REPRESENTING AND REASONING ABOUT OBJECT FUNCTIONALITY: TOWARDS AN INTEGRATED APPROACH

M.A. SUTTON
msutton@dcsuwf.dcsnod.uwf.edu
Department of Computer Science, University of West Florida,
11000 University Parkway, Pensacola, FL 32514-5751, USA

L. STARK
stark@napa.eng.uop.edu
Department of Electrical and Computer Engineering, University of the Pacific,
3601 Pacific Avenue, Stockton, CA 95211, USA

K.W. BOWYER
kwb@csee.usf.edu
Department of Computer Science and Engineering, University of South Florida,
4202 East Fowler Avenue, Tampa, FL 33620-5399, USA

This work summarizes research which uses various aspects of functionality as a means to achieve generic object recognition. We examine several approaches which attempt to integrate computer vision and robotics for the purpose of achieving recognition of functional classes of objects. Such integrated approaches include those which incorporate both the knowledge of the potential functionality of an object and the steps to confirm said functionality through interaction. An overview of our system, **GRUFF-I** (**G**eneric **R**ecognition **U**sing **F**orm, **F**unction and **I**nteraction), is presented. This system reasons about and generates plans for interaction with 3-D shapes from the categories furniture and dishes.

1 Introduction

One area of computer vision research which has been active for a number of years now is the development and use of generic object models. A model which is "generic" is meant to capture a category of objects, as opposed to one particular instance. Part of the inspiration for this approach has come from the field of psychology, where researchers such as Rosch [1,2,3] have done extensive work on how humans develop category concepts. Categorization by humans is attributed to the idea that all possible combinations of attributes are not possible in the physical world. Therefore, humans develop an organization of the world in terms of hierarchical category concepts, grouping objects with similar characteristics. In Rosch's hierarchy, there are three major levels of specificity. For example, a basic (or entry level) category is the name most commonly give to an object, such as "cup." To represent a refinement or specialization of a basic category there are subordinate categories, like "mug" or "juice glass." Finally, a generalization of a basic category would be a superordinate category description, such as "dishes."

Researchers in artificial intelligence were the first to begin incorporating ideas about such object classes into recognition systems. Well-known early work in the area was done by Winston, Binford and co-workers. [4,5] They pointed out that there can be an infinite number of different shape descriptions for objects in a category as simple as *cup*, but that a single functional description can be used to represent all cups in a concise manner. The input to their system consisted of a category definition and a segmented, interpreted description of the object as a semantic network. The system incorporated learning and generalization using examples and analogy.

As work in the area of generic object recognition has continued, a number of researchers have begun investigating the idea of incorporating functional characteristics into the representations of objects, for the purposes of recognition, manipulation and/or navigation. For example, Hodges [6,7] developed the EDISON system which used computational models that could solve problems and reason creatively about mechanical devices. In his Functional Ontology for Naive Mechanics (FONM) representation model, Hodges stressed the importance of structural characteristics in addition to a representation of experience based on function. This is because mechanical improvisation scenarios require the problem solver to recognize in a device some functional capacity which it may not have been used for previously, or was not designed for.

Brand [8] has described a system using causal and functional analysis to interactively build a model of relationships between parts in a scene. Additional work with Cooper, *et al.* [9] focused on understanding image scenes from the point of view of providing a causal explanation of the scene to explain how an agent could interact with it. Three systems (BUSTER, Fido and MugShot) addressed the issues of directing focus of attention, handling occlusion and grasp planning. The MugShot system used a fixed interaction strategy based on a causal scene explanation developed by the system designers, in order to direct a robot arm to lift a mug placed in a workspace. This work stressed the importance of including in the representation the potential for action with the object, which some other agent could perform.

2 Function From X

More recently, a number of methods, which can be termed *function from X* techniques, have been developed for extracting "functional" information from input images, models, or interaction. For example, research in *function from shape* uses the idea that the 2-D or 3-D shape of an object provides some indication of its function. *Function from motion* attempts to recognize the function some object is serving, as the system observes a task being performed with the object. Finally, *function from manipulation/navigation* addresses recovering the function of an object by manipulating or navigating around it.

2.1 Function from Shape

With respect to function from shape, early work by Brooks [10] on the ACRONYM system recognized the relevance in the world of man-made objects of including classes of objects based on function discernible by analysis of geometric structure. Objects were modeled as subpart hierarchies of generalized cones, where variations in size, structure and spatial relationships were permitted for recognition.

In a more generalized approach to modeling functional categories, a system designed by Stark, *et al.* [11] described the 3-D functional requirements for objects to be members from the category chair. These requirements included *provides sittable surface* and *provides stable support*, which could be confirmed by analyzing a complete 3-D model of a given object. This system was later expanded to include functional descriptions for rigid objects in the superordinate categories furniture,[12] dishes [13] and handtools. [14] The most recent version of this system was tested on a set of 400+ 3-D objects. Green, *et al.* [15] later continued work on the system and expanded it to include analysis of the 3-D functional requirements of some non-rigid objects, whose function depended on parts joined by articulated connections. Examples from this hierarchy of objects included pliers and scissors. From a sequence of complete 3-D shape descriptions of an object, an articulated shape model was created which was composed of part descriptions, connections and linkage relations which explained the observed motion sequence. This articulated model was then analyzed for such functional requirements as *provides opposing finger grasp* or *provides opposing cutting blades*. Experimental results for this system were provided for 24 different shapes, with varying degrees of membership for the category scissors.

Rivlin, *et al.* [16] have done extensive work in the area of reasoning about the functionality of an object's parts, recovered from 2-D or 3-D image data. After part recovery, reasoning was performed about the functionality of the parts and interactions, such as relative orientation, size, or motion. Top-down analysis of an intensity image of a functional mallet was presented to demonstrate the use of their approach in expected object recognition, where knowledge of the stored internal model of a hammer was used to constrain image search for potentially functional regions. Bottom-up analysis of an intensity image of a non-functional hammer was presented to demonstrate the use of the system for unexpected object recognition, where the task was to segment the image into regions and recover qualitative shape information and relations to generate possible object hypotheses.

Finally, Kim and Nevatia[17] have used functional descriptions to describe objects when creating a map for a mobile robot. The input to their system was derived from a set of images, and knowledge of the world was based on functional categories. For example, a door was defined as an opening that allows passage through it and a means of closing it, and a desk was defined as an object that a human may work comfortably at and can have objects placed on it. Characteristics in the intensity image, such as size, height, shape and orientation of edge segments were used for recognition.

