

Part-Based 3D Descriptions of Complex Objects from a Single Image

M. Zerroug, *Member, IEEE Computer Society*, and R. Nevatia, *Fellow, IEEE*

Abstract—Volumetric, 3D, part-based descriptions of complex objects in a scene can be highly beneficial for many tasks such as generic object recognition, navigation, and manipulation. However, it has been difficult to derive such descriptions from image data. There has been some progress in getting such descriptions from range data or from perfect contours, but analysis of a real intensity image presents many difficulties. The object and part boundaries do not completely correspond to image boundaries. The detected boundaries are often fragmented and many boundaries due to surface markings, shadows, and noise are present. In addition, inference of 3D from a 2D image is difficult. This paper describes a method to compute the desired descriptions from a single image by exploiting projective properties of a class of generalized cylinders and of possible joints between them. Experimental results on some examples are given.

Index Terms—3D shape descriptions, part-based representations, object segmentation.

1 INTRODUCTION

ABILITY to automatically detect and describe complex curved objects is important for many visual tasks including generic object recognition, navigation, robotic grasping, and object learning. An important design issue in solving this problem is the use of an adequate shape representation scheme. It is well-accepted that a good way to represent a complex 3D object is by decomposing it into volumetric parts and describing the parts and the relationships between them. This type of representation is rich, stable to small irregularities in shape, and allows the representation of objects even under occlusion or missing components [3], [14]. This is highly useful for object comparison, even with large differences in surface properties between objects, and under different viewpoints. There is also evidence, such as from psychological experiments performed by Biederman [2], that a similar scheme is used by the human visual system.

The use of simpler parts to describe more complex objects has a long history in computer vision [3], [11], [14], [16], [18]. However, the use of part-based representations in computer vision systems has been limited. This is, we believe, largely due to the difficulty of automatically computing part-based descriptions from a real image: The part decomposition hierarchy is not given a priori, we must infer it from the observable features in the data. In addition, we need to first separate an object from its background (or other objects in the scene) and infer 3D shape. Consider the example shown in Fig. 1. Humans readily see a teapot with some other surfaces

that are considered to be part of the “background.” Humans naturally describe this teapot as consisting of a container, a handle, a lid, and a spout and infer 3D shapes of these parts. Note that this does not require prior knowledge of the objects expected in the image.

It should be clear that if such descriptions can be obtained automatically, the tasks of object recognition and manipulation would become much easier. However, while it is easy for humans, this process presents many computational difficulties. Consider computing the desired descriptions from the edges detected from the image and shown in Fig. 2. Lack of direct 3D measurements makes it difficult to determine discontinuities that may characterize object and part boundaries. Instead, we must work with intensity boundaries which may correspond to depth boundaries, but also to markings, shadows, specularities, and noise. Furthermore, object boundaries are not complete due to both poor edge localization and occlusion. We need to aggregate the appropriate boundaries to segment the teapot from the background, to find the significant parts of the teapot and to infer their 3D shape. In this paper, we present a method to generate such descriptions from a single intensity image.

Some of the computational problems become easier if dense range data is available. Ideally, the only boundaries then correspond to surface discontinuities and surface markings become invisible. The availability of 3D data also helps with the task of part decomposition and shape inference. Early work on extracting part-based descriptions from a range image of complex curved objects is described in [14]. In [16], Pentland presents a system which derives part-based descriptions using super-quadratics. This method assumes that the image has been presegmented into regions corresponding to parts (skeletonization) and focuses on the fitting problem, which is addressed using a coarse-to-fine method. Thus, in a sense this method addresses the description problem, treating the segmentation as a separate and

- M. Zerroug is with Adept Technology Inc., 17800 Castleton St., Suite 175, City of Industry, CA 91748.
- R. Nevatia is with the Institute for Robotics and Intelligent Systems, Department of Computer Science, University of Southern California, Los Angeles, CA 90089-0273. E-mail: nevatia@usc.edu.

Manuscript received 6 Aug. 1997; revised 8 Feb. 1999.

Recommended for acceptance by K. Ikeuchi.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107571.



Fig. 1. Image of a compound object.

prior issue (such as a manual segmentation, for example). More recently, there have been efforts to describe a range image in terms of geons, (qualitative generalized cylinders) such as in [7] and [15]. An effort to recover qualitative and quantitative part descriptions from range data is given in [6]. While range data is very helpful in solving the problems of object segmentation and description, it is rather difficult to acquire: Most sensors work well with only certain kinds of surfaces, ambient illumination may need to be controlled and the sensors are expensive.

Single intensity images, on the other hand, can be acquired easily and inexpensively in most environments and, hence, attractive to use. Some previous work has attempted to address generic part and object segmentation from a 2D image. Many of these methods use synthetic data or assume that perfect object outlines are given [1], [5], [9]. Under these conditions, it is possible to identify parts and relations by using low-level features such as junctions, but it is not clear how these approaches can be generalized to work with fragmentation and clutter common in real images.

There have also been systems that attempt to work with real images. Early work include Brook's Acronym system [4] where a model-based system is described for segmenta-

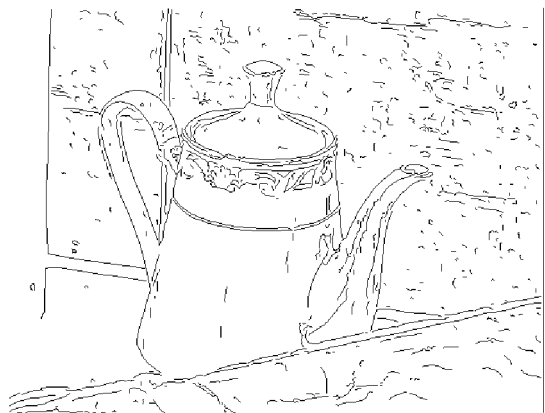


Fig. 2. Edges detected from the image shown in Fig. 1.

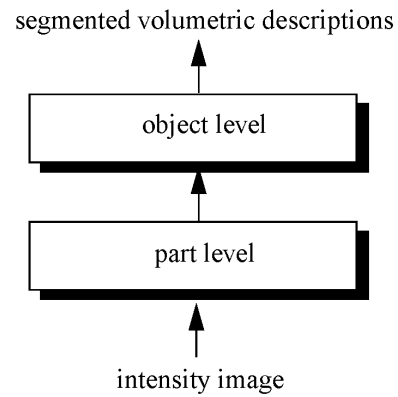


Fig. 3. The two major levels of the method.

tion/recognition of airplane images. Rao and Nevatia [17] have described a system that uses more generic models and uses ribbons to describe parts. Mohan and Nevatia [12] used a perceptual grouping approach to form descriptions through a hierarchy of features. These methods use intuitive constraints to build descriptions which may not necessarily correspond to properties of the viewed 3D objects and no explicit 3D recovery is attempted.

More work has have taken the approach of using rigorous, but generic models and attempting to find them in images from expected projective properties. These models have been typically subclasses of generalized cylinders (GCs). Commonly used classes are straight homogeneous generalized cylinders (SHGCs), where axis is straight but cross section size may vary, and planar right generalized cylinders (PRGCs) which allow axis to be curved but planar. Sato and Binford [19], and Gross and Boulton [8] describe a system for recovery of SHGCs. In our own previous work, we have described methods for detecting SHGCs [23] as well as PRGCs [24] from real images; these systems recover 3D descriptions of the detected objects as well. These methods, however, are designed to work only with objects consisting of a single part (i.e., simple objects).

