USC IMAGE UNDERSTANDING RESEARCH: 1994-1995

R. Nevatia, K. Price and G. Medioni*

Institute for Robotics and Intelligent Systems
Computer Science Department
University of Southern California
Los Angeles, California 90089-0273

Abstract

This paper summarizes the USC Image Understanding research projects and provides references to more detailed descriptions of each one. We have undertaken a broad range of research with a major effort directed toward developing image understanding techniques to extract and use high-level descriptions from images. These techniques allow us to infer three-dimensional shape descriptions from one or more intensity images, or from range data. The resulting descriptions are used for recognition and pose estimation. In addition, we briefly report our progress on building detection and change analysis supporting the RADIUS program, incorporation of knowledge representation techniques in computer vision programs, and an effort in implementing vision algorithms on parallel systems.

1 Introduction

This paper summarizes the progress in our research projects since the last Image Understanding Workshop. Much of this work is described in detail in other papers in these proceedings, with this overview giving only a brief description of the detailed efforts.

Our research covers a broad range of separate tasks in image understanding, but the different tasks are highly inter-related and share many common techniques. The long term goal is to develop generic vision systems, that is, systems capable of working in a variety of domains, or at least capable of being reconfigured rapidly to do so. Such systems are needed to reduce the cost of developing highly tailored systems for each task, and to achieve robust performance for even a single task if the imaging conditions change significantly.

We work on several levels of the image understanding task, from the extraction of descriptions directly from images to object recognition using high level descriptions. Some of the work reported here represents efforts at a preliminary stage while others are the results of completed research efforts. The specific topics we discuss in this overview are:

- 3-D descriptions from intensity images,
  - from multiple images,
  - from 3 close views
- Object Recognition:
  - recognition and learning for large datasets,
  - pose estimation of compound objects
- Analysis of range data:
  - from sparse, noisy input,
  - from dense covering

Our RADIUS related efforts involving building detection with one or more images and change analysis, along with the work in knowledge representation for computer vision, and parallel implementations of vision algorithms are covered briefly here with more discussion in the RADIUS overview paper in these proceedings [Nevatia *et al*. 1996].

We first give a very brief outline of our work over the last 15 months, then present results from many of the individual projects with references to the papers that appear elsewhere in these proceedings.

3-D Descriptions from Intensity Images

Generating 3-D descriptions is one of the most difficult, but important, tasks in IU. Scene segmentation is hard, as different types of features such as object boundaries, surface orientation discontinuities, surface markings, shadows and noise cannot be immediately distinguished. We are building on earlier perceptual grouping techniques for high level descriptions and segmented generalized cylinder descriptions. We report on two efforts here:

- Using contour information and three views of a scene to infer the complete shape of objects in terms of generalized cylinders. See Section 2.1.

* This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Grant No. F49620-95-1-0457 and F49620-95-1-0522; the Rome Laboratories under Contract No. F30602-93-C-0064; or the U.S. Army Topographic Engineering Center under Contract No. DACA76-93-C-0014. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

- Using multiple views to infer the structure of a scene. See Section 2.2

Object Recognition

The emphasis of our research is on generic object recognition, that is recognition of an object from a class. Consider, for instance, two different airplanes which have similar features but different geometries. This forces us to use methods beyond the alignment of geometric features, sufficient to estimate the location and orientation of an exactly known rigid object in a scene. Our approach to solve this difficult problem is to derive high level volumetric descriptions from the scene, then to reason about parts and their arrangements. We first show how such descriptions allow us to efficiently retrieve similar objects from a large database of objects, and also to learn new objects by dynamically updating the database. We then demonstrate how these high level descriptions can also be used to estimate the pose of a recognized object.

We report on two efforts:

- Recognition and learning for large datasets. See Section 3.1
- An alignment-like pose estimation technique for a large class of curved objects. See Section 3.2

Analysis of Range Data

The goals of our effort in range image understanding are to generate rich descriptions from sensed 3-D data. These descriptions should be segmented and capture different aspects of the shape of an object, such as volumetric and surface information. One immediate and useful application is the automatic generation of 3-D models from multiple range images.

We are conducting a number of efforts in the analysis of range data, three are described in more detail here:

- Obtaining surface information from very sparse 3-D data, even in the presence of noise and outliers. See Section 4.1
- Using deformable models for surface approximation and object segmentation of complex objects using dense range data. See Section 4.2
- Generating a surface description coupled with the detection of discontinuity edges in the data. See Section 4.3

RADIUS Related Efforts

Our work in aerial image analysis is currently supported under the RADIUS program. This research includes model-based change detection applied to buildings in the scene and detection and description of buildings using one or more views.

The change detection system registers a site model to the input image, then validates the individual objects in the model. Any changes or differences located in the validation step are analyzed to determine if they represent actual changes in the scene. Detected changes are then used to update the site model.