2.2 Function from Motion

Research in function from motion addresses determining the suitability of some object for satisfying a particular function, and/or recognition of what particular function an object is serving, as it is observed by a system.

Early work in this area by Bogoni, *et al.* [18] investigated manipulatory interactions such as piercing. Three different tools were placed within a Puma-560 gripper, equipped with a force sensor, and were subsequently used to pierce three target objects whose shape and hardness varied (e.g., objects composed of Styrofoam, balsa wood and pine). As the operation was being performed, data was gathered on the visual height of the tool, the force signal orthogonal to the object surface and the position of the end effector. Analysis of this information indicated whether or not the object was capable of performing the operation of piercing.

Duric, *et al.* [19,20] have done work on examining motion sequences of known objects as they were being used. Using information about the object, including its typical uses and the analysis of this motion, this system attempted to recognize functions such as jabbing, chopping, stabbing, scooping, hitting, tightening and hammering. Objects were modeled as shapes composed of primitive parts (i.e. sticks, strips, plates and blobs), and sequences of images of the object were analyzed for primitive motions (i.e. rotation and translation). For example, for the knife category of objects, once the object was located in the image (using the average of all edge points for which the normal flow was computed), motion of its main axes relative to some action surface (termed the *actee*) allowed the recognition of manipulation tasks such as chopping and stabbing.

2.3 Function from Robot Manipulation/Navigation

In the field of robotics, a number of researchers have incorporated vision components and addressed techniques for grasping unfamiliar objects. For example, Hager, *et al.* [21] have worked on a calibration-free system using a Zebra Zero robot arm with a PC controller and cameras to direct robot positioning for picking up arbitrary objects. Hager [22] has also used edge data derived from a single intensity image to consider all possible ways to approach the object and grasp it stably by solving equations based on force, torque and friction. Ade, *et al.* [23] used intensity images derived from three CCD cameras (two lateral silhouettes and an image from above) to analyze collections of dishes and cutlery pieces on cafeteria trays. In later work, [24] they avoided the use of explicit object models and used range data to recover geometric information which served to detect grasping opportunities for a two-fingered gripper in order to perform object removal. In both cases, the goal of these systems was to grasp randomly placed objects.

Realizing that geometric information alone might not be enough to generate an effective interaction or grasping plan for some objects, Krotkov [25] has investigated using vision and robotics to perceive the object's material properties. In this work, a *push and feel* interaction determined the compliance and shear strength of natural terrain (sand, sawdust, or soil) using a six-axis force-torque sensor and joint position sensors attached to a robot leg which exerted compressive and shearing forces on a terrain sample. A *tap and listen* interaction used acoustic sensors (i.e. a microphone, amplifier and filter circuit) to classify materials (wood, concrete, clay, zinc, or ceramic) when they were struck with a cane. A *strike and watch* interaction estimated the mass and the coefficient of sliding friction of a material based on observing its trajectory when struck with a wooden pendulum. More recent work by Krotkov [26] has focused specifically on analysis of acoustic input using a microphone and rods composed of wood, brass, aluminum, glass and plastic (each in two lengths). Each rod was suspended and then struck with a well-damped solid object. The focus of this work has been to determine what acoustic information can be used to diagnose the material and is invariant to object size and shape. Krotkov pointed out that visual cues such as surface luminance can be a key to material properties (e.g., coefficient of friction), but can also be disguised in real-world environments.

Work by Stansfield [27,28] has addressed the types of information a gripper can extract just from exploring an arbitrary object placed in front of it. This information was then used to perform recognition by matching the recovered characteristics to prototypical exemplars. This use of robotics and vision emphasized the modeling of generic object categories, in the context of combined visual/tactile sensing. The concept of a six-sided *spatial polyhedron* was also introduced to represent a class of objects in terms of the spatial configuration of required parts and function-based features that must be able to be sensed along six directions by the end effector. The model provided a means of classifying objects with wide structural variations of a particular feature, provided that the feature was sensed in the spatial location indicated by the model.

More recently, Rivlin, *et al.* [29] discussed the issue of classes of navigational functionalities, where objects were classified as threats or obstacles, prey or food, or landmarks. Categorization was based on characteristics such as object motion, size, or comparison to a reference view. In the context of performing the task of cleaning the corridor floors of a building, their navigating agent, Sisyphus, was designed to recognize independently moving objects as potential threats, large object masses as obstacles, small object masses as litter (prey or food), and specific places (e.g., the corridor walls and ends of the main cleaning area). Each of these recognition tasks corresponded to a different class of object functionalities for the navigating agent, depending on how the class of objects impeded or facilitated its motion in the environment.

3 Experimental Platform for GRUFF-I

We propose a system, **GRUFF-I** (**G**eneric **R**ecognition **U**sing **F**orm, **F**unction and **I**nteraction), which is differentiated from other work in this field by performing shape-based labeling of the functional *areas* of the object to guide later interaction. An overview of this system is provided in Figure 1. The system checks for movement in a workspace and subsequently acquires registered intensity and range images of the scene. The range image is segmented and further analyzed to obtain a 3-D boundary representation of any object in the workspace. This model is provided to a Shape-based Reasoning Subsystem which considers shape-suggested functionality by applying concepts of physics and causation to symbolically label the object's potential functionality (e.g., *provides containment* or *provides sittability*). This labeling is then used to produce a plan for interaction for the object, using a robot arm. Analysis of subsequent 2-D intensity images taken during the interaction determines the object's successful use in a task consistent with its shape-predicted functionality.