In this paper, we describe a method that detects and describes complex objects (i.e., those containing multiple parts). The method we present builds on our recent work [23] and [24] to recover descriptions of single-part objects in terms of subclasses of generalized cylinders (GCs). It addresses the figure/ground problem and part-based shape description problem at the same time and in a generic way (i.e., without prior knowledge of the specific objects being viewed). Compound objects introduce many new issues related to the use of joint relationships to effect object-level segmentation and shape description, including 3D shape inference. We address these issues through the exploitation of geometric (scaled orthographic) invariant and quasi-invariant properties as well as structural properties of the contours of GCs mentioned above. One of the significant result of using such properties is that the derived descriptions give projective attributes of 3D shape, not just intuitive ones. The paper is organized as follows. Section 2 provides an overview of the method. Section 3 contains details of the method

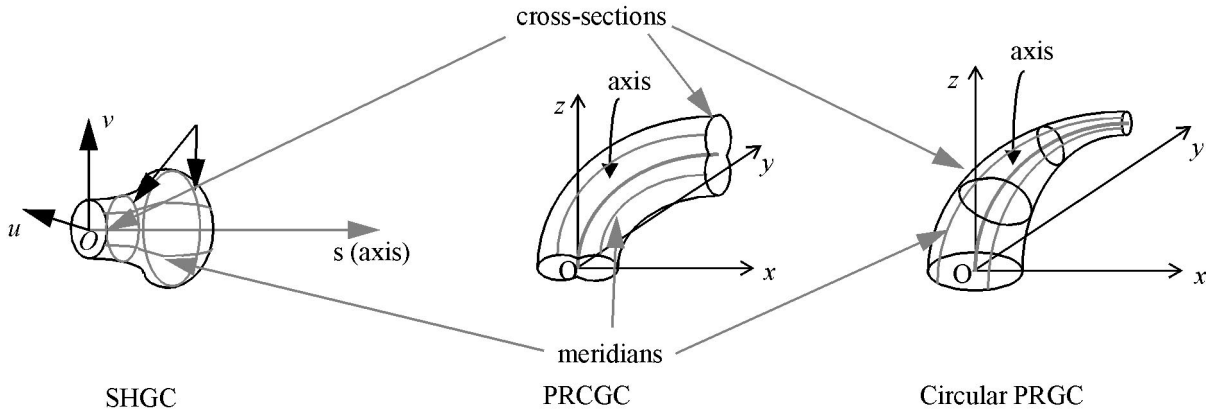


Fig. 4. Generalized cylinders used to describe parts in the method.

for detecting complex objects and describing them by projections of parts and relations between them. Section 4 gives a method for 3D shape reconstruction from the 2D descriptions. Section 5 has a discussion of the strength, limitations, and complexity of the method. A conclusion is given in Section 6.

2 OVERVIEW OF THE METHOD

Our system takes a single intensity image as input. Edges and curves are detected from this image and used for further processing. The segmentation and description method may be viewed as consisting of two main levels, the part level and the object level as shown in Fig. 3. The part level generates part hypotheses using relations between curve segments which should satisfy some projective properties of the modelled part classes. This may result in multiple hypotheses for the same part. The object level analyzes the joint relationships between the hypothesized parts to build a geometric context used to hypothesize compound objects and resolve ambiguities. 3D descriptions are then generated for the parts of the object; the relations between the parts can help in the process of 3D inference.

The scope of this method are objects which consist of the arrangement of parts which can be (approximately) described by one of the following classes (see Fig. 4):

1. **Straight Homogeneous Generalized Cylinders (SHGCs):** These are generalized cylinders (GCs) with a straight axis. The cross sections are homogeneous (i.e., of the same shape), but may vary in size along the axis.
2. **Planar Right Constant Generalized Cylinders (PRCGCs):** Here, the axis is allowed to be any planar curve, however, the cross sections are fixed, both in shape and size. The cross sections must be orthogonal to the local axis direction.
3. **Circular Planar Right Generalized Cylinders (circular PRGCs):** These also allow any planar curve for axis and the cross section size may vary along the axis. However, the cross sections are limited to being circular and orthogonal to the local axis directions.

We will refer to the combined class of PRCGCs and circular PRGCs as PRGCs.

These primitives can join to form complex objects in a number of ways. We restrict the joints between two parts to be one of the following two classes (as shown in Fig. 5):

1. **End-to-End Joint:** Here, one of the two ends of each of the joining parts are attached to each other, as in the case of the lid and the container in Fig. 5. As the ends are planar, this implies that the two cross sections share a common plane and some common area; however, the two need not be of the same size.
2. **End-to-Body Joint:** Here, the end of one part is attached to the "side" of another, as is the case for the spout and the handle joined to the container in Fig. 5. Strictly speaking, a planar end cannot be mated precisely with a curved side and some "glue" is required in between. We do not explicitly model this glue.

Obviously, not all objects in our environment are well modeled by the above types of parts and joints. However, we do believe that they cover a large and interesting class of objects of our everyday experience. We have selected these classes based on their wide applicability but also because it has been possible to derive strong projective properties for them.

The target detection level of the system is described in detail in [23] and [24]. The approach consists of computing a hierarchy of description levels obtained by boundary grouping, symmetry grouping, and surface patch grouping. A hypothesize-verify process is used which proceeds in a

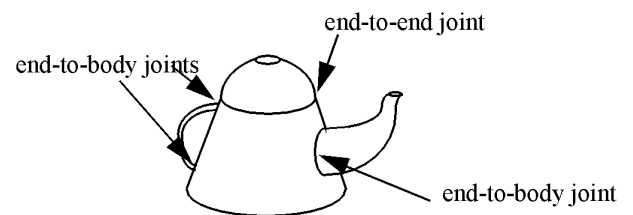


Fig. 5. Examples of joints between parts.

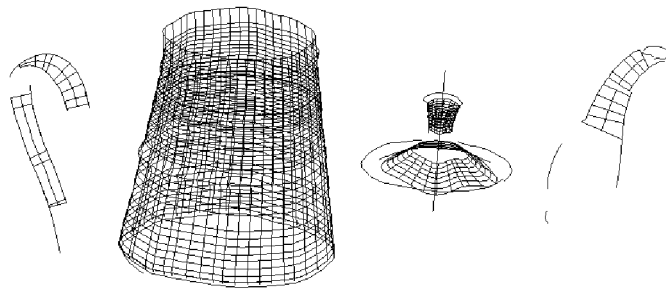


Fig. 6. Parts detected from the edges shown in Fig. 2.

bottom-up, local-to-global fashion. This process detects local features (such as surface fragments), groups those likely to project from the same part, and verifies the resulting part hypotheses. An important aspect of the method is that the constraints used in these steps are derived from rigorous geometric invariant and quasi-invariant, and structural, properties of the contours of the subclasses of GCs. This allows a systematic account of the effects of partial occlusion, image fragmentation, and extraneous boundaries (markings, shadows, and so forth). Random surface markings, for example, are unlikely to satisfy the strong constraints at all levels of the hierarchy. We omit the details of this process as they have been reported in previous work cited above.

For the example of the image in Fig. 1, the part level detects four parts as shown in Fig. 6. For each part, we display its axis and cross sections at selected intervals; even though such depiction appears three-dimensional to us, only 2D projections have been computed so far. Note that

there are some gaps in the descriptions of the handle and the lid and for the lower part of the spout, no axis or cross section are determined even though the limb boundaries are found (this is partly due to the fact that we do not see the cross section at the lower end).

The part level detects SHGCs and PRGCs separately and it is possible that the same part allows descriptions as both types of classes. In some situations, this type of ambiguity can be resolved at the part level. For example, the large cone of Fig. 6 is both an SHGC with a circular cross section and linear sweep rule, and a PRGC with a straight axis and linear sweep rule. In this case, the SHGC interpretation is favored (i.e., when the axis is straight, an SHGC interpretation is favored). However, in other situations disambiguation needs a more complete context involving other part descriptions in the vicinity of the ambiguous part, this is described in Section 3.2.

