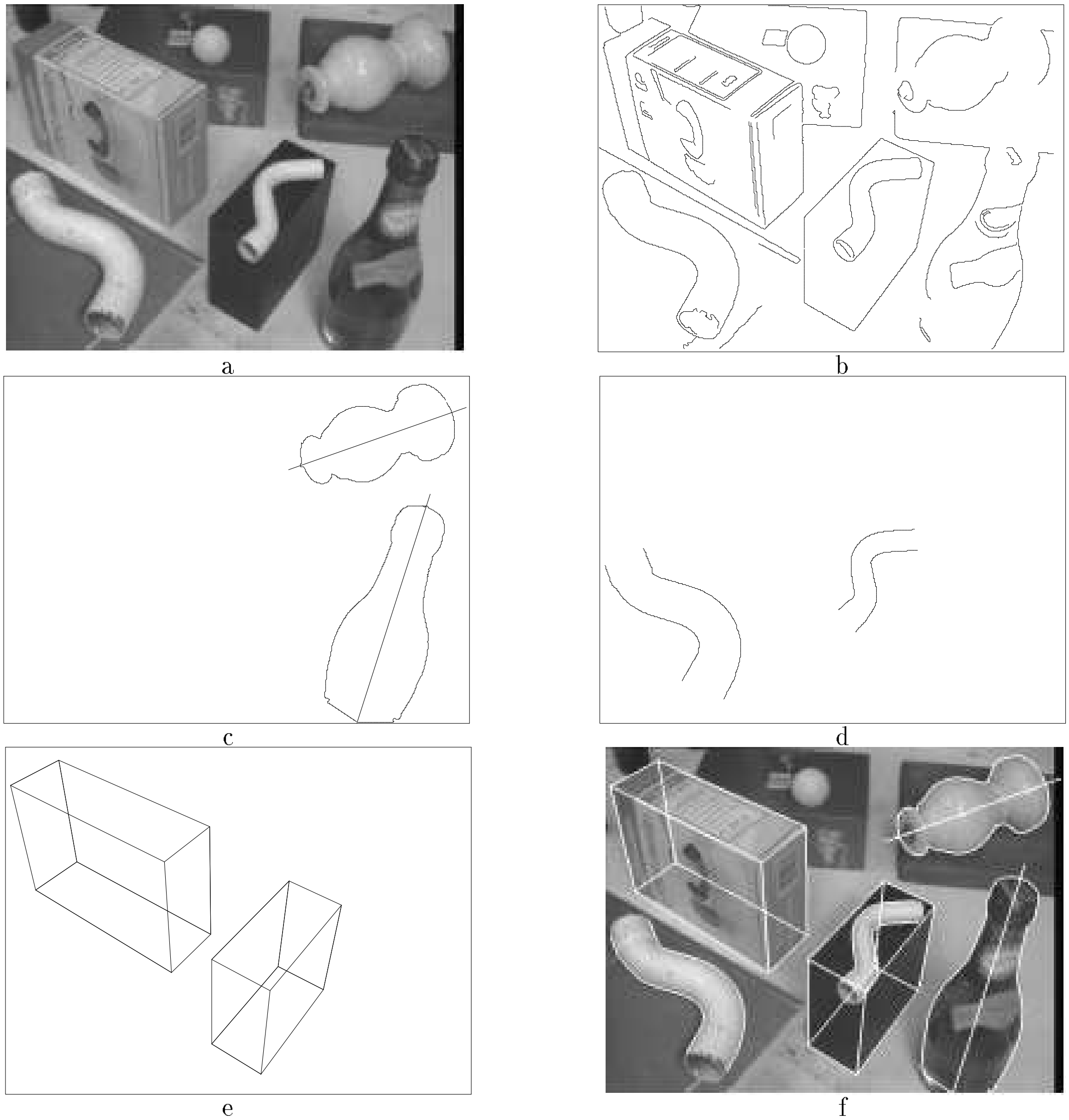

Figure 12: (a) Original image containing two SORs, two canal surfaces, and two polyhedra. (b) The linked edges computed from (a). Profiles are extracted and grouped from these linked edges by the class based groupers. (c) Extracted SOR profiles with axes. (d) Extracted canal profiles. (e) Extracted polhedra outlines. (f) Extracted profiles superimposed on original image. Again, all the correct instances of a class have been grouped, and no false instances grouped.

7 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. We have demonstrated "recognising" the *class* - by employing class based groupers to determine and group curves belonging to putative profiles for that class. An object is first grouped and *recognised* as an SOR, before it is classified as a particular vase or bottle. 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.

The grouping methods developed here also have appropriate tolerances i.e. they are not overly sensitive, for example failing catastrophicly due to object imperfections (objects in the real scenes are not perfectly SOR and certainly not canal surfaces), and imaging imperfections — there is significant radial distortion towards the periphery of the image.

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. 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 [17] for skewed ribbons, or the semi-local invariant [22] could be used to initialise matching.

References

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

[2] Binford, T.O., "Survey of model-based image analysis systems," *IJRR*, **1**, 1, 18-63, 1982.

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

[4] 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.

[5] 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.

[6] Gross A.D. and Boult T.E., "Recovery of Generalized Cylinders from a Single Intensity View", *Proc. of Image Understanding Workshop*, 1990.

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

[8] Kanade T., 'Recovery of Three Dimensional Shape of an Object from a Single View'. *Artificial Intelligence*, vol. 17, nos. 1–3, pp. 409–460, 1981.

[9] Koenderink, J.J., *Solid Shape*, MIT Press, 1990.

[10] Lamdan Y., Schwartz J.T. and Wolfson H.J., 'Object Recognition by Affine Invariant Matching'. *Proc. CVPR*, pp. 335–344, 1988.

[11] 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.

[12] Mukherjee D.P., Zisserman A. and Brady J.M., 'Shape from symmetry—detecting and exploiting symmetry in affine images'. To appear, *Proc. Royal Soc.*, 1994.

[13] Mundy, J.L. and Zisserman, A. (editors), "Geometric Invariance in Computer Vision,", MIT Press, Cambridge Ma, 1992.

[14] Nalwa V., 'Line drawing interpretation: Bilateral symmetry' *IEEE PAMI*, vol 11, 10, 1117-1120, 1989.

[15] Pillow N., Utcke S. and Zisserman A. 'Viewpoint-Invariant Representation of Generalized Cylinders Using the Symmetry Set', *Proc. British Machine Vision Conference*, 1994.

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

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

[18] 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.

[19] Rothwell, C.A. *Recognition Using Projective Invariance*, D.Phil. Thesis, Department of Engineering Science, University of Oxford, 1993. To appear OUP, 1994.'

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

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

[22] 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.

[23] 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.