3.1 System Setup and Coordination

Our system runs on a SunSPARC Ultra 1, with RS-232 connections to a Microbot Alpha robot arm and the vision system (see Figure 2). The vision system is composed of a K2T Gray-Code Range Finder and Sony CCD Monochrome Video Cam-

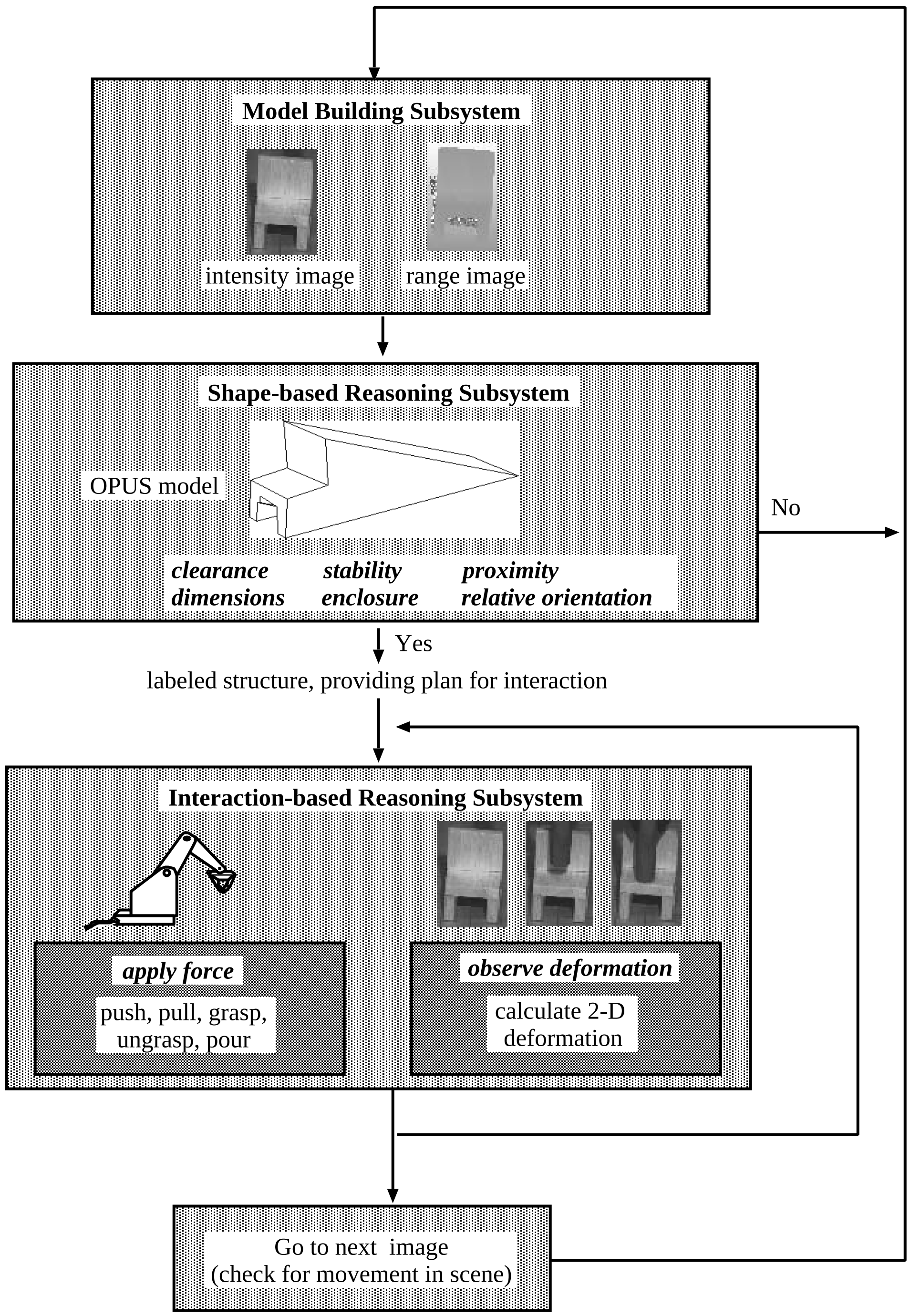


Figure 1: A methodology for recognizing and interacting with generic classes of objects.

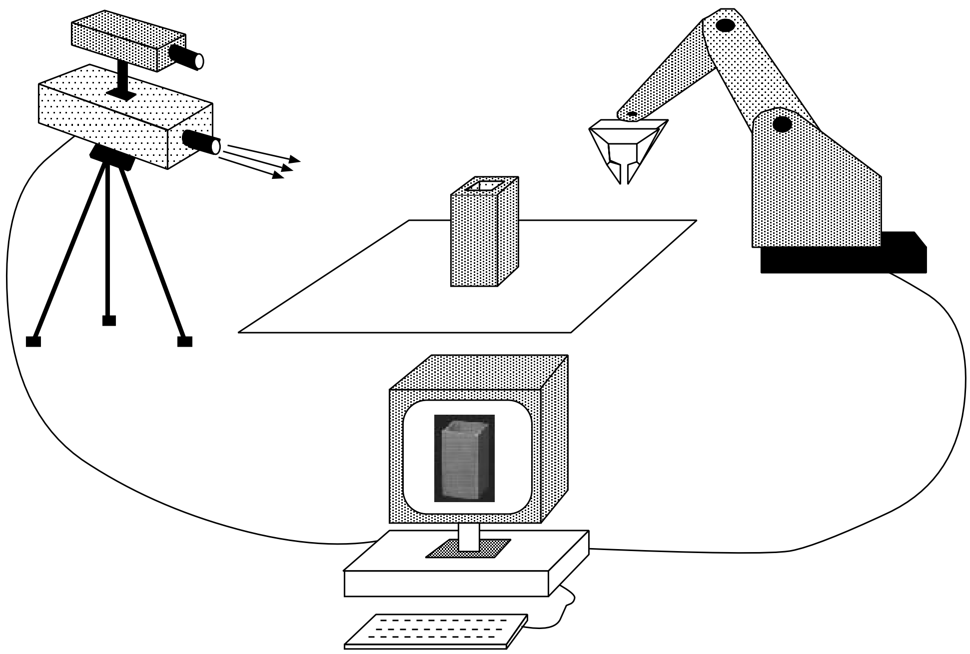


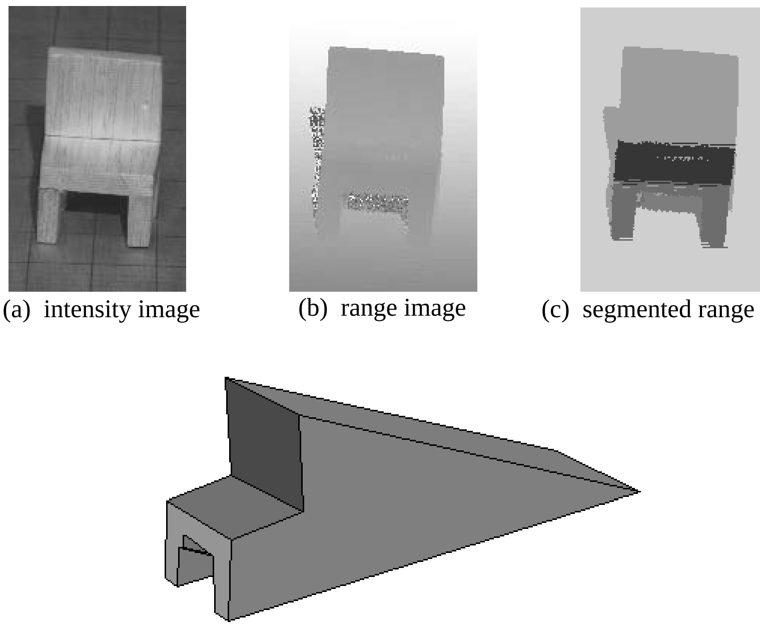
Figure 2: The experimental setup includes a Sun workstation, a Microbot robot arm, a K2T GRF-2 structured light scanner and external circuitry to transmit/receive sensor signals.