At the object level, described in this paper, additional analysis are performed to generate compound objects

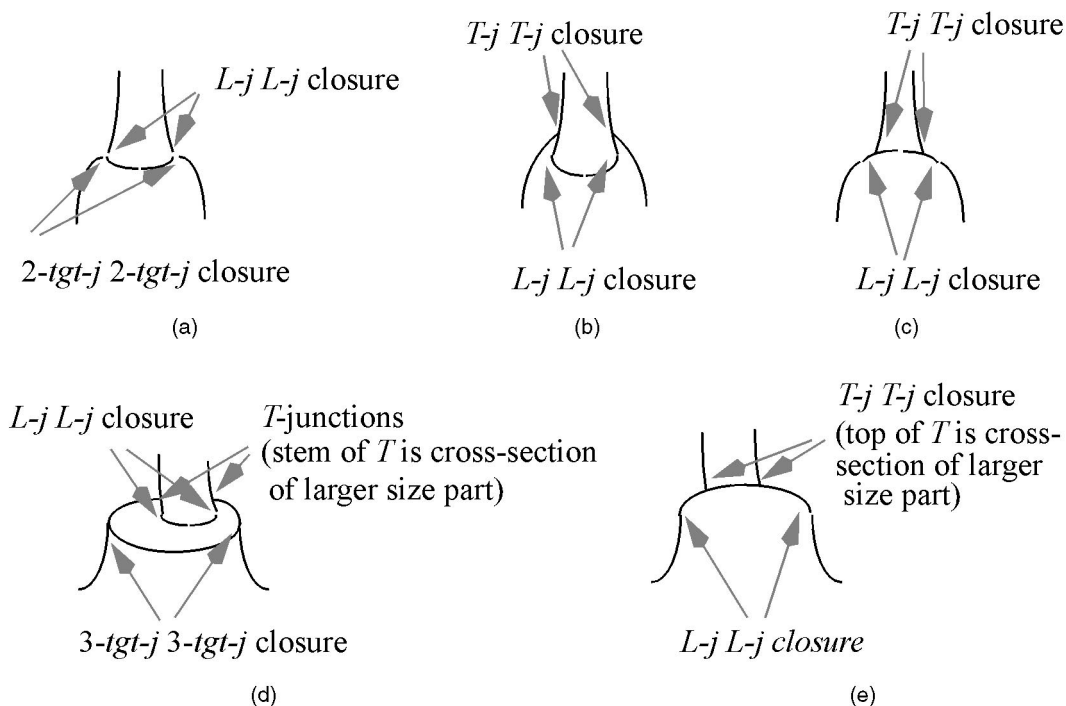


Fig. 7. Structural relationships for end-to-end joints. Equal-size ends (a through c) and different-size ends (d and e).

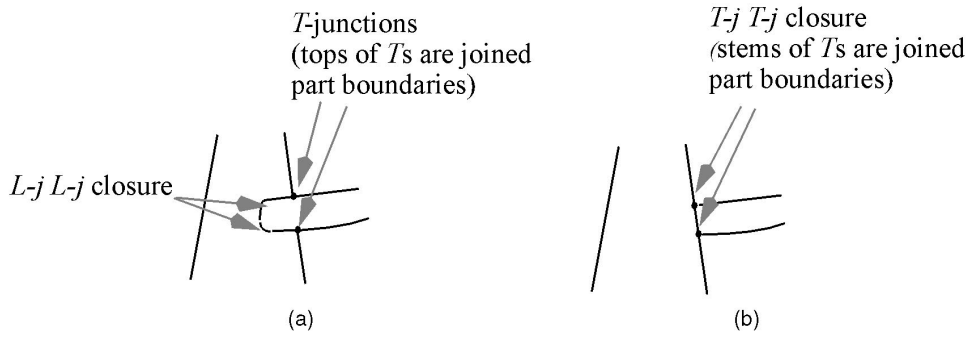


Fig. 8. Structural relationships for end-to-body joints. (a) Visible joint curve. (b) Nonvisible joint curve.

descriptions. The object level extends the geometric context created by parts hypotheses to one where the relationships between parts are exploited to identify compound objects. For two parts to be described as being connected, the proposed joint(s) between them should satisfy the properties expected from joining such parts, as described in Section 3. The type of joint is also determined by this process. In addition, the analysis of relations between parts helps further reduce the multiple individual interpretations that may have survived from the parts level and combine constraints from both parts and joints in completing missing shape information. The result of the object-level analysis are part-based descriptions in the form of a graph where nodes are projected GC descriptions of parts and arcs encode observed joint relationships between the parts. As the 2D descriptions correspond to some 3D GCs of a specific class, we can also recover their 3D shape descriptions. The 3D shape inference method combines constraints on the 3D parameters of each GC part from its projective GC description and from the joint relationships between parts. The joint relations can often help reduce the degrees of freedom inherent in making the 3D inferences. For example, if the projective descriptions of two joined parts indicate that they have parallel cross section cuts (say, from some symmetry relationships), then their 3D cross sections (or axes) must lie on parallel planes.

We next describe the process of detecting compound objects and of inferring 3D shape in detail.

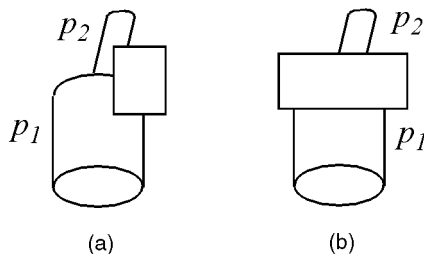


Fig. 9. Joint detection allows for partial occlusion. (a) A joint is marked between parts p_1 and p_2 . (b) No joint is marked.

3 DETECTION OF COMPOUND OBJECTS

At this stage, we have already detected the parts and their 2D descriptions. The parts (or subparts) may have multiple descriptions, some of which may be eliminated when their relations with other parts are examined. The detection of compound objects proceeds in two steps: the analysis of joints and the analysis of ambiguities.

3.1 Analysis of Joints

The objective of this step is to identify potential joint relationships between hypothesized parts (whether there is physical contact between parts cannot be firmly concluded from an image). For this, it is useful to analyze the generic image events corresponding to the two types of joints of Fig. 5 and use them to hypothesize those relationships.

3.1.1 Properties of Joints

The structural properties of these two types of joints consist of the closure patterns of the joined parts and the observed junction relationships between the joined parts' boundaries. The observed structure depends on the viewing direction, shapes of the parts (sweep derivatives and axes curvatures) at the joints and the visibility of their *joint curve* (intersection of their surfaces). In the following, we discuss first the possible image events for each type of joint, then how fragmentation and partial occlusion are dealt with.

1) End-to-End Joints. There are two cases for an end-to-end joint: the two parts have the same size at their contact (Figs. 7a, 7b, and 7c) or have different sizes (Figs. 7d and 7e). For any viewing direction (except the specific one which lies in a part's cross section plane), at the joined ends, one part will have its cross section facing "toward" the camera and the other will have its cross section facing "away" from the camera. The observed closure junctions of a part when it faces "toward" or "away" from the camera are combinations of the set of junctions of [10]: *Three-tangent junctions* (3-*tgt-js*), *L-junctions* (L-*js*), *T-junctions* (T-*js*), and *cusps*.¹ For example, a closure with two L-*js* indicates a cross section facing "away" from the camera (for example, the bottom of the image of a cylinder), and a closure with two 3-*tgt-js* indicates a cross section facing "toward" the

1. In the current analysis, Arrow and Y-junctions are not used, though, they do not present any conceptual difficulties.

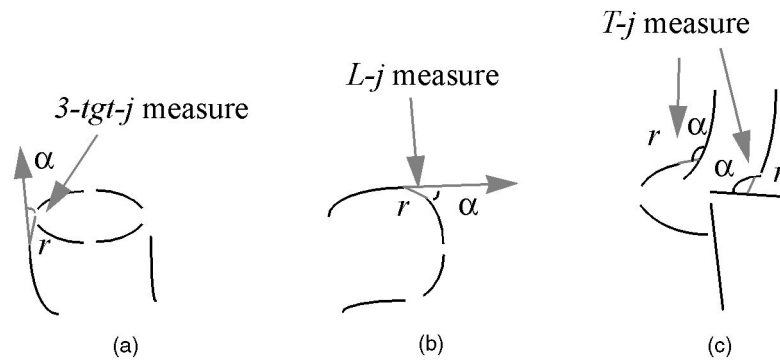


Fig. 10. Junction measures.

camera (the top of the image of the cylinder). In each case, there are other junction combinations involving “cusps” and T -js. In the analysis below, we use the L -j L -j combination to represent any of the closure patterns that result from the cross section facing “away” from the camera and the 3- tgt -j 3- tgt -j combination to represent any of the closure patterns that result from the cross section facing “toward” the camera. Both combinations can be replaced by any of the set of combinations they represent.

An additional analysis is needed for the observed junctions at occlusion regions between the joined parts. To analyze the effects of joints, we separate the cases of equal-size ends joints and different-size ends joints. In each case, we first consider the case where the intersection curve between the parts (*joint curve*) is visible, then the case where it is not visible.

