Interactive Confirmation of Object Functionality

Melanie Sutton and Kevin Bowyer
Dept of Computer Science & Engineering
University of South Florida
Tampa, FL 33620-5399
sutton or kwb @csee.usf.edu

Louise Stark
Dept of Electrical & Computer Engineering
University of the Pacific
Stockton, CA 95211
lstark@uop.edu

Abstract

The premise of our most recent work is that a recognition system can and should incorporate both the symbolic labeling of the potential functionality of an object and the steps to confirm said functionality through interaction. Hence, the task at hand is as follows. A researcher selects an object and places it in the observation area of a robot arm. An initial intensity and range image are acquired. This initial state is the input to a two-stage recognition system which first performs the symbolic labeling of the object's potential functionality and produces a plan for interaction for the object. The second stage involves the interaction tests, guided by the plan for interaction, to confirm the object's functional *use* in a task. This paper describes the current state of the system.

Introduction

Our investigation into the use of reasoning about functionality as a means toward general object recognition began in 1987. The initial paradigm for the **GRUFF** (Generic Recognition Using Form and Function) project focused on the essential first step of being able to use functionality to recognize an object by reasoning about its 3-D shape. Initial versions of **GRUFF** aimed to answer the question– given a description of the complete 3-D shape of an object, can a system appropriately categorize the object according to the function(s) that it can serve? A sequence of increasingly more competent versions of **GRUFF** has demonstrated that the answer to this question is yes. Versions of the system have evolved to appropriately categorize a domain of over 400 shapes from the super-ordinate object categories "furniture", "dishes" and "hand tools" (Stark & Bowyer 1991, Stark & Bowyer 1992, Sutton, Stark & Bowyer 1993, Stark & Bowyer 1996).

Once the shape-based reasoning labels the object to indicate areas of functional significance, a *plan for interaction* is created to direct a robot arm to interact with the object and confirm its suggested functionality. The following sections will discuss related work in this area along with our proposed method of experimentation in this direction and some preliminary experimental results.

Related Work

The ideas behind grasping objects and reasoning about function for the purpose of recognition are not new. A number of researchers are actively studying this area. One body of work has investigated performing interaction with objects in a given scene without requiring any model reconstruction or recognition (Ade, Rutishauser & Trobina 1994, Bendiksen and Hager 1994). Other researchers are focusing on using interaction to determine the material properties of an object, again, without explicitly analyzing object shape (Bogoni & Bajcsy 1995, Krotkov 1994). The work which is most closely related to ours includes research where models are recovered and interaction is performed based on hypothesized areas of functional significance (Connell, J.H. 1994, Kim & Nevatia 1994, Rivlin, Rosenfeld & Perlis 1993, Rivlin & Rosenfeld 1995, Stansfield 1992). We propose a method which is differentiated from other work in this area by performing shape-based functional object labeling first, as a means to guide later interaction.

Function-based object recognition

The premise of our most recent work is that a recognition system can and should incorporate both the symbolic labeling of potential functionality of an object and interatction to confirm said functionality. In order to accomplish this task there are three stages in the evaluation:

- model building
- shape-based reasoning - static labeling
- interaction-based reasoning - dynamic confirmation

Model building

In order to create the 3-D model needed for recognition, we rely on a model-building subsystem which builds a 3-D polyhedral boundary representation called an OPUS (Object Plus Unseen Space) model from a range image (Stark & Bowyer 1996). The resulting model includes

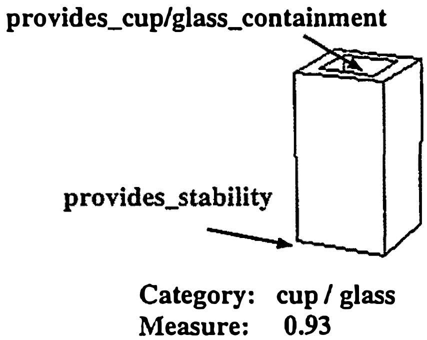


Figure 1: Analysis of Object Shape

the locations of actual object faces and occlusion surfaces. From this information we build a boundary representation model which can be provided to our recognition system for further analysis.