Our automated building detection and description for one or more images) uses features (lines, junctions, L shapes, U shapes, rectangles, etc.) detected in the images by a perceptual grouping process. These provide initial hypotheses that are verified using shadows, walls, height, etc.

These efforts are described in more detail in the RADIUS overview paper in these proceedings [Nevatia 1996], which also provides pointers to the more detailed papers on these topics.

Knowledge Representation in Computer Vision

This work focuses on integrating advanced knowledge representation technology (as represented by Loom) with image understanding technology to develop advanced tools for the generation of vision systems. It is aimed at enhancing current knowledge representation technology to support computer vision tasks in the realm of spatial reasoning and bringing modern knowledge representation technology to computer vision systems through the use of higher level representations. While this work is performed under a separate contract from the RADIUS effort, it is described in more detail in the RADIUS overview [Nevatia 1996] because it shares the same application domain.

Our recent effort has been in recognizing events in a sequence of images (e.g. a field training exercise). Loom is used to describe the event in terms of its subevents. The Loom query mechanism is used to recognize the event when and select which images contribute to the event when given a number of images. This work is described in detail in [Russ *et al.* 1996].

Parallel Implementation of Vision Algorithms

Under another contract, we are exploring the implementation of standard computer vision algorithms on parallel machines. The goals of this work is to develop scalable (both with image size and number of processors) and portable (within similar classes of machines) algorithms. In this project we are implementing parallel versions of the steps of the building analysis programs. Since the application is the building analysis domain, we describe this work in more detail in the overview of RADIUS related efforts [Nevatia 1996].

Ph. D. Graduates

A major product of our research is graduates. In this period, we have had two graduates: C. Liao [Liao 1995], G. Guy [Guy 1995].

Further Information

We make descriptions of our current research efforts available on-line through the WWW, see the URL: "http://iris.usc.edu/USC-Computer-Vision.html." This page contains pointers to current summaries of our research, example results, recent papers, much of our data, and descriptions and code for many of our programs.

2 3-D Descriptions from Intensity Images

It is necessary to compute high level, segmented, volumetric descriptions from intensity images in order to perform further reasoning, such as recognition or learning. This is clearly a challenging task. A promising methodology is to adequately implement principles of perceptual grouping. To be generic, perceptual grouping methods must work not only on imperfect images, but for large classes of objects as well. We view the description module as consisting of a bottom-up process. However, this process makes use of expectations about image properties, which can be thought of as top-down generic knowledge.

An important aspect of this process is its organization. We do not believe it can be solved at a single feature level, for example by finding the best edge detector or boundary grouping method. There is a direct link between the hierarchical nature of the representations and the organization of the description process. This latter should proceed in successive stages where features are detected then grouped to form higher-level descriptions. The feature groups become themselves features of a higher level type which in turn can be grouped, and so on. Each level increases the scope of the interpretation process by building a geometric context based on previous feature groups. Thus, feedback loops can be used from higher levels to lower ones in order to refine the features or their groupings based on the information gathered from the increased scope.

Figure 1 illustrates the organization of the description process. The description module starts from an intensity image and ends at a level where structured representations in terms of parts (generalized cylinders) and their relationships are obtained.

The tasks that need to be performed at each feature level are:

- extraction of features
- grouping of features
- selection (and refinement) of feature groups

Each of the above tasks constitutes a real challenge in itself. Important issues here are which features to extract, what criteria to use in grouping them into meaningful clusters, and how to know which ones are indeed meaningful and which are not.

The role of the generic shape representation scheme in this process is crucial. Knowledge of this scheme allows the derivation of generic constraints (in the form of expectations) in most of the above tasks, especially as we go higher in the hierarchy, to overcome noise and other image imperfections. Part of these constraints are image properties that any element of the class of objects captured by the shape scheme must satisfy if it is observed. In using these properties to extract object descriptions, the extraction process takes explicit account of the shape representation (the desired output of the system) including its dimensionality. In the case of 3-D representations, the expectations are in the form of *projective properties* which bear the 3-D attributes of the shape of the object.

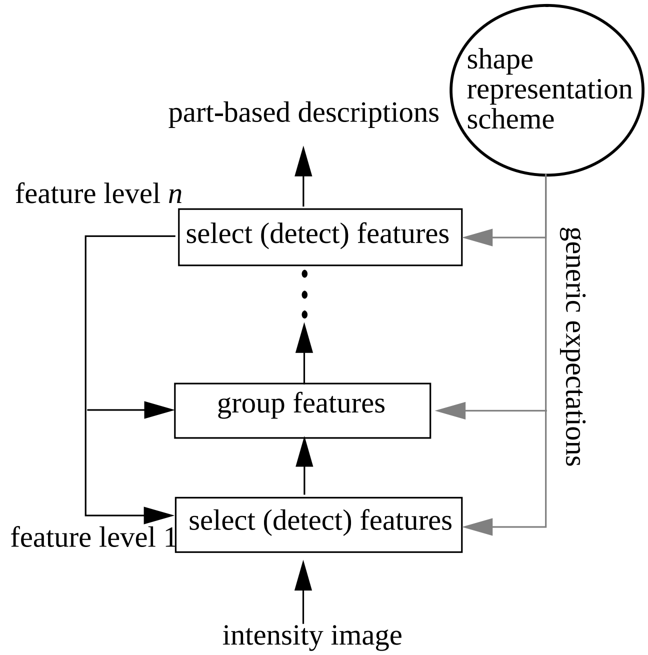


Figure 1. Hierarchical organization of the description extraction process.