Equal-Size Ends: *Case 1. The Joint Curve is Visible.* In this case, the joint curve is the common cross section. Therefore, the part with its cross section facing away from the camera will have a regular L -j L -j closure (top part in Figs. 7a and 7b). The part with its cross section facing toward the camera cannot have 3- tgt -j 3- tgt -j closure because the other part partially occludes its cross section. Only the portion of its cross section that coincides with the visible joint curve will be observed. The observed junctions of this part depend on whether its junction points are close to the other part’s or not.

In case they are, then this has the effect of preventing the upper branch of the 3- tgt -j to be visible, resulting in a sort of “two-tangent junction” (henceforth, 2- tgt -j) as can be seen for the bottom part of Fig. 7a. In the case where the junction points are relatively distant (such as due to a flaring of the part with its cross section facing toward the camera), then the part with its cross section facing away will partially occlude the other part’s closure, causing this latter part to have T -j closures (bottom part of Fig. 7b).

Case 2. The joint curve is not visible. The only case where the joint curve is not visible is when the part with its cross section facing away from the viewer occludes the other part (Fig. 7c). This happens, for example, when the occluding part has a flaring (high-derivative sweep). This case corresponds to the case of Fig. 7b seen “from behind.”

Different-Size Ends. *Case 1. The joint curve is visible.* In this case, both parts have regular closures, i.e., the one facing away has L -j L -j closure and the one facing toward

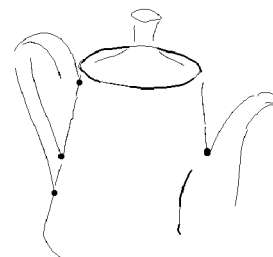
the camera has 3- tgt -j 3- tgt -j closure. However, unless the former part is too small compared to the cross section of the latter part, T -js are observed between the side boundaries of the smaller-end part and the cross section boundaries of the large-end part. See Fig. 7d.

Case 2. The joint curve is not visible. This occurs when the larger part’s cross section occludes the smaller part, causing the former to have L -j L -j closure and the latter to have T -j T -j closure. See Fig. 7e.

2) End-to-Body Joints. In the discussion below, we will call joined part the part joined at its end and supporting part the one joined at its body. As before, we consider two cases depending on the visibility of the joint curve.

Case 1. The joint curve is visible. In this case, the joint curve will constitute the joined part’s closure. The latter part cannot have its cross section facing toward the camera. This is because the surface of the supporting part faces the viewer (otherwise the surface and the joint curve will not be visible). Thus, the joined part will have L -j L -j closure (as before, any of the closure patterns corresponding to the cross section facing away from the camera can be observed instead). Further, unless the joined part is too small, it will partially occlude the supporting part’s side boundaries, causing T -js, where the top of the T s are the joined part’s side boundaries. This case is illustrated in Fig. 8a.

Case 2. The joint curve is not visible. In this case, the supporting part occludes the joined part, resulting in T -j closures for the latter where the stems of the T s are the joined parts side boundaries. This case is illustrated in Fig. 8b.

Fig. 11. Detected joints for the image of Fig. 1. (T -junction points are displayed with large dots and junction curves in bold lines).

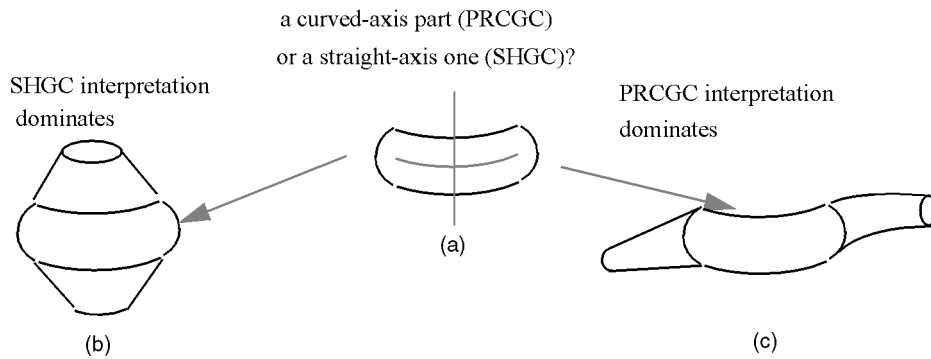


Fig. 12. An example of geometric ambiguity.

3.1.2 Detection of Joints

The detection of joints involves finding the structural patterns described previously at the end of each part with other parts in its vicinity. The analysis between a pair of parts is based on the parts closure patterns, the junctions between the parts' boundaries and on an "extent analysis" described below as a set of rules. We use the terms "away closure" to denote a closure pattern corresponding to a cross section pointing away from the camera and "toward closure" to denote a closure pattern corresponding to a cross section pointing toward the camera.

- If two parts share closing boundaries such that one has an "away closure" and the other "2-tgt-j closure" and "similar" sizes at their ends then an end-to-end joint with equal-size ends is hypothesized (pattern of Fig. 7a).
- If a part has T -j closures with another part's (occluding part) side boundaries such that the T -junctions are on opposite sides of the occluding part and the occluding part has "away closure" with "similar" sizes at the point of contact between the two parts, then an end-to-end joint with equal-size ends is hypothesized (pattern of Fig. 7b).
- If a part (occluding part) has "away closure" and its side boundaries partially occlude another part's (occluded part) cross section boundaries such that the occluded part has "toward closure" and the closure boundaries of the occluding part lie completely within the region delimited by the occluded part's cross section, then an end-to-end

joint with different-size ends is hypothesized (pattern of Fig. 7d).

- If a part (occluded part) has T -j closures such that the top of the T s are another part's (occluding part) cross section boundaries and the occluding part has "away closure" with a larger size end than the occluded part's, then an end-to-end joint with different-size ends is hypothesized (pattern of Fig. 7e).
- If a part (occluding part) has "away closure" and has T -junctions with another part's (occluded part) side boundaries such that the stems of the T s lie on the same side of the occluded part and the closing boundaries of the occluding part lie inside the surface of the occluded part, then an end-to-body joint is hypothesized (pattern of Fig. 8a).
- If a part (occluded part) has T -j closures such that the top the T s are another part's (occluding part) side boundaries lying on the same side of the occluded part, then a joint of type end-to-body is hypothesized (pattern of Fig. 8b).

Notice that the structural patterns of Figs. 7c and 7e are similar. Thus, the above algorithm will only hypothesize the one of Fig. 7e, i.e., does not mark the joint as "same-size ends" due to the lack of evidence. This is not critical though, since the main decision here is the end-to-end interaction.

The conditions on the part's closures above allow for partial occlusion at one side, at most, of a part's end. For example, the case of Fig. 9a is accepted as a joint, whereas the one of Fig. 9b is not.

In order to handle real image imperfections such as fragmented edges, the junction detection algorithms make

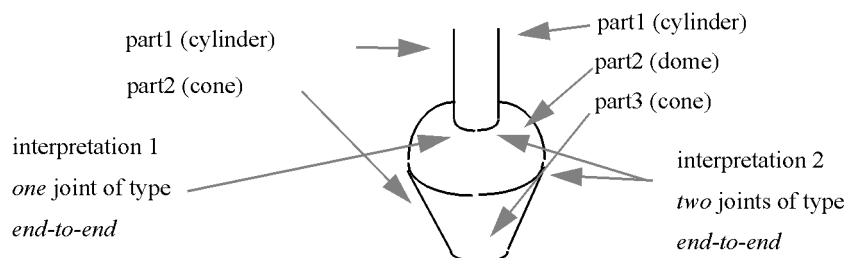


Fig. 13. An example of structural ambiguity.

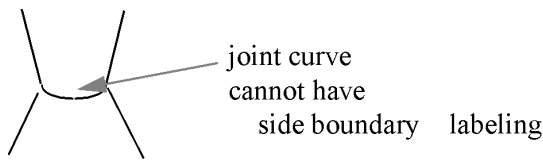


Fig. 14. Visible joint boundaries suggest "part-end" boundary labelling.

use of junction measures (based on angular and proximity relationships) so as to account for possible boundary breaks. Fig. 10 illustrates these for the three types of junctions used (3-*tgt-j*, *L-j* and *T-j*). For each junction, both proximity (relative measure of the junction gap) and the angle must be smaller than given thresholds (system constants).