Shape-based reasoning - static labeling

Our system analyzes the 3-D model of the object to determine, based on shape alone, if the object can tentatively function as a member of a particular generic category of objects, such as furniture, dishes, or hand tools. Each class of objects demands specific shape-based functional requirements be met, such as provides_graspability or provides_cup/glass_containment (see Figure 1). This analysis supplies one (or more) "potential" category labels for the objects, with some measure(s) of goodness, and the identification of those portions of the shape with specific uses.

The step by step process leading to shape-based recognition involves the following levels of analysis:

- **Evaluations of the shape** - Each evaluation returns a measure of certainty that some constraint on the shape was met.
- **Integration** - A series of shape evaluations is necessary to determine the compatibility of the shape with a target functional requirement or category.
- **Category decisions** - When faced with a set of alternative interpretations of the shape, categorical decisions are determined by the highest final *association measure* to some category.

In order to use the functional requirements to recognize objects, the functional requirements must be converted into calls to appropriate operators which act on the shape to recover relevant information. There are six operators (termed *knowledge primitives or KPs*) which can be applied to determine functionality:

- **clearance** - of some area within or around the object.
- **dimensions** - lengths, areas of surfaces.
- **enclosure** - locations of the concavities in the object.
- **proximity** - of two surfaces or characteristics of the object.
- **relative orientation** - of two surfaces of the object.
- **stability** - in a given orientation.

A series of knowledge primitives (KPs) are invoked to operate on selected portions of the shape to determine if requirements are met.

Interaction-based reasoning - dynamic confirmation

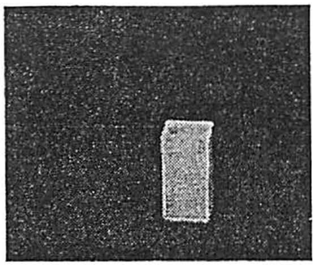
Such analysis would assume that the material properties of the object are sufficient. In reality, this information is difficult to infer reliably from shape alone without actually using the object in some sort of task. The results of the preliminary shape-based reasoning provide us with the symbolic labeling we need to create a plan for interaction, based on the areas of the object with functional significance. Such a plan is developed by the system and commands are generated to direct interaction using a robot arm (see Figure 2).

In order to determine the success or failure of the interaction, new knowledge about how to "apply force" and "observe deformation" is needed. For example, it is possible for a cup-like object composed of paper to pass shape-based analysis. However, during the interaction-based analysis, the limitations of this object should be detected, since the object is not functional. Figure 3 shows the results of interacting with a rigid wooden cup, and a non-rigid paper cup. One way to study the decrease in functionality is by analyzing the deformation of the object boundary regions derived from two sequential intensity images taken during the interaction.

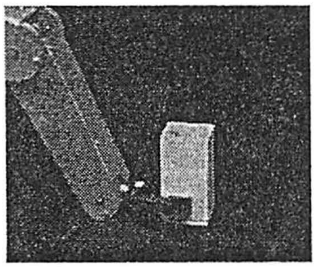
Current Implementation

The experimental paradigm discussed here involves using an integrated vision and robotics system to distinguish between functional and non- functional exemplars, based on either shape alone, or the results of interactions which are suggested by successful shape analysis. We are actively working toward the complete implementation of such a system. With regards to the Model Building subsystem, we currently assume that multiple views have been combined off-line to provide an almost complete 3-D model of the object resting in its expected orientation in the scene. With regards to the Shape-Based Reasoning subsystem, we currently have a system which can analyze 3-D models of objects from the categories furniture, dishes, and hand tools. The resulting analysis provides a symbolic labeling of the object which is used to direct a robot arm to interact with the object. We are currently implementing a set of interaction-based primitive tests to fulfill the dynamic confirmation of the object functionality. These tests include:

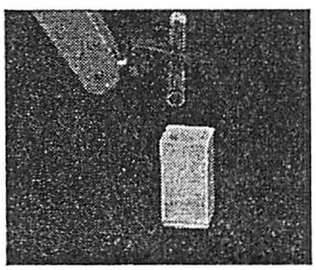
- **change_orientation** - The purpose of this primitive is to re-orient the object so that additional views can



Is the object stable?