era Module with a K2T V300 RGB Digitizer and Display Board. Intensity images are taken as a sequence of light patterns is projected on the scene area, and depth is recovered via triangulation. The scene area consists of a section of a table with a dark surface to diminish the effects of inter-reflectance. The camera is located roughly 1.5 meters from the scene with the light pattern projector mounted just above it. The field-of-view limitations of this camera required test objects to be scaled up to 1/10 of their normal size such that the object fit within a workspace centered by a 0.05 m cube.

3.2 Overview of the **GRUFF-I** *System Components*

Model Acquisition

After movement has been detected in the workspace, the resulting range image is processed using two algorithms developed by Hoover, *et al.*[30] The first algorithm segments the range image into approximate face regions. The second algorithm fits planar surfaces to these regions and connects them together to obtain a 3-D boundary representation model of the scene. An example of the data we acquire from the Model Building Subsystem is shown in Figure 3. The final model acquired from the above steps is what Hoover describes as an Object Plus Unseen Space (OPUS) model. The OPUS model captures the visible real faces and introduces faces along the line of sight referenced as occlusion faces. The resulting OPUS



(a) intensity image (b) range image (c) segmented range image

(d) OPUS model (side view)

Figure 3: The range image in (b) is segmented into approximate face regions (c), from which an OPUS model is created (d).

model of the object is a valid solid model enclosing the object and its surrounding occluded space. Defined in terms of faces and vertices, these models have an average space requirement of about 10 KB each.

Shape-based Reasoning

The Shape-based Reasoning Subsystem analyzes 3-D models of objects from the categories furniture, dishes and handtools to determine if the shape satisfies the functional requirements of some category of objects. This subsystem is written in C, with an executable program size of approximately 2.5 MB. Total shape-based processing time for each object is on the order of seconds to minutes. The resulting analysis provides a categorization of the object and symbolic labeling of object areas with functional significance. There are six operators (termed *knowledge primitives or KPs*) which can be applied to determine functionality:

- **clearance** - This primitive can be used to check that there is a specified volume of unobstructed space in a particular location relative to a particular area of the shape. The volume is represented by a *clearance polyhedron* which is specified by a set of faces and vertices.
- **dimensions** - This primitive can be used to determine, for example, if the width or depth of a surface or group of surfaces lies

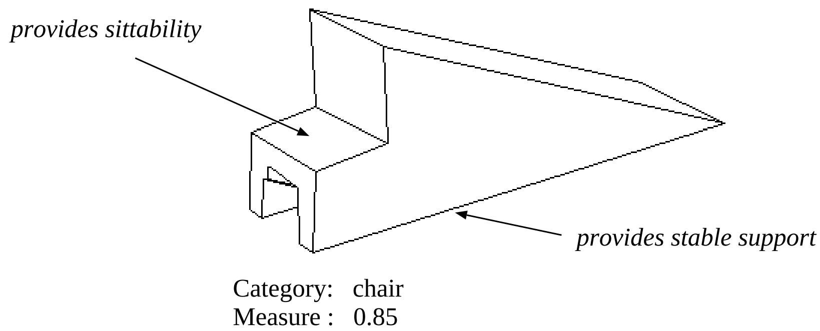


Figure 4: The Shape-based Reasoning Subsystem attaches a category label and indicates the object areas with functional significance.

within a specified range.

- **enclosure** - This primitive is used to determine if there exists a concavity in the shape which can be "closed" by a single plane introduced parallel to the support plane in a given orientation. Because an infinite number of planes could be introduced at different *levels*, each enclosing a different volume, it is desired not only to confirm that a concavity can be enclosed, but also to find the maximal volume that can be enclosed.
- **proximity** - This primitive can be used to check qualitative relations between faces or areas of the object, such as *above*, *below* and *close to*.
- **relative orientation** - This primitive determines if the angle between two normalized vectors falls within a desired range.
- **stability** - This primitive checks that a given shape is stable in the given orientation, with a (possibly zero) force applied to its center of mass. It is assumed that the object has homogeneous density, so that the center of mass may be calculated directly from the shape description.

A series of knowledge primitives is invoked to operate on selected portions of the shape to determine if requirements are met. For example, when analyzing a 3-D shape description to determine if it can satisfy one of the functional requirements for the chair category, *provides sittability*, this subsystem will determine if there exists some surface within the shape with an appropriate size (an invocation of *dimensions*) which is essentially parallel to the ground plane (an invocation of *relative orientation*) and accessible (an invocation of *clearance*) for sitting on (an invocation of *stability*). The results of providing the model in Figure 3-(d) to the Shape-based Reasoning Subsystem would be the categorization and labeling indicated in Figure 4.