Our studies [Zerroug & Nevatia 1994] indicate that we can use these properties to segment objects, fill gaps even in presence of occlusion and infer 3-D shapes from monocular images, as shown in Figure 5.

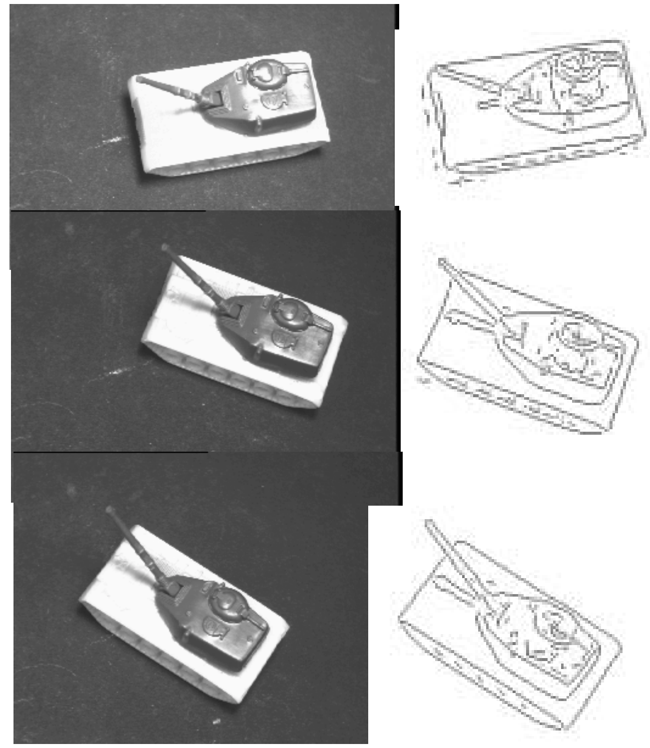
We extend this analysis here where we observe 3 close-by views of a scene. In such a case, it is possible to classify the boundaries observed as belonging to true (physical) contours or to limb (self-occlusion) contours. Also, since some 3-D information is available from matching, it becomes possible to infer more complex classes of GCs than in the monocular case.

2.1 Segmented, Volumetric Shape from Intensity Images

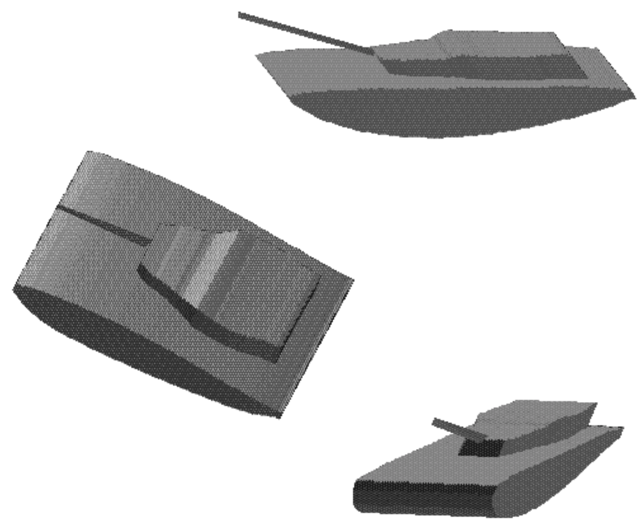
The descriptions we generate in the methodology presented above are volumetric. As such, they are richer than the traditional ones consisting of points, lines, and surfaces. They are also more difficult to extract from the image data.

Whereas the methods described in [Zerroug & Nevatia 1994] addressed the inference of shape from a single image, we use three weakly calibrated (known epipolar geometry) images of an object from slightly different viewpoints as input. In such images, the object is only partially visible. There are parts of the object whose surfaces are facing away from the camera, parts of the object are self occluded. Arriving at volumetric descriptions in these cases, where only partial data is available, requires the development of strong inference rules. Our description is in terms of Generalized Cylinders, and the inference rules are based on the local properties of GCs.