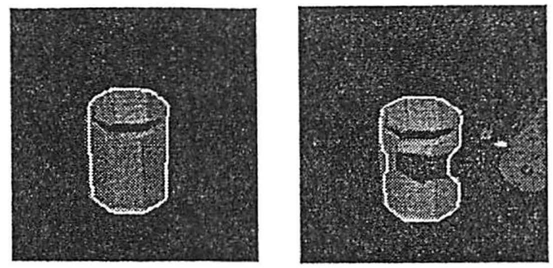


Is the object graspable?



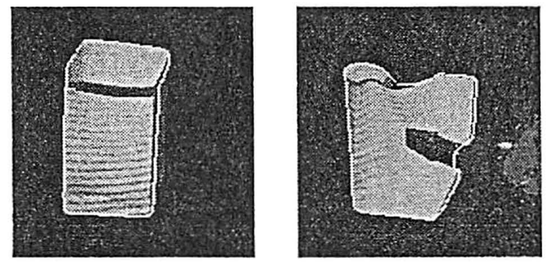
Can the object contain something?

Figure 2: Analysis of Object Interaction



before grasp boundary after grasp boundary

(a) Analysis of a rigid object



before grasp boundary after grasp boundary

(b) Analysis of a deforming object

Figure 3: Analysis of Object Deformation

be taken to build up a more complete 3-D model. where a greater portion of the actual faces of the object and scene are known.

- apply_force - This primitive guides the robot arm during interaction, using the following operations:

  PUSH - push on a given point on the object

  GRASP/UNGRASP - grasp/ungrasp objects at a given point

  TRANSPORT - pull, raise, or lower the object

  POUR - pour a substance into the object at a specified point

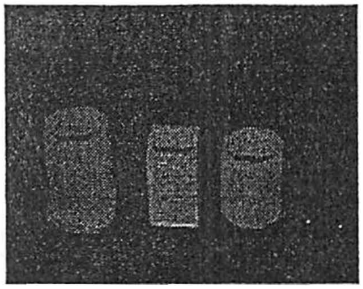
- observe_deformation - Using feedback from both the robot and vision components, the goal of this routine is to determine if the object is deforming during the interaction, in comparison to a reference instance of the shape. The deformation can be analyzed from the information available in the 3-D range image or from comparison of sequential 2-D intensity images taken during the interaction.

Ten object models were created from a variety of materials (Styrofoam. wood, paper, sponge. and foamcore) for the initial run of experiments (see Figure 4). The purpose for varying the material composition of the objects was to test the completeness of our deformation analysis. These objects were also created with a variety of functional capabilities. Our goal was to design objects which could fail at various points in the recognition processing. Analysis of two example objects is depicted in Figure 3.

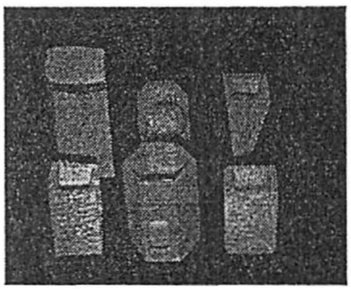
The system we have presented is work in progress. and several of our important components for recovering 3-D information are off-line. As we begin to acquire these models during actual interaction, we hope to be able to use this information to determine when deformation has occurred and additionally to help correct path planning for the robot arm. However, even with the assumption of an almost complete 3-D model prior to interaction. the analysis of 2-D intensity images during the interaction to determine deformation has provided promising results.