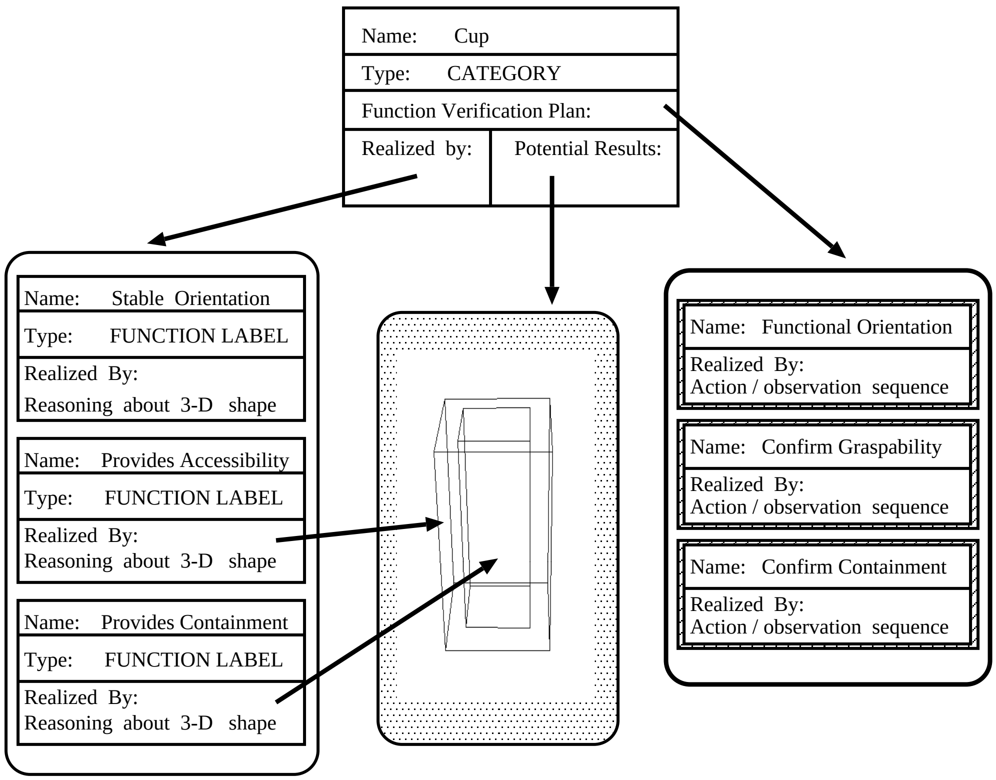


Figure 5: Functional properties that involve only reasoning about abstract shape are shown on the left, as the *hypothesized functional plan*. Functional properties which reason about physical interaction are shown on the right, as the shape-suggested *function verification plan*.

Provided shape-based analysis is successful, a function verification plan is created, which contains a representation of how reasoning about physical interaction should occur. An example of the results from this stage for a cup-like object is shown in Figure 5.

Interaction-based Reasoning

The output of the Shape-based Reasoning Subsystem provides a symbolic labeling of the object. This information can provide a means to direct a robot arm to interact with the object, in a manner consistent with its shape-predicted functionality. In other words, rather than testing all possible ways of sitting on something which looks like a chair, the Interaction-based Reasoning Subsystem can test the specific orientation in which this object is assumed to *function as* a chair. During interaction, the verification plan template, which stores a representation of the shape-based potential result and pointers for interaction, is instantiated. This plan involves invoking the robot arm to interact with the object and taking successive intensity images to monitor the interaction, as shown in Figure 6.

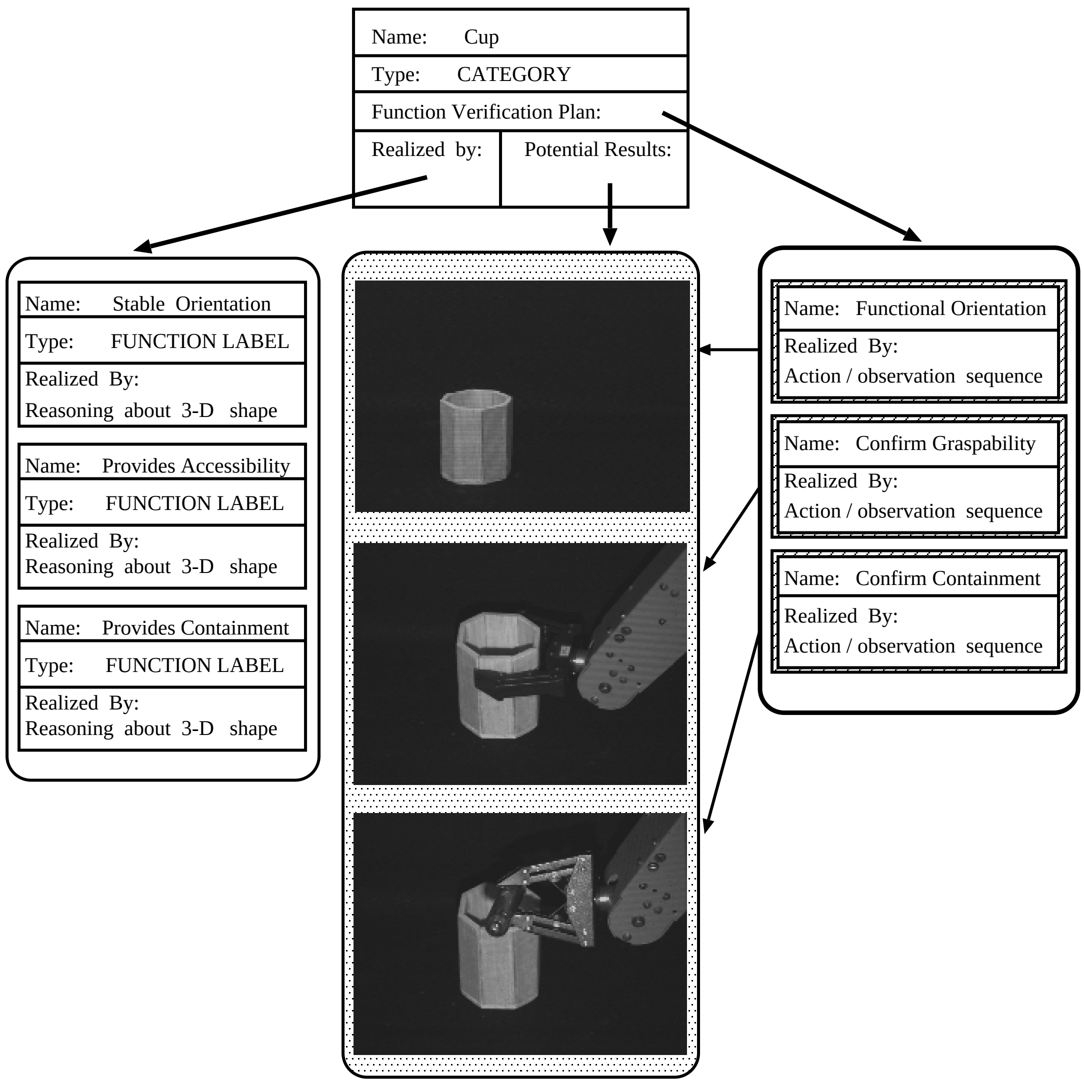


Figure 6: Verification of the function plan involves instructing the robot arm to interact with the object in a manner consistent with its shape-suggested functionality.