Fig. 11 shows the result of the joint detection on the parts found in Fig. 6. *T*-junction points are displayed as large dots and the junction curves in bold lines. Four joints are found: One end-to-end joint, with equal-size ends and visible joint curve, between the lower cone (the pot) and the upper SHGC (the lid), one end-to-body joint with visible joint curve between the right-most circular PRGC (the spout), and the pot and two end-to-body joints with nonvisible joint curve between the left-most PRGC (the handle) and the pot. The joint labels can be seen in Fig. 15.

3.2 Analysis of Ambiguities

This step attempts to identify conflicting hypotheses (interpretations) and, if possible, disambiguate among them. One source of ambiguities is the conflict between geometric properties, for example, the same symmetry relationships for different GC primitives, as shown in Fig. 12. Another source of ambiguity is the similarity between the structural properties of a part and those of a joint as shown in Fig. 13.

Many ambiguities can be resolved by the joint relationships between hypothesized parts. Joints with visible joint curves can be seen as highly regular structures. They involve not only regular (verified) parts but with specific structural relationships. For such relationships to be observed in the image, without the parts actually being joined in the scene, requires first the existence of a viewing direction that would accidentally produce these relationships, then that the scene be viewed from that specific viewing direction. Thus, the joints can be thought of as non-accidental relationships whose presence in the image suggests certain labeling of boundaries. In case of ambiguities, joint curves are interpreted as parts ends, as shown in Fig. 14. The single part of Fig. 12a taken by itself could be interpreted as either an SHGC or a PRGC. When considered in the context of Fig. 12b, it can only be interpreted as an SHGC and when considered in the context of Fig. 12c as a PRGC part.

If, after application of this filtering, some ambiguities still remain, different object interpretations are generated (as those of Fig. 13). Each such object interpretation is represented as an attributed graph (one for each compound object) whose nodes are the parts and whose arcs are the joints between the parts. The arcs are labeled partially by the type of joints they represent. Although our approach allows for multiple interpretations, in the examples given in this paper, only one interpretation is

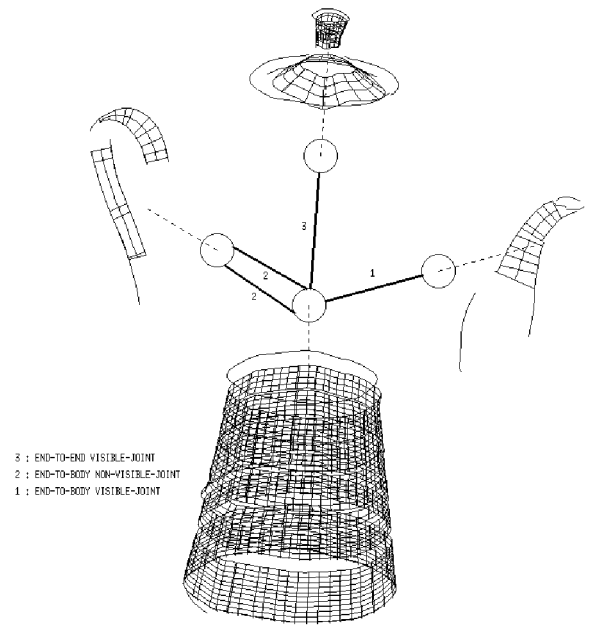


Fig. 15. Resulting graphical representation from the hypothesized parts of Fig. 6.

found for each image. Fig. 15 shows the graph constructed for the parts of object in Fig. 1.

Another example is shown in Fig. 16. Here, the image contains two compound objects (mugs). There is more texture on the objects and in the background than for the teapot example, and there is significant occlusion of one object by the other. Still, the method is able to detect the two mugs and describe each as having two parts. The computed 2D GC descriptions of the parts are also shown in the figure.

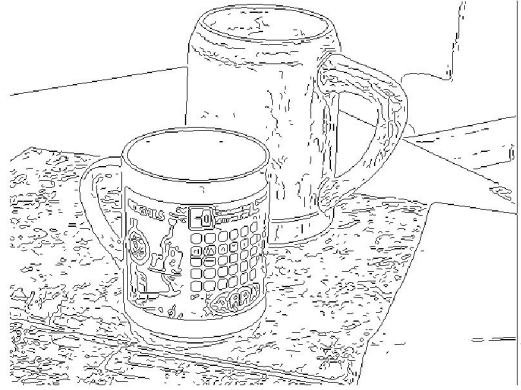
4 INFERENCE OF 3D SHAPE

Inferring the 3D shape of an object requires finding the 3D intrinsic GC description of each of its parts, i.e., its 3D cross section, its 3D axis and its scaling function. A complete method for such 3D recovery would use the information given by each part (its projective description built so far) and by the joints between parts. This is because, in general, the problem is under-constrained and needs the use of additional assumptions (preference criteria). Those assumptions should be consistent with the observed interactions between the surfaces of a part and between the parts of an object.

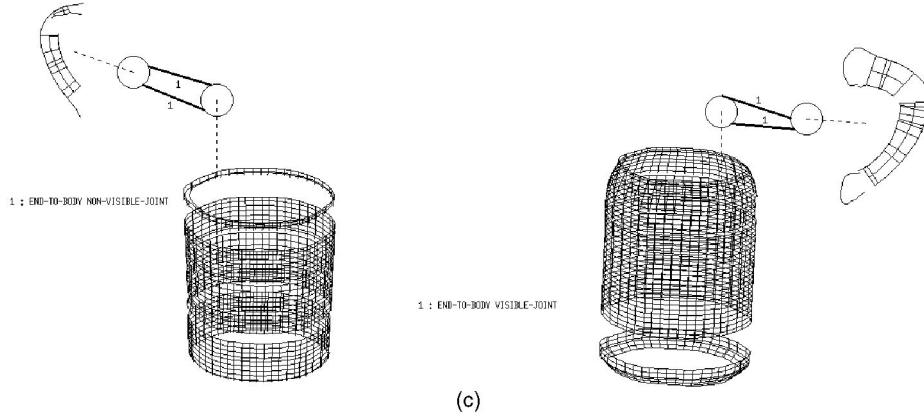
The 3D shape inference method we use exploits both types of interactions. It uses intrapart constraints as well as interpart constraints. The intrapart constraints consider each part as though it were isolated. They come from relationships between the observed contours such as those forming symmetries and relations between limbs and cross section curves. We omit a description of these as they have been presented in detail in past work [23], [24]. The interpart constraints establish relationships between 3D shapes (orientations) of joined parts' cross sections. We will discuss the use of joints for this



(a)



(b)



(c)

Fig. 16. Another result of the compound object segmentation method.

purpose and also for shape completion in Section 4.2. But first, in Section 4.1, we discuss the classification of parts, a useful step in inferring 3D shape.

4.1 Classification of Parts' Cuts

This step uses the parts' descriptions to infer useful information about their cuts (properties of their end cross sections) when the cross section is fully visible. Such information includes whether a visible part's cut is likely to be cross sectional² or planar. This classification helps decide what constraints a part's image description imposes on its 3D shape. The cut classification can be inferred from the observed image properties of a part as follows (the properties are known to hold for the image of 3D parts; here, they are used to make the inverse inference):

- If the part is an SHGC with *linearly parallel symmetric*³ ends then the cuts at both ends are assumed to be cross sectional (implicitly planar).

2. A part may not be cut along its cross section (for example, an oblique cut of a cylinder).

3. Linear parallel symmetry is the extension of straight line parallelism to curved lines such that the symmetric curves are similar up to scale and translation [20].

- If the part is an LSHGC with *line-convergent*⁴ symmetric ends then the cuts are assumed to be planar (not both cross sectional).
- If the part is an SHGC with neither of the above symmetries, then if it has a full cross section it is assumed to be a cross sectional cut (for example, when the bottom end is completely occluded). See Fig. 17.
- If the part is a PRGC with an elliptic cross section⁵ at some end, and with an extremal 2D axis tangent orthogonal to the major axis of the ellipse then the cut is assumed to be cross sectional (implicitly planar). See Fig 17b.
- If the part is a PRGC with an elliptic cross section not orthogonal to the extremal 2D axis tangent then the cut is assumed to be planar but not cross sectional. See Fig 17c.