We approach the problem by detecting groups in each image. Grouping is based on proximity, parallelism and symmetry. The GCs, whose volumetric description is to be inferred, give rise to groups in the image. Recovering these groups in the image helps in the volumetric inference of the object parts. The groups in the three images are matched and their contours are labelled as “true” and “limb” edges. Information about the groups, the label associated with their contours (which could be all “true,” all “limb” or a combination of both) to recover visible surfaces. We then use local properties of generalized cylinders to obtain the position of the GC axis and the cross sections to make a volumetric inference. The final descriptions are volumetric and in terms of parts. We demonstrate results on real images of moderately complex objects with texture and shadows. An example is shown below on Figure 2. This work is reported on in more detail in [Havaldar & Medioni 1996].



(a)Three views of a tank and detected edges



(b) Inferred volume from different viewpoints

Figure 2. Shape inference from three views.

2.2 Structure and Motion from a Sparse Set of Views

We address the problem of building complete 3D models for objects from a few intensity images taken at various unknown viewpoints. We aim at recovering the location of all the points visible in the input images with respect to a single reference frame. In contrast to most of the methods that use multiple images, we do not need dense, ordered, image sequences. The input images are taken from viewpoints that show many different aspects of the object. While the burden of processing a large number of images is alleviated, we have to deal with larger change in aspect of the object and larger motion between views. In this case, most features only show up in a couple of views. Moreover, due to large motion between views, it is harder to establish feature correspondences. We propose an algorithm that uses nonlinear least squares fitting to solve for structure and motion parameters without a priori knowledge of the shape or motion, while handling partial or wrong feature matches. The least squares formulation allows us to deal with various camera models, partial or uncertain feature matches across images, and different kind of features, such as points and lines, simultaneously. However, as with all other nonlinear problems, the solution to this minimization involves an iterative process that requires, and is sensitive to initial guesses.

We approach the problem as follows. Images that are taken from close-by viewpoints are clustered together. For each cluster of images, we use the Levenberg-Marquardt algorithm, a general purpose optimization technique, to solve for structure and motion parameters locally by minimizing the image error function. Initial guesses are obtained analytically from pairs of images. Our experiments showed that while the linear motion recovery algorithm sometimes gives good initial guesses that improve convergency, its sensitivity to noise makes it even secondary to simple initialization. We thus use a more robust motion estimation algorithm in our initialization stage. Since the correctness of the initial guesses varies, the correctness of the result for each image cluster, as measured by the image error function, may vary. Good estimates are propagated among clusters to modify the initial guess for local structure and motion estimation. This process iterates until all local estimates are reasonably good. Results from different clusters are then merged together to form a global structure which is then refined by the least squares fitting again. This is illustrated on Figure 3. We describe this work in more detail in [Lee & Medioni 1996].

3 Object Recognition

The broad meaning of "object recognition" in psychology refers to the identification of an object as belonging to a *class* (i.e. a "chair," a "house"), with no implication about specific instances or exact geometry [Biederman 1987]. Developing systems capable of recognizing objects generically is highly useful for computer vision applications, for instance, recognizing two different vehicles which have similar features but different geometries. First, this relieves us from the requirement to use exact models (which may not be available), or even to precisely model all objects of interest. Such systems would account for unknown objects, those which underwent changes (such as articulation, or removal of parts), and would handle a large set of objects (provided adequate representations are used).

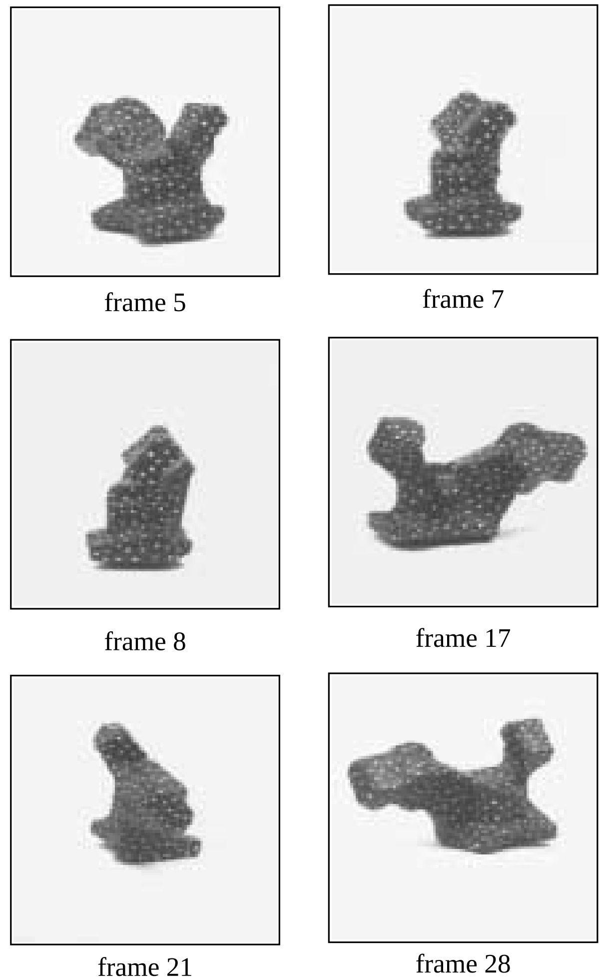
We propose that the way to solve this difficult problem is to reason about parts and their arrangements, and that such descriptions can be efficiently used to perform pose estimation once the object is recognized.