In order to determine the success or failure of the interaction, new knowledge about how to "apply force" and "observe deformation" is needed. The creation of an interaction plan is accomplished by establishing a set of interaction-based primitive tests, as follows:

- **apply force** - This primitive guides the robot arm during the interaction, using the following operations:
  - GRASP/UNGRASP - grasp/ungrasp objects at a given point
  - TRANSPORT - push, 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 primitive is to determine if the object is deforming during the interaction, in comparison to a reference instance of the shape, using consecutive 2-D intensity images taken during the interaction.

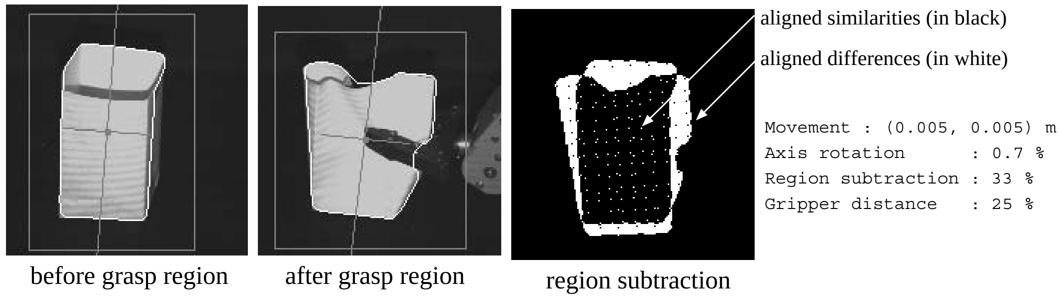
Just as the shape-based functional requirements (such as *provides sittability*) were tested by invoking the shaped-based primitives, interaction-based functional requirements are confirmed for a given object using sequences of invocations of these two new primitives. For example, the requirement *confirm sittability* involves calls to *apply force* to position a weight on the potential seat and subsequent calls to *observe deformation* to determine if the interaction is successful.

In each intensity image taken during the interaction, the object region is segmented from the background based on thresholding. Calculating the centroid of each region provides the approximate object location. The distance between centroids in successive images is monitored to detect when the object falls outside a given range. Additional image processing occurs depending on the current category. When testing containment for cups, for example, we need to analyze both images and compare changes in the size, number and angle of the principle axes of the object regions, in addition to the locations of pouring substance regions. Feedback from the gripper also indicates when an object is clenched beyond its predicted width, when confirming graspability. An example of this type of information is shown in Figure 7.

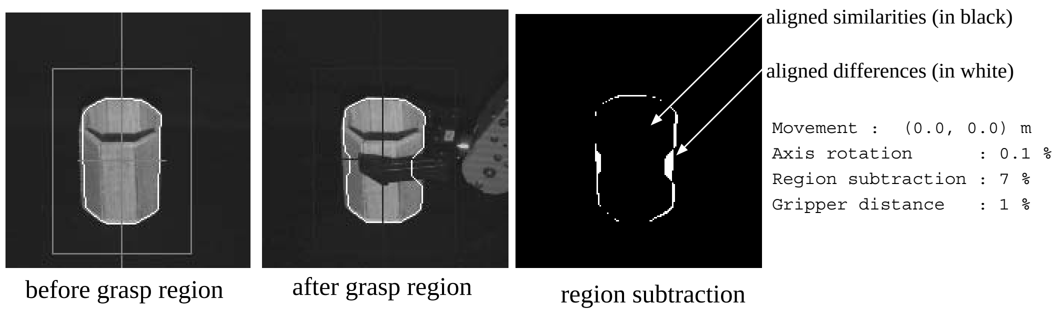
The goal of the Interaction-based Reasoning Subsystem is therefore to determine the *suitability* of the material property requirements of the input shape, as opposed to the *identification* of its particular material composition. For example, based on interaction, this subsystem would be able to determine that a paper mache chair is not functional due to its tendency to deform when weight is applied, as opposed to determining this based on the fact that the object is composed of paper. If interaction is successful, this subsystem displays a measure of goodness, then waits for movement, or the next object to be placed in the scene area.

4 Experimental Results

Twenty-one object models were created from a variety of materials (Styrofoam, wood, paper and sponge) for the initial run of experiments (see Figure 8). The pur-



(a) Results of deformation analysis for a deforming object

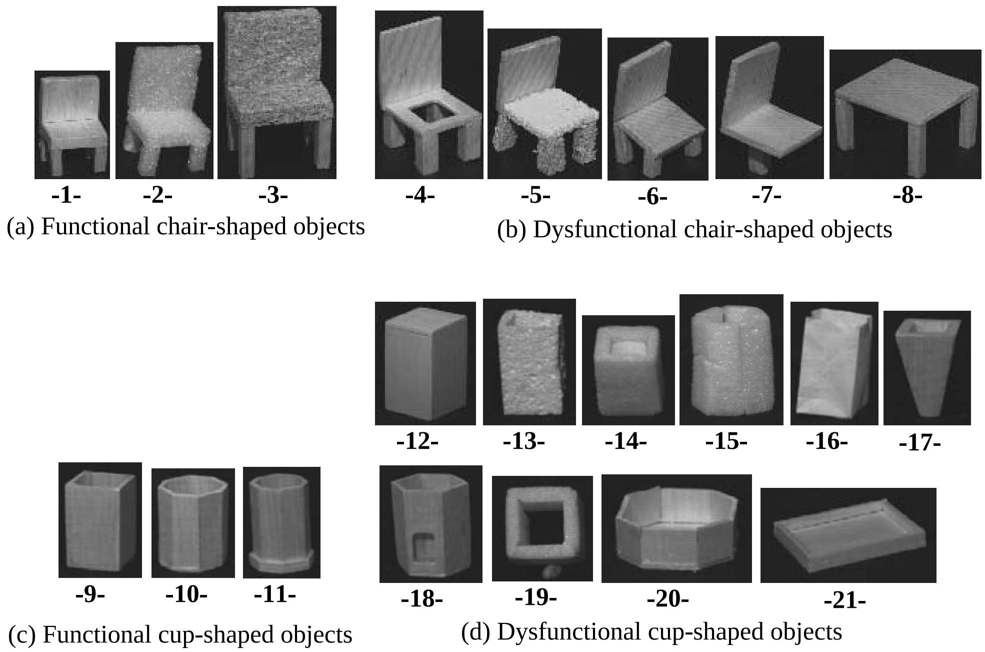


(b) Results of deformation analysis for a rigid object

Figure 7: Once the object regions and principle axes are determined for each image, the *after grasp* region can be aligned to the *before grasp* region for image subtraction and comparison. Additionally, the expected and actual gripper distance before and after grasping the object can be compared.