4. Line convergence is a form of symmetry whereby tangent lines at symmetric points of two curves intersect along a line [22]. It can be thought of as the generalization of point incidence to line incidence.

5. A curve is judged to be (approximately) elliptic if the fitting error is small. This error is chosen to be the minimum of the least squares fit or the scatter fit [20].

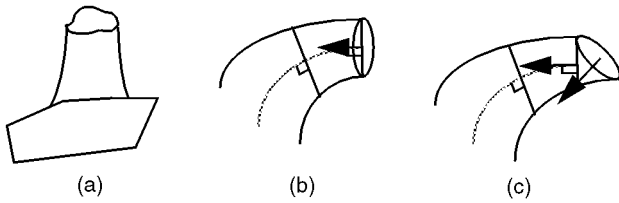


Fig. 17. Examples of classification of fully visible parts' cross sections.

The results of this classification are cross section labels (cross sectional and/or planar cut) for each part of every object. This label is used in the intra-part constraint generation for 3D shape recovery. The constraints depend also on each part's primitive type (SHGC, PRCGC, or circular PRGC) [20], [22], [23], [24].

4.2 Exploiting Joints to Infer 3D Shape

Joined parts set mutual constraints on one another's 3D shape. For example, well-defined differential geometric relationships exist between the surface orientations of intersecting surfaces. Their surface normals, along their intersection curve, are both orthogonal to the tangent to that curve [20]. The problem of exploiting such constraints from a single image of arbitrary joint relationships between parts is a research topic in its own right (see [21] for an approach that addresses objects with multiple Zero-Gaussian Curvature surfaces). For this reason, we limit the analysis to the use of end-to-end joints with visible joint curves as they can be exploited from a single image.

There are two ways end-to-end joints can be exploited in 3D shape recovery. The first is to set mutual constraints on the cross section plane orientations of joined parts. The second is to infer invisible (or not completely visible) cross sections for parts with incomplete descriptions. We discuss them separately.

4.2.1 Using End-to-End Joints to Constrain 3D Cross Section Orientations

End-to-end joints with visible joint curves indicate that the joined parts are cut by the same plane. Therefore, a 3D description of the joined parts must be consistent with this constraint. For this, sets of joined parts' ends which must have the same plane orientation at their ends are constructed (some of these sets may contain only one element).

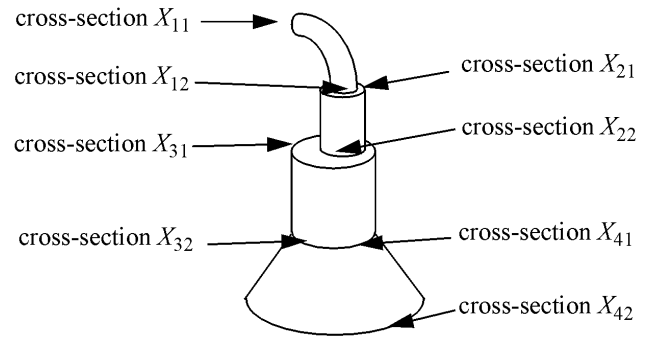


Fig. 18. Finding parts' ends which have the same plane orientations.

This is done by traversing the object graph through the arcs labeled as "end-to-end with visible joint curve" and through nodes labeled as having "parallel cuts," such as SHGCs with linearly parallel symmetric ends. An example is shown in Fig. 18. In this example, two sets are obtained

$$S_1 = \{X_{11}\}$$

$$S_2 = \{X_{12}, X_{21}, X_{22}, X_{31}, X_{32}, X_{41}, X_{42}\}.$$

To estimate the common plane orientation of the parts' ends of each of the sets so obtained, the method first estimates the orientation of each part's end, independently of the others, using a method based on ellipse fitting described in [20]. Then the common orientation of those parts' cross sections is set to be the average of all the obtained orientations. The common cross section orientation of the lid and the pot of Fig. 21a is so computed.

At this stage, all visible cross section orientations have an estimate of their plane orientations consistent with the observed end-to-end joints with visible joint curve. Parts with at least one (almost) completely visible cross section facing toward the camera can be used to recover their 3D shape by the methods of [20], [22], [23], [24]. The remaining parts do not have a fully visible cross section at either end. For some of these parts, some way of inferring the missing cross sections is needed in order to be able to recover their 3D shapes. For this, we find the end-to-end joints with *equal-size ends* (and visible joint curve) to be useful. This is discussed below.

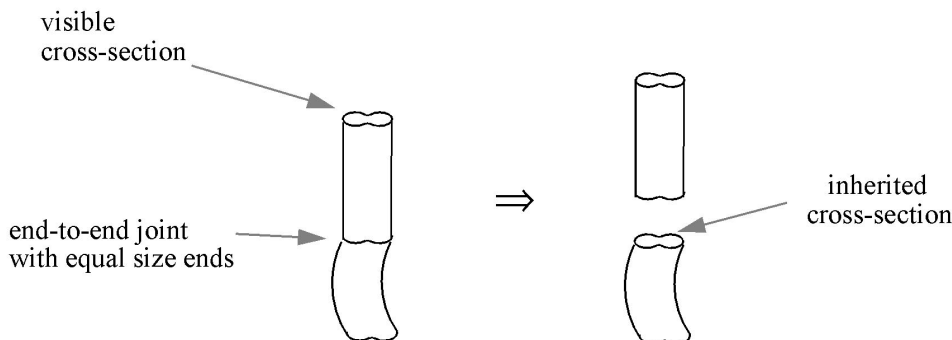


Fig. 19. Using end-to-end joints to infer (full) cross sections for parts with partially visible cross sections.

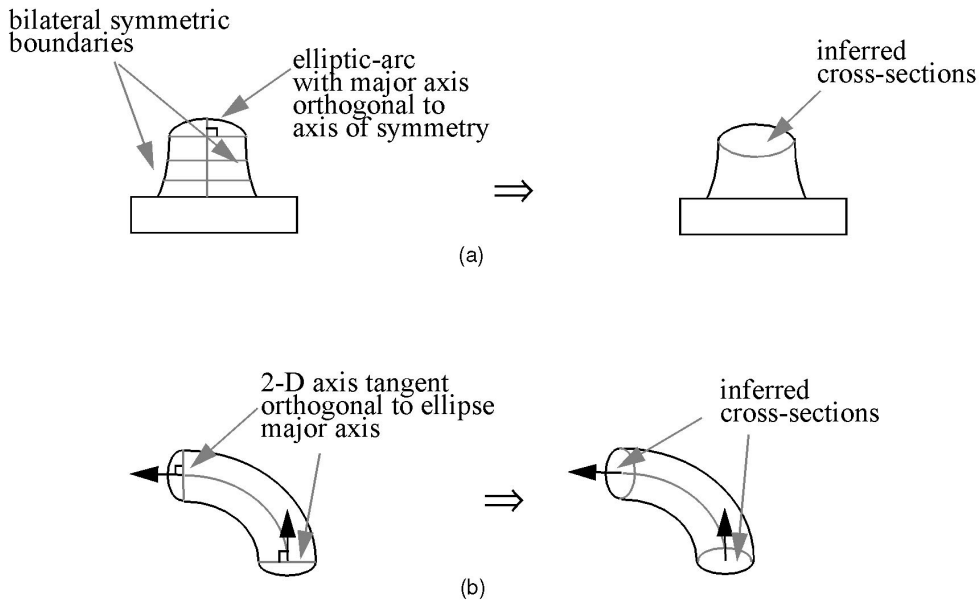


Fig. 20. Assumption of cross section circularity from surface symmetry of parts with partially visible (elliptic) cross sections. (a) Bilateral symmetry for SHGCs. (b) Ellipse major axis orthogonal to extremal 2D axis tangent.

4.2.2 Using End-to-End Joints with Equal-Size Ends to Complete Parts' Descriptions

Parts with only partially visible cross sections and which are involved in an end-to-end joint with equal-size ends may have their cross sections "inherited" from the joined part if the latter has a complete 3D description. Thus, parts whose 3D shape cannot be recovered if considered isolated, may be recovered after this inference is made. In the case such joint-based inference is not possible, assumptions about the circularity of the cross sections are made whenever consistent. The different cases are discussed below.