This aspect of recognition is therefore complementary to the now well understood problem of pose estimation of rigid objects for which exact models exist in the database (see [Grimson 1990] and references therein). Such approaches are very appropriate when evolving in a controlled environment, where the number of possible objects is small and their geometry is precisely known.

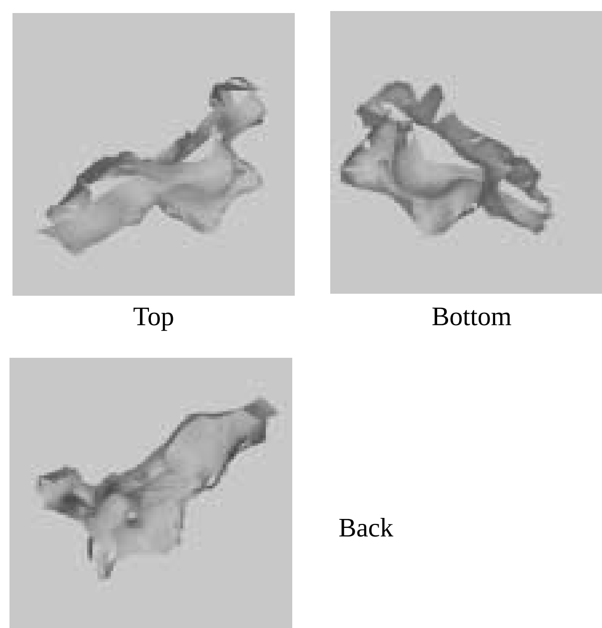
The two sections below describe efforts in implementing a combined learning-recognition system for planar shapes described in terms of 2-D ribbons (GCs in 2-D), and in pose estimation of multi-part objects described in terms of GCs and their arrangement.

3.1 Recognition and Learning for large datasets

Once "good" descriptions (i.e. part-based, volumetric, hierarchical) are extracted from the image, we need to define a recognition engine to manipulate them. Each new object needs to be stored into the database (learning operation) in such a way that it can be retrieved efficiently when a candidate description is proposed for recognition. This is quite different from traditional approaches, in which a static database of models is created off-line, and these models are compared with the ob-



(a) Sample Input images



(b) Three views of the recovered structure

Figure 3. Reconstruction of a 3-D model of an automobile part from a few intensity images

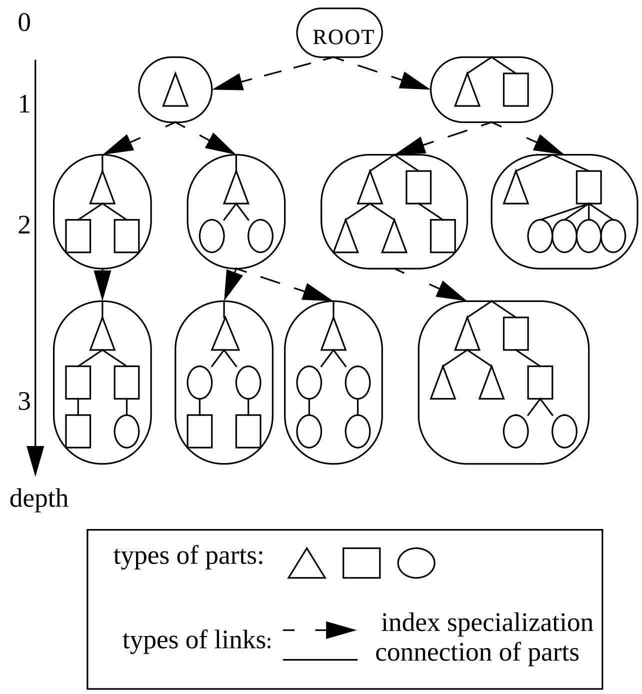


Figure 4. Hierarchical organization of the database for dynamic and efficient access.

served descriptions. We therefore identify two issues:

- the organization of the database (learning and indexing based on high-level representations), and
- the retrieval methodology (matching high-level representations).

One possible solution to these problems is to organize the database hierarchically also, using the structure of the descriptions and the local description of parts. This allows the database to evolve dynamically with minimum cost to update the index. This is schematically illustrated in Figure 4.

The retrieval step needs to proceed from partial matches to take into account possible occlusions and uncertainty. A partial match, based on the connection structure and the aggregation of dissimilarities between parts, is computed incrementally level by level between the new shape and the possible candidates. The combination of the incremental process with the hierarchical indexing effectively decreases the number of shapes processed at each step, therefore dramatically reducing the average complexity of the retrieval. The selected retrieved shape(s) are used to give a classification for the new shape.