pose of varying the material composition of the objects was to test the completeness of our deformation analysis and to have objects with a variety of functional capabilities. Two sets of experiments were run, where each object was placed in front of the scene two times. The first set of experiments examined the effect of interaction with furniture-like objects, using all three subsystems (Model Building, Shape-based Reasoning and Interaction-based Reasoning). The second set of experiments examined the effect of interaction with dish-like objects, using the final two subsystems (Shape-based Reasoning and Interaction-based Reasoning). This latter set of experiments assumed a complete 3-D model as input for the shape description, defined in terms of faces and vertices in the robot arm-centered coordinate system. This assumption will be relaxed in future versions of the **GRUFF-I** system, as problems with the added complexity of surface descriptions of concavities within the 3-D shape are worked out in the Model Building Subsystem.

The results of these experiments are summarized in Table 1. Two unexpected results occurred in the first set of objects (for the category chair). In this case, the Model Building Subsystem failed to to create a valid 3-D model due to extensive over/under segmentation and/or shadow areas in the range images for Objects 3



(a) Functional chair-shaped objects (b) Dysfunctional chair-shaped objects

(c) Functional cup-shaped objects (d) Dysfunctional cup-shaped objects

Figure 8: Functional and dysfunctional object models were created for each category.

and 4. Shadow areas are an inherent problem when using a structured light scanner, which contains two optical paths. Whenever some surface of the object obstructs the view of either sensor (CCD or light projector), the resulting range pixels are invalid and are often not handled properly by the segmentation and model building algorithms.

Chair-like objects failed shape-based reasoning in this category if they were determined to be unable to meet all the functional requirements of the chair category. For example, Object 6 did not have a suitable, essentially parallel sittable surface, and the size of Object 8 fell outside the dimensional constraints of chairs. Objects 5 and 7 failed interaction-based reasoning for chairs due to the detection of deformation during the interaction.

For the cup category of objects, Object 12 failed shape-based reasoning due to a lack of a concavity, while Objects 20 and 21 failed due to shape-based dimensional constraints. Objects 13 and 16 deformed significantly during grasping due to material composition (sponge and paper). Object 15 fell apart during grasping, while Object 17 was determined to be unstable. Objects 14, 18 and 19 were determined to have inadequate concavities to function as cups.

5 Scalability and the Future of the GRUFF-I System

Although the objects shown in the previous section are polyhedral, the Model Building Subsystem can be used to recover polyhedral approximations of curved surfaces,

Table 1: Summary of Experimental Results

Object Number	Recognition: Chair Shape / Interaction	Recognition: Cup Shape / Interaction	Comment
1	PASS / PASS	–	object is a chair
2	PASS / PASS	–	object is a chair
3	SYSTEM ERROR	–	Model Building Error
4	SYSTEM ERROR	–	Model Building Error
5	PASS / FAIL	–	object is non-rigid
6	FAIL / –	–	no sittable surface
7	PASS / FAIL	–	object is unstable
8	FAIL / –	–	object is a non-chair
9	–	PASS / PASS	object is a cup
10	–	PASS / PASS	object is a cup
11	–	PASS / PASS	object is a cup
12	–	FAIL / –	no concavity found
13	–	PASS / FAIL	object is non-rigid
14	–	PASS / FAIL	insufficient concavity
15	–	PASS / FAIL	object falls apart
16	–	PASS / FAIL	object is non-rigid
17	–	PASS / FAIL	object is unstable
18	–	PASS / FAIL	insufficient concavity
19	–	PASS / FAIL	insufficient concavity
20	–	FAIL / –	object is a non-cup
21	–	FAIL / –	object is a non-cup

by fitting a set of planar surfaces.[31] For the work described in this paper, the use of polyhedral models was driven by the constraints of path planning and determining grasping possibilities for a two-fingered gripper. As the scene area is increased in future systems, we expect to incorporate a set of actual manufactured objects (in addition to the set of scaled custom-made models), in order to more thoroughly investigate the handling of real-world data.

In addition, although the Model Building Subsystem requires 'dense' range data to build a 3-D model for the Shape-based Reasoning Subsystem, the Interaction-based Reasoning Subsystem is based entirely on analysis of sequences of 2-D images. This constraint was imposed due to the time required to segment the range image and build a model, which is on the order of minutes. One possible extension would be to develop a set of 2-D shape-based knowledge primitives, which could also act on the set of intensity images during limited time constraint situations.

Other extensions include incorporating additional sensors into the system and integrating the information between sensors, such as using vision to aid path planning. An additional robot arm would allow interaction with articulated objects. Force and weight sensors would be useful to provide additional feedback when testing requirements such as *provides sittability* and *confirm containment*. In addition, the use of an alternative 3-D data acquisition system could overcome the object material/appearance restrictions of the structured light scanner components.

A more extensive project would involve integrating multiple views of an object in the scene to explore the shape as much as possible, prior to any interaction. This would allow the system to postulate the view or views which contain the "maximal function information" for model building. This leads to the the need for the development of another interaction-based primitive which could be used to re-orient the object and/or the porting of the system to a mobile platform.

Finally, in the current implementation, isolated objects are assumed. However, more interesting applications are possible with scenes of objects. Toward this end, Hoover[31] has designed a system for acquiring what he calls a *space envelope representation* of the world. This offers an exciting opportunity to consider implications of the "super-functionality," or context, of a collection of objects, and suggests the possibility of performing object segmentation based on functional properties. Such "functionality in the large" applications also offer the opportunity to explore higher level processing mechanisms for scene analysis. For example, from the functional processing of a local neighborhood of objects, the system may determine the location is a "kitchen." This information could then be used to further index into or more appropriately rank subsequent processing mechanisms.

6 Conclusion

This paper has dealt with summarizing current approaches to generic object recognition and the integration of function-based reasoning at the representation and confirmation levels. By using classes of objects, rather than particular instances, the quantity of distinct models is greatly reduced for creating a general purpose vision system. One of the strongest drawbacks of previous approaches in the field of object recognition has been the reliance on vision alone to achieve recognition. To develop more successful systems, there are a number of additional sensing modalities a complete system should incorporate.