Scalability, Complexity

The system we have presented is directly scalable to categories of manufactured objects whose functionality is mostly physical. The corresponding category definition tree grows slowly with the extension into new domains since the additional knowledge is based on functional descriptions for each new *category of objects*,



(a) Functional Objects



(b) Dysfunctional Objects

Figure 4: Test Object Models

rather than on defining new *object instance models*. Additionally, our interaction routines (apply force, observe deformation, etc.) are designed to be generic enough to apply to other categories of objects in the dishes. furniture, and handtools domains. An analysis of the complexity of the shape-based processing is presented in (Sutton, Stark & Bowyer 1994).

Conclusion

This version of GRUFF represents a system operating under what might be termed an *expectancy* paradigm. An initial hypothesis is formed for recognition using bottom-up visual and shaped-based reasoning alone, expecting that material properties of an object are sufficient. However, the results of this preliminary analysis (yielding locations of important functional elements) can provide the guidance to instantiate further (more expensive) top-down exploratory modules to selectively attend to these areas. Just as a diversity of underlying mechanisms support human cognition and learning, we are designing GRUFF as a multi-stage recognition process encompassing overlapping and partially independent modules which operate on different types of information.

Acknowledgments

This research was supported by AFOSR grant F49620-92-J-0223 and NSF grant IRI-9120895. We would like to thank Senthil Kumar and Maha Sallam for providing source code and valuable advice.

References

Ade, F., Rutishauser. M. and Trobina, M. Grasping Unknown Objects, *Dagstuhl-Seminar 9443*, 1994.

Bogoni, L. and Bajcsy, R. Interactive Recognition and Representation of Functionality, *Computer Vision and Image Understanding*, Vol. 62, No. 2, 194-214, 1995.

Connell, J.H. Get Me That Screwdriver! Developing a Sensory-Action Vocabulary for Fetch-and-Carry Tasks, *IBM CyberJournal Research Report RC 19473*, April 1994.

Bendiksen, A. and Hager, G. A Vision-Based Grasping System for Unfamiliar Planar Objects, *IEEE Conference on Robotics and Automation*, 2844-2849, 1994.

Hoover, A., Goldgof, D. and Bowyer, K.W. Extracting a valid boundary representation from a segmented range image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 920-924, September 1995.

Kim, D. and Nevatia. R. A Method for Recognition and Localization of Generic Objects for Indoor Navigation. *IEEE Workshop on Applications of Computer Vision*, 280-288, 1994.

Krotkov, E. Perception of Material Properities by Robotic Probing: Preliminary Investigations, presentation at *The CVPR Workshop: The Role of Functionality in Object Recognition*, 1994

Rivlin, E. and Rosenfeld, A. and Perlis, D. Recognition of Object Functionality in Goal-Directed Robotics, in *Working Notes on Reasoning About Function*, 126-130, 1993.

Rivlin, E. and Rosenfeld, A. Navigational Functionalities, *Computer Vision and Image Understanding*, Vol. 62, No. 2, 232-244, 1991.

Stark, L., and Bowyer, K.W. Achieving generalized object recognition through reasoning about association of function to structure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 3, No. 10, 1097-1104, 1991.

Stark, L., and Bowyer, K. *Generic Object Recognition using Form and Function*, Series in Machine Perception Artificial Intelligence, Vol. 10, World Scientific, New York, 1996.

Sutton, M., Stark, L. and Bowyer, K.W. Function-based generic recognition for multiple object categories, in *Three-dimensional Object Recognition Systems*, A.K. Jain and P.J. Flynn, editors, Elsevier Science Publishers, 447-470 (1993).

Sutton, M.A., Stark, L. and Bowyer, K. GRUFF-3: Generalizing the Domain of a Function-Based Recognition System, *Pattern Recognition*, Vol. 27, No. 12, 1743-1766, December 1994.