We have implemented the core of this recognition engine for 2-D shapes descriptions as produced in [Rom & Médioni 1993]. Experimental validation indeed confirms that retrieval time is nearly independent of the number of objects stored in the database, even with thousands of entries.

It should be noted that the described approach does not obviate the need for geometric reasoning and pose estimation, but rather proposes to invoke them once the object has been recognized. The purpose is then to estimate pose, but also to emphasize differences with existing stored models. This pose estimation task is presented below for multi-part 3-D objects.

3.2 Pose Estimation of Multi-Part Curved Objects

Recognizing 3-D objects from a 2-D image is important for many visual tasks. Part of this problem is the estimation of the 3-D pose of the viewed objects. This task, when dealing with complex, curved, objects, is more difficult because of the possible dependency of the outlines with changes in viewpoint.

Here, we show that alignment-like techniques can be used for a large class of complex, curved, multi-part objects provided adequate features and representations are used. More specifically, we demonstrate that high-level descriptions, based on a part-based formalism using generalized cylinders, provide means to establish *quasi-invariant* correspondences (meaning that they are almost exact over almost all viewpoints) between image and model shapes. These correspondences are in terms of powerful intrinsic quantitative shape attributes such as the axis, the scaling function and the cross-section of a part. The idea is that although the outlines may be viewpoint dependent, or may not have distinguished points, the derived shape descriptions in terms of the above attributes (and their combinations) provide viewpoint independent entities which can be put into correspondence with models so represented. We believe this to be an important demonstration of the usefulness of high-level, part-based, descriptions in extending the classes of shapes which can be handled.

For this, the extracted GC-based image descriptions are first matched to model objects in order to establish object-object and part-part associations. This process is based on graph matching (subgraph isomorphism) using qualitative attributes derived of the image descriptions. Once such associations have been established, a quantitative pose estimation step is invoked for all matched image objects. These results are illustrated in Figure 5. Further details of this work is given in [Zerroug & Nevatia 1996].

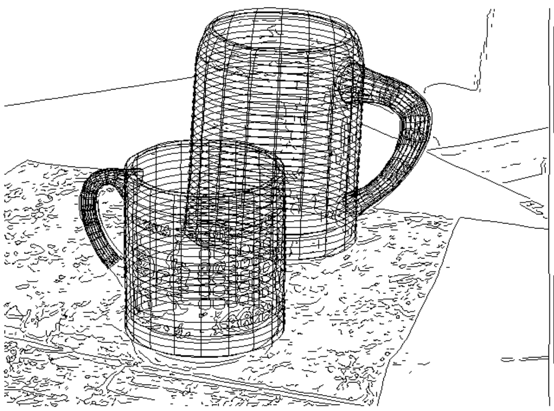
4 Analysis of Range Data

The goals of our effort in range image understanding are to generate rich descriptions from sensed 3-D data. These descriptions should be segmented and capture both the volumetric and surface information related to objects.

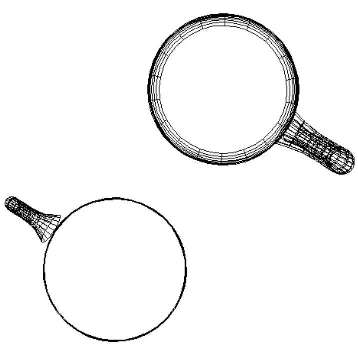
4.1 Inference of Surfaces from Sparse 3-D Data

We address the problem of inferring high-level descriptions (in terms of 3-D junctions, space curves, and sur-

(a) Image



(b) Overlay of objects on edge image.



(c) Top View of the objects

Figure 5. Pose estimation results, matching models to an image.

faces) from an unstructured set of 3-D data, in the presence of noise and outliers. The input may take the form of points, curve elements, or patches, and the output should describe both the surfaces and the surface orientation discontinuities.

This is a difficult problem, as local methods may fail if surfaces are close to each other, and global methods need further assumptions, such as the existence of a single object with a known topology.

A voting process is employed, where each input site casts a vector vote everywhere in a 3-D array. These votes represent preferred orientation and likelihood for a surface passing through a given point in space. The collection of votes at each site of the space is then interpreted in terms of saliency, or a relative likelihood value. Salient surfaces appear as “maximal surfaces.” A subsequent application of the “Marching Cubes” algorithm produces a description in the form of triangular patches.

A similar process is invoked to infer space curves, or lines of intersection between two surfaces, and 3-D junctions. This is illustrated on Figure 6, which shows the object, the sample 3-D data, the extracted surfaces and junctions.

The process is non-iterative, parameter free, and allows for any number of objects in the scene, each with any genus, subject to resolution limitations. This work is described in more detail in [Guy & Medioni 1996b].