The **Gruff-I** system is predicated on the idea that due to the volume of data an autonomous agent may encounter in any real-world scene, judicious use of resources is important. Shape-based reasoning prior to manipulation provides the ability to suggest possible areas of functional significance, where further robotic manipulation can be used. Since robotic control can be expensive in terms positioning, re-calibration and time, pre-recognition of the shape and the design of an interaction plan to confirm those areas with functional significance provides an efficient methodology. In other words, we expect to visually explore the object as much as possible, to develop a *plan of minimal work* for interaction. The cost of such shape-based analysis is directly proportional to the complexity of the object. This analysis is therefore not computationally cheap in an absolute sense, but appears significantly cheaper relative to the costs involved with interacting independent of object shape. Further information on this project can be accessed via http://marathon.csee.usf.edu/.

Acknowledgments

Adam Hoover created the range image segmenter and OPUS model builder. Maha Sallam provided image processing source code and valuable advice. This work was supported in part by AFOSR grant F49620-92-J-0223 and NSF grant IRI-9120895. Melanie A. Sutton was supported by a NASA Space Grant Consortium Fellowship.

References

1. E. Rosch, C.B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.
2. E. Rosch. Human categorization. In Neil Warren, editor, *Studies in Cross-cultural Psychology*, pages 1–49. Academic Press, 1977.
3. S. Palmer, E. Rosch, and P. Chase. Canonical perspective and the perception of objects. In J. Long and A. Baddeley, editors, *Attention and Performance IX*, pages 135–151. Lawrence Erlbaum Associates, Publishers, 1981.
4. T.O. Binford. Survey of model-based image analysis systems. *International Journal of Robotics Research*, 1:18–64, 1982.
5. P. Winston, T. Binford, B. Katz, and M. Lowry. Learning physical descriptions from functional definitions, examples, and precedents. *AAAI*, pages 433–439, 1983.
6. J. Hodges. Naive mechanics–a computational model of device use and function in design improvisation. *IEEE Expert*, pages 14–27, February 1992.
7. J. Hodges. Functional and physical object characteristics and object recognition in improvisation. *Computer Vision and Image Understanding*, 62(2):147–163, September 1995.
8. M. Brand. Vision systems that see in terms of function. In *Working Notes on Reasoning About Function*, pages 17–22, 1993.
9. P.R. Cooper, L.A. Birnbaum, and M.E. Brand. Causal scene understanding. *Computer Vision and Image Understanding*, 62(2):215–231, September 1995.
10. R.A. Brooks. Model-based computer vision, UMI Research Press, Ann Arbor. *MI*, 85, 1984.
11. L. Stark and K.W. Bowyer. Generic recognition through qualitative reasoning about 3-D shape and object function. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 251–256, 1991.
12. L. Stark and K.W. Bowyer. Function-based recognition for multiple object categories. *CVGIP: Image Understanding*, 59(1):1–21, January 1994.
13. M. Sutton, L. Stark, and K.W. Bowyer. GRUFF-3: Generalizing the domain of function-based recognition system. *Pattern Recognition*, 27(12):1743–1766, December 1994.
14. L. Stark and K.W. Bowyer. *Generic Object Recognition Using Form and Function.* World Scientific Publishing, New Jersey, 1996.
15. K. Green, D. Eggert, L. Stark, and K. Bowyer. Generic recognition of articulated objects through reasoning about potential function. *Computer Vision and Image Understanding*, 62(2):177–193, September 1995.

16. E. Rivlin, S.J. Dickinson, and A. Rosenfeld. Recognition by functional parts. *Computer Vision and Image Understanding*, 62(2):164–176, September 1995.
17. D. Kim and R. Nevatia. A method for recognition and localization of generic objects for indoor navigation. In *IEEE Workshop on Applications of Computer Vision*, pages 280–288, 1994.
18. L. Bogoni and R. Bajcsy. Interactive recognition and representation of functionality. *Computer Vision and Image Understanding*, 62(2):194–214, September 1995.
19. Z. Duric, J.A. Fayman, and E. Rivlin. Function from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):579–591, June 1996.
20. Z. Duric, J. Fayman, and E. Rivlin. Recognizing functionality. In *Proceedings of International Symposium on Computer Vision*, pages 247–252, 1995.
21. G.D. Hager, W. Chang, and A.S. Morse. Robot feedback control based on stereo vision: Towards calibration-free hand-eye coordination. In *Proceedings of IEEE Conference on Robotics and Automation*, pages 2850–2856, 1994.
22. A. Bendiksen and G. Hager. A vision-based grasping system for unfamiliar planar objects. In *Proceedings of IEEE Conference on Robotics and Automation*, pages 2844–2849, 1994.
23. F. Ade, M. Rutishauser, M. Trobina, and A. Yla-Jaaski. A 3-D vision system for deriving gripping information for a robot. In *Proceedings of 8th Scandinavian Conference on Image Analysis*, pages 329–336, 1993.
24. F. Ade, M. Rutishauser, and M. Trobina. Grasping unknown objects. In H. Bunke, editor, *Proceedings of Dagstuhl Seminar: Environment Modeling and Motion Planning for Autonomous Robots*, pages 445–459. World Scientific, 1995.
25. E. Krotkov. Perception of material properties by robotic probing: Preliminary investigations. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 88–94, August 1995.
26. E. Krotkov, R. Klatzky, and N. Zumel. Analysis and synthesis of the sounds of impact based on shape-invariant properties of materials. In *Proceedings of the International Conference on Pattern Recognition*, pages 115–119, August 1996.
27. S. A. Stansfield. Haptic perception with an articulated, sensate robot hand. *Robotica*, 10, 1992.
28. S.A. Stansfield. Representing generic objects for exploration and recognition. *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1090–1096, 1988.
29. E. Rivlin and A. Rosenfeld. Navigational functionalities. *Computer Vision and Image Understanding*, 62(2):232–244, September 1995.
30. A. Hoover, D. Goldgof, and K.W. Bowyer. Extracting a valid boundary representation from a segmented range image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(9):920–924, September 1995.
31. A. Hoover. *The Space Envelope Representation for 3D Scenes.* Ph.D. Dissertation, University of South Florida, Tampa, 1996.