- If a part P_1 has an end-to-end joint with equal-size ends with another part P_2 and if the cross section of P_1 at the joined end is known (complete) then it is used as the complete cross section of P_2 at its joined end (see Fig. 19).
- If an SHGC part has an elliptic arc as the partially visible cross section and bilaterally symmetric side boundaries with the axis of symmetry orthogonal to the ellipse major axis, then use the ellipse as the part's cross section (see Fig. 20a).
- If a PRGC part has an elliptic arc as the partially visible cross section and an extremal tangent to the 2D axis (almost) orthogonal to the ellipse major axis, then use the ellipse as the part's cross section (see Fig. 20b).

An example of the first rule is given in Fig. 19, where the "bottom" cross section of the cylinder is known (since it can be recovered by the method of [23]) and thus can be used as the "top" cross section of the lower part. We find the other types of joints not strong enough to be exploited in a similar way from a single image.

The second rule is based on the bilateral symmetry property of surfaces of revolution [13] and the third rule is

based on a similar symmetry derived for curved-axis primitives in [24]. Both rules infer circular cross sections (elliptic in the image) from the symmetry and the partial cross-section observed. Fig. 20 gives some examples. If the symmetries, for example, are not observed between the part's side boundaries then no inference is made and the part's description remains in terms of its surface correspondences only, without complete information about its cross section. Its description remains projective.

Each time the cross section of a part is inferred in any of the above ways, the 3D shape of the part can be recovered and the new information is propagated through the visible end-to-end joints with equal-size ends. In doing so, full advantage of such joints is taken for shape inference.

Fig. 21a shows the recovered 3D shapes for the parts of the object in Fig. 15. The recovered 3D parts are displayed in terms of cross sections, meridians and axes and are shown for different 3D poses. Fig. 21b shows the recovered 3D parts for the objects in Fig. 16. Notice that the handle of the teapot of Fig. 15 and the handle of the left-most mug of Fig. 16 could not be recovered in 3D because their cross sections are completely occluded by the pot and the cup, respectively, (the 3D recovery methods of [20], [22], [23], [24] require visibility of both body and cross section).

Completion methods are used when no joint constraints exist for a part, such as the spout of the teapot shown in Fig. 21. In this case, certain assumptions are made about the missing part shape if either one end or one side is visible. The 3D shape completion interpolates the gaps in the axis description by a conic (whenever geometrically possible) and the gaps in the scaling function by a linear scaling function. This process (completion of parts based on intrapart information) is described in more detail in [24]. In some cases, 3D

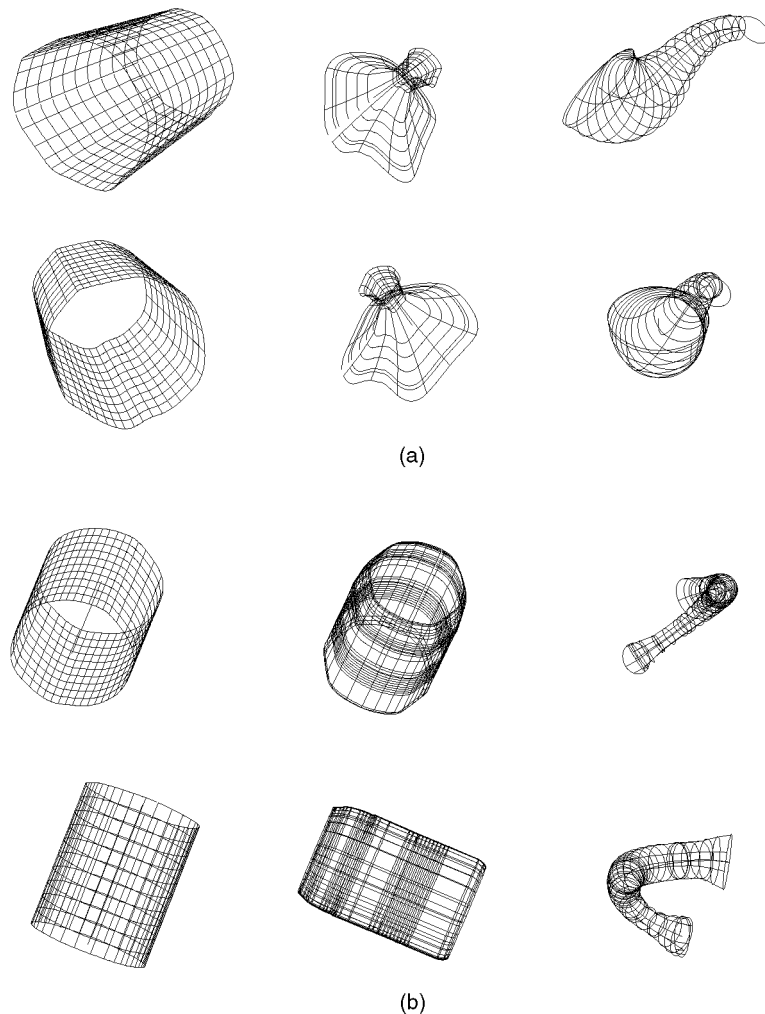


Fig. 21. Recovered 3D descriptions, displayed from two viewing directions each, for the parts in Fig. 6 and for the parts in Fig. 16b.

descriptions may remain incomplete but, their projective descriptions still provide important shape information.

5 DISCUSSION

The object level exploits the projective structural properties of the joints in order not just to infer compound objects, but also to provide global context for both segmentation and shape description. The use of joints with visible joint curves for filtering out false hypotheses, which the part level accepted, is an example of their use for a more complete solution to the figure/ground problem. The use of end-to-end joints with equal-size ends and visible joint curves to infer missing cross sections and the use of end-to-end joints with visible joint curves to estimate common 3D cross section orientations also exploit that context for (3D) shape description.

However, the method may not handle two-dimensional compound objects (involving flat GCs) where the joints do not form any of the patterns discussed in Section 3.1 as shown in Fig. 22. This type of object may better described by a method suited to 2D objects such as one given in [17]. A

combination of both methods would increase the set of objects that can be handled.

The performance of the whole method, including the SHGC and PRGC modules, on the images shown is summarized in Tables 1, 2, 3, 4, 5, and 6. The tables show the evolution of the number of hypotheses of both modules and of key steps of the object level and the corresponding run times (on a SUN SparcStation 10 using COMMONLISP). The final number of parts given is the one obtained after the part verification step between both SHGC and PRGC parts.

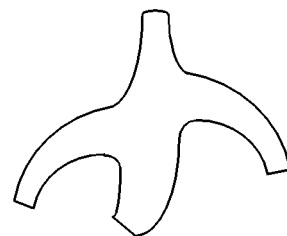


Fig. 22. A flat object.

TABLE 1

Evolution of the Number of Hypotheses of the SHGC Module

scene	boundaries	hypothesized cross-sections	local SHGC patches	parts hypotheses	verified parts hypotheses
fig. 3.5	850	3	9	6	2
fig. 3.10	2057	6	18	17	2

TABLE 2

Evolution of the number of Hypotheses of the PRGC Module

scene	parallel symmetries	local PRGC patches	parts hypotheses	verified parts hypotheses
fig. 3.5	3610	84	68	2
fig. 3.10	3783	105	89	2

TABLE 3

Evolution of the Number of Hypotheses of the Object Level

scene	number of parts	number of joints	number of objects
fig. 3.5	4	4	1
fig. 3.10	4	4	2

It can be seen that the object level processing itself is very fast. This is to be expected because there are few parts hypotheses given as input to that level, making the joint detection fast. Also, the 3D inference methods are purely numerical and take little time to execute. The bulk of the processing takes place at the part level. The second scene (Fig. 16) takes more time to interpret due to its higher complexity (more markings both within the objects and the background).

Finally, the parameters used in the method, both in part-level and object-level processing, are all constant. The robustness of the method to changes of key parameters has also been tested by modifying their values by up to 50 percent of their default values. The final results were unchanged.