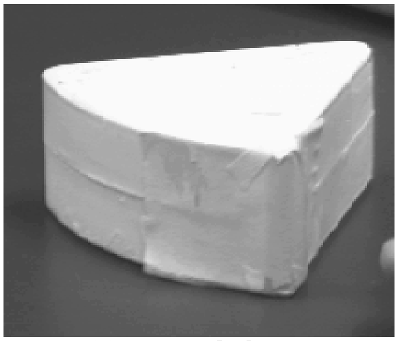
4.2 Surface Approximation and Segmentation of Objects

Deformable models have been widely used to approximate objects from collected data points, mostly for geometrically and topologically simple objects. To handle complex objects (with holes, for example), the topology of the target has to be known beforehand to set up a correct initial guess (curve or surface). This is sometimes inappropriate

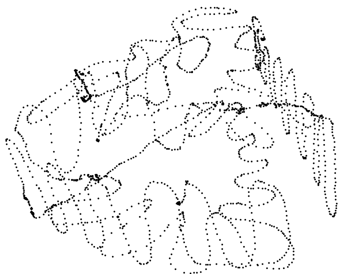
We propose an approach which can fit simultaneously more than one curve or surface to approximate multiple topologically complex objects by using (1) the residual data points, (2) the badly fitting parts of the approximating surface, and (3) appropriate Boolean operations. In 2-D, B-snakes are used to approximate each object (pattern). In 3-D, an analytical surface representation, based on the elements detected, is presented. The global representation of a 3-D object, in terms of elements and their connection, takes the form of B-spline and Bézier surfaces. A Bézier surface is also used to connect different elements, and the connecting surface itself conforms to the data points nearby through energy minimization. This way, a G^1 continuity surface is achieved for the underlying 3D object. We have experimented with both synthetic and real data sets. Shown on Figure 7 is the approximation and decomposition of a teapot. This work is described in detail in [Liao & Medioni 1996].

4.3 Deformable Surface Reconstructions with Discontinuity Edge Detection

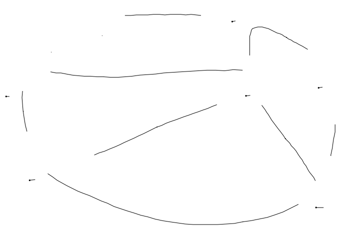
Surface approximation of a dense set of 3-D data points is an ubiquitous problem, which is generally solved by minimizing some form of local error functional, and adaptive subdivision when the fit cannot be improved. This approach is appropriate for smooth surfaces, or surfaces with precomputed sharp edges. This edge detection problem, however, is itself difficult. We propose here to use deformable surfaces embedded with active



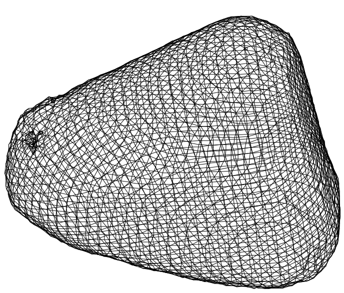
(a) Original object.



(b) Input point set

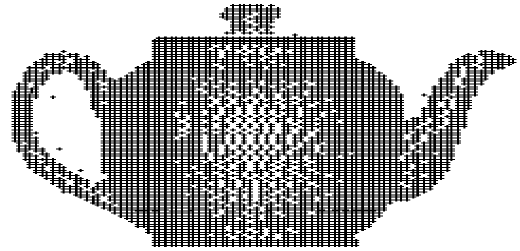


(c) Inferred curves and junctions.

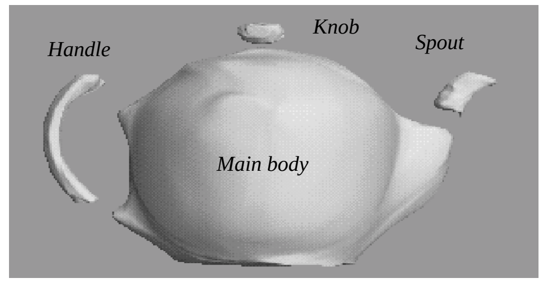


(d) Inferred surfaces.

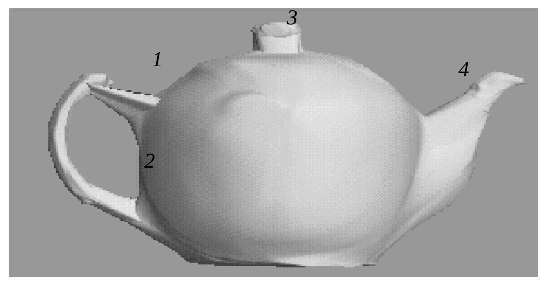
Figure 6. Surface inference from range data.



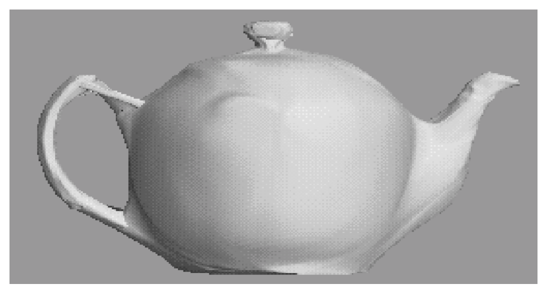
(a) Data points



(b) Result of part segmentation



(c) four Initial connecting surfaces



(d) Final surface fit

Figure 7. Surface description of a Genus 1 object (teapot).

edges, that is to couple the edge detection and surface fitting processes. Our so-called “winged B-snakes” are reconstructed as follows: from an initial, arbitrary triangular tessellation, the triangular B-spline deformable surfaces (wings) flap up and down to fit the range data, and the embedded active edges (snakes) slide around to align with the actual discontinuity edges; then the borders between all pairs of wings are tested to see if they should be continuous or discontinuous; finally the knots are pulled apart for C^1 continuous regions, or allowed to stay together or collinear to respect discontinuity edges and junctions.

We use quadratic triangular B-spline meshes as the surface representation primitives, as they are the state-of-the-art model in CAGD and graphics, and possesses many desirable properties such as arbitrary topology, automatic continuity, completeness, convex hull, fast evaluation, and affine invariance. Since the active edges can slide to align with the actual discontinuity points, we do not have to subdivide the triangles into many smaller ones near discontinuity edges, and thus require much fewer triangles than previously. Also, edge linking and junction detection are only necessary in the segment-level. More details are given in [Han & Medioni 1996]

5 ACKNOWLEDGMENTS

This paper represents the work of many different current and former students, visitors, staff members and associated faculty. M. Zerroug worked on object descriptions from intensity images, and on pose estimation using these descriptions. A. François worked on the generic recognition and learning system, C. Liao worked on surface approximation and segmentation of range images. G. Guy worked on inference of surfaces from sparse and noisy data. P. Havaldar worked on inference of shape from intensity images. M. Lee worked on structure and motion from a sparse set of views. S. Han worked on surface reconstructions coupled with edge detection.

REFERENCES

[Biederman 1987] I. Biederman. “Recognition by Components: A Theory of Human Image Understanding,” *Psychological Review*, 94(2):115-147. 1987.

[Grimson 1990] W. Grimson, *Recognition by Computer-The Role of Geometric Constraints*, Cambridge, MA:MIT Press, 1990.

[Guy & Médioni 1995a] G. Guy and G. Médioni, “Inference of Surfaces, 3D Curves and Junctions from Sparse 3D Data,” in *Proc. IEEE Symposium on Computer Vision*, Miami Beach, Fl, November 1995.

[Guy & Médioni 1995b] G. Guy and G. Médioni, “Inference of Surfaces, 3D Curves and Junctions from Sparse 3D Data,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.

[Guy 1995] G. Guy, “Inference of Multiple Curves and Surfaces from Sparse Data,” Ph. D. Thesis, University of Southern California, 1995.

[Havaldar & Médioni 1996] P. Havaldar and G. Médioni, “Inference of Segmented, Volumetric Shape from Intensity Images,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.

[Han & Médioni 1996] S. Han and G. Médioni, “Deformable Surface Reconstruction Coupled with Discontinuity Edge Detection,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.

[Liao & Médioni 1996] C. Liao and G. Médioni, “Surface Approximation and Segmentation of Objects with Unknown Topology,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.

[Liao & Médioni 1995] C. Liao and G. Médioni, “Surface Approximation of Complex Multipart Objects,” in *Proc. Physics Based Computer Vision Workshop*, Cambridge, MA, June 1995.

[Liao 1995] C. Liao, “Surface Approximation of Complex Objects,” Ph. D. Thesis, University of Southern California, 1995.

[Lee, et al. 1995] M. Lee, G. Médioni and R. Deriche, “Structure and Motion from a Sparse Set of Views,” in *Proc. IEEE Symposium on Computer Vision*, Miami Beach, Fl, November 1995.

[Lee & Médioni 1996] M. Lee and G. Médioni, “Structure and Motion from a Sparse Set of Views,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.

[Nevatia *et al*. 1996] R. Nevatia. “USC RADIUS Related Research: An Overview,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.

[Rom & Médioni 1993] H. Rom and G. Médioni, “Hierarchical decomposition and axial shape description,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, pp. 973-981, October 1993.

[Russ *et al*. 1996] T. Russ, R. MacGregor, K. Price, and R. Nevatia, “VEIL: Combining Semantic Knowledge with Image Understanding,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.

[Zerroug & Nevatia 1994] M. Zerroug and R. Nevatia, “From an Intensity Image to 3-D Segmented Descriptions.” in *Proceedings of the IEEE International Conference on Pattern Recognition*, Jerusalem 1994.

[Zerroug & Nevatia 1996] M. Zerroug and R. Nevatia, “Pose Estimation of Multi-Part Curved Objects,” in *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, February, 1996, these proceedings.