6 CONCLUSION

Although the extraction of 3D part-based descriptions from a single 2D image is an old problem, past work has either assumed perfect and segmented boundaries or addressed 2D shapes. The method described in this paper relaxes these limiting assumptions and directly exploits the 3D nature of the viewed objects. It has successfully extracted generic 3D volumetric descriptions composed of certain subclasses of generalized cylinders and of their joint relationships, in the presence of noise, boundary breaks, markings, shadows, and partial occlusion. This capability is achieved by the use of a bottom-

TABLE 4

Run Times of the SHGC Module (min:sec)

scene	cross-section finding	local SHGC patch detection	grouping of local patches	verification of SHGC parts
fig. 3.5	0:55	0:15	0:00.6	0:06
fig. 3.10	4:30	0:42	0:00.9	0:27

TABLE 5

Run Times of the PRGC Module (min:sec)

scene	parallel symmetry	local PRGC patch detection	grouping of local patches	verification of PRGC parts
fig. 3.5	1:40	0:54	0:14	3:31
fig. 3.10	2:49	3:52	0:25	15:10

TABLE 6

Run Times of the Object Level (min:sec)

scene	joint detection	3-D SHGC inference	3-D PRGC inference
fig. 3.5	0:00.5	0:17	0:00.1
fig. 3.10	0:00.6	1:24	0:00.1

up hierarchical approach where a local-to-global geometric context is built and refined. The method also takes explicit account of the 3D nature of the objects by exploiting rigorous geometric and structural properties of their projections. This allows the inference of 3D shapes when enough structure is visible in the image.

The scope of the method is large, we believe, as the subclasses of generalized cylinders it addresses capture the (approximate) 3D shape of many man-made (and some natural) objects in our environment. The resulting descriptions, either the 3D intrinsic elements of a GC or their projective descriptions, can be used to provide powerful, view-insensitive, indexing keys to large databases of object models for object recognition. They also provide means to analyze similarities and differences in shape, a highly useful capability for learning. The descriptions also have applications in grasping, where the shapes can be used to plan for the grasp and preshape the hand.

The method can be extended in many ways. First, a parallel implementation should not be difficult given the organization of the search processes at the part and object levels. It can also be extended to exploit multiple cues in order focus the search or reduce ambiguity in 3D recovery. The ideas and principles of the method can be used in a stereo (or motion) framework where the high-level descriptions obtained in each image can be used to

match features across views and reconstruct, with fewer assumptions, the 3D shapes of the detected objects.

ACKNOWLEDGMENTS

This work was supported by a grant from the Defense Advanced Research Projects Agency, Grant No. F49620-93-1-0620, monitored by the Air Force Office of Scientific Research.

REFERENCES

- [1] R. Bergevin and M.D. Levine, "Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, pp. 19-36, 1993.
- [2] I. Biederman, "Recognition by Components: A Theory of Human Image Understanding," *Psychological Review*, vol. 94, no. 2, pp. 115-147, 1987.
- [3] T.O. Binford, "Visual Perception by Computer," *IEEE Conf. Systems and Controls*, Miami, Fla., Dec. 1971.
- [4] R.A. Brooks, "Model-Based Three-Dimensional Interpretations of Two-Dimensional Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 140-150, 1983.
- [5] S.J. Dickinson, "3D shape Recovery using Distributed Aspect Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 174-198, 1992.
- [6] S.J. Dickinson, D. Metaxas, and A. Pentland, "The Role of Model-Based Segmentation in the Recovery of Volumetric Parts from Range Data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 259-267, 1997.
- [7] A. Lejeune and F.P. Ferrie, "Finding the Parts of Objects in Range Images," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 230-247, Sept. 1996.
- [8] A.D. Gross and T.E. Boulton, "Recovery of SHGCs from a Single Intensity View," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 161-180, Feb. 1996.
- [9] J.E. Hummel and I. Biederman, "Dynamic Binding in a Neural Network for Shape Recognition" *Psychological Review*, 1992.
- [10] J. Malik, "Interpreting Line Drawings of Curved Objects," *Int'l J. Computer Vision*, vol. 1, no. 1, pp. 73-103, 1987.
- [11] D. Marr and H.K. Nishihara, "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes," *Proc. Royal Physical Soc.*, pp. 269-294, 1977.
- [12] R. Mohan and R. Nevatia, "Perceptual Organization for Scene Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1992.
- [13] V. Nalwa, "Line Drawing Interpretation: Bilateral Symmetry," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 1,117-1,120, 1989.
- [14] R. Nevatia and T.O. Binford, "Description and Recognition of Complex Curved Objects," *Artificial Intelligence*, vol. 8, no. 1, pp. 77-98, 1977.
- [15] Q.L. Nguyen and M.D. Levine, "Representing 3D Objects in Range Images Using Geons," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 158-168, Jan. 1996.
- [16] A. Pentland, "Recognition by Parts," *Proc. Int'l Conf. Computer Vision, ICCV*, pp. 612-620, 1987.
- [17] K. Rao and R. Nevatia, "Describing and Segmenting Scenes from Imperfect and Incomplete Data," *Computer Vision, Graphics and Image Processing J.*, vol. 57, no. 1, pp. 1-23, Jan. 1993.
- [18] L. Roberts, *Machine Perception of Three-Dimensional Solids*. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1965.
- [19] H. Sato and T.O. Binford, "Finding and Recovering SHGC Objects in an Edge Image," *Computer Vision Graphics and Image Processing*, vol. 57, no. 3, pp. 346-356, 1993.
- [20] F. Ulupinar and R. Nevatia, "Perception of 3D Surfaces from 2D Contours," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 3-18, 1993.
- [21] F. Ulupinar and R. Nevatia, "Recovery of 3D Objects with Multiple Curved Surfaces from 2D Contours," *Artificial Intelligence*, vol. 67, no. 1, pp. 1-28, May 1994.
- [22] F. Ulupinar and R. Nevatia, "Shape From Contour: Straight Homogeneous Generalized Cylinders and Constant Cross Section Generalized Cylinders," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 120-135, Feb. 1995.
- [23] M. Zerroug and R. Nevatia, "Volumetric Descriptions from a Single Intensity Image," *Int'l J. Computer Vision*, pp. 11-42, Oct. 1996.
- [24] M. Zerroug and R. Nevatia, "Three-Dimensional Descriptions Based on the Analysis of the Invariant and Quasi-Invariant Properties of Some Curved-Axis Generalized Cylinders," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 237-253, Mar. 1996.



M. Zerroug received the PhD degree in computer science from the University of Southern California in 1994. He had previously received the Diplome d'Ingenieur in computer science from the University of Science and Technology of Algiers in 1988. He was a research assistant during 1990-1994 and a research associate during 1994-1995 at the Institute for Robotics and Intelligent Systems of the University of Southern California. He has several publications

in the area of segmentation, description, and recognition of 3D shapes and has received several academic awards during his undergraduate and graduate student years. Dr. Zerroug is currently with Adept Technology Incorporated and is active in the research and development of real-time vision-based systems for industrial automation. He is also involved in the areas of real-time client/server architectures, distributed applications, and digital signal processors. His interests include computer vision, graphics, visualizations, distributed processing, and real-time systems. He is a member of the IEEE Computer Society.



R. Nevatia received his BS degree from the University of Bombay, India, and his MS and PhD from Stanford University, Stanford, California, all in electrical engineering. He has been with the University of Southern California, Los Angeles, since 1975, where he is currently a professor of computer science and electrical engineering and the director of the Institute for Robotics and Intelligent Systems. He has authored two books, *Machine Perception* and

Computer Analysis of 3D Curved Objects, and contributed chapters to several others. He has been a regular contributor to the literature in computer vision. His research interests include computer vision, artificial intelligence, and robotics. Dr. Nevatia is a fellow of the American Association for Artificial Intelligence and a member of the Association for Computing Machinery. He is an associate editor of *Pattern Recognition and Computer Vision, Graphics and Image Processing: Image Understanding*. He has served as an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and as a technical editor in the areas of robot vision and inspection systems for the *IEEE Journal of Robotics and Automation*. He also served as a co-general chair of the IEEE Conference on Computer Vision and Pattern Recognition in June 1997. Dr. Nevatia is a fellow of the IEEE.