

Function from visual analysis and physical interaction: a methodology for recognition of generic classes of objects

Melanie Sutton^{a,*}, Louise Stark^b, Kevin Bowyer^c

^a*Department of Computer Science, University of West Florida, Pensacola, FL 32514, USA*

^b*Department of Electrical and Computer Engineering, University of the Pacific, Stockton, CA 95211, USA*

^c*Department of Computer Science and Engineering, University of South Florida, Tampa, FL 33620, USA*

Received 18 October 1996; revised 17 September 1997; accepted 5 November 1997

Abstract

This paper presents an overview of the GRUFF-I (Generic Recognition Using Form, Function and Interaction) system, a nonpart-based approach to generic object recognition which reasons about and generates plans for interaction with three-dimensional (3D) shapes from the categories furniture and dishes. The system operates as follows. A researcher selects an object and places it in an observation area. An initial intensity and range image are acquired. These are input to a three-stage recognition system. The first stage builds a 3D model. The second stage receives as input a 3D model and considers the shape-suggested functionality of this shape by applying concepts of physics and causation (e.g. to infer stability) to label the object's potential functionality. The third stage uses this labeling to instantiate a plan for interaction to confirm the object's functional use in a task by incorporating feedback from both visual and robotic sensors. Results of this work are presented for eighteen chair-like and cup-like objects. Major conclusions from this work include: (1) metrically accurate representations of the world can be built and used for higher level reasoning; (2) shape-based reasoning prior to interaction-based reasoning provides an efficient methodology for object recognition, in terms of the judicious use of system resources; and (3) interaction-based reasoning can be used to confirm the functionality of a categorized object without explicitly determining the object's material composition. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Visual analysis; Physical interaction; Generic recognition; GRUFF-I; Computer vision; Robotics

1. Introduction

Research in computer vision has addressed a problem humans seem to solve quite easily hundreds of times each day; that is, recognition of objects in the world around us. Initial work in this area concentrated on the use of explicit geometric models for object recognition. However, as the need for more general purpose vision systems has grown, there has been a shift toward the development and use of generic object models. Such models are meant to capture a category of objects, as opposed to one particular instance. In addition, as researchers in the fields of psychology, artificial intelligence and computer vision have gained a greater understanding of the complexity of object recognition, the idea of using functional characteristics of an object to distinguish classes of objects to aid recognition has further developed. In brief, functional analysis of an object relies on analyzing the object to determine if it can satisfy the

requirements of some object category. For example, one could describe members of the category 'cup' in terms of functional characteristics, such as provides graspability or provides containment. As a result, function-based reasoning as a methodology for object recognition can be viewed as an approach applicable to environments in which the objects which need to be recognized were designed or are used with specific purposes in mind.

This paper describes the implementation of a function-based object recognition system which operates to recognize items by integrating feedback from both vision and robotics. The system is called GRUFF-I (Generic Recognition Using Form, Function and Interaction). The input to the system is a set of range and intensity images, and the output is an analysis of whether the object can function as a member of a particular generic category of objects such as furniture or dishes. Each class of objects demands that specific shape-based functional requirements be met, such as provides graspability or provides cup/glass containment. Once the reasoning about the three-dimensional (3D) shape

* Corresponding author. E-mail: msutton@argo.cs.uwf.edu

of the object indicates 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 layout of the remainder of the paper is as follows. Section 2 explains related research in function-based object recognition. Section 3 details the motivation for work in this area. Section 4 provides an overview of the experimental platform and equipment upon which the GRUFF-I system is based. Sections 5 and 6 summarize the evaluation techniques and experimental results of the implemented system. Finally, Section 7 provides a conclusion and directions for future research.

2. Background

In psychology, there has been ongoing research to determine how humans form category concepts. This work has been based on the assumption 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 [1–3].

In the field of artificial intelligence, a number of researchers have investigated how systems might use functional characteristics to discriminate among a set of objects. These systems often begin with a symbolically labelled object and attempt to classify it. For example, Dey et al. incorporated ideas about recognition and learning of generic classes of objects in designing a system to create a knowledge base for the handtools domain [4]. In the system designed by Winston et al., functional descriptions were used to represent objects from the cup category [5]. Minsky has argued for a stronger connection between structure and function, using the category chair as an example [6].

As work in the area of generic object recognition has continued, a number of researchers have been investigating the idea of incorporating functional characteristics into the representations of objects, for the purposes of recognition, manipulation and/or navigation. For example, Hodges developed the EDISON system which uses computational models that can solve problems and reason creatively about mechanical devices [7,8]. Brand has described a system using causal and functional analysis to interactively build a model of relationships between parts in a scene [9]. Additional work with Cooper and Birnbaum focused on understanding image scenes from the point of view of providing a causal explanation of the scene to explain how an agent can interact with it [10].

A number of additional 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 2D or 3D 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.

With respect to function from shape, early work by Brooks on the ACRONYM system recognized the relevance of modeling classes of objects based on function discernible by analysis of geometric structure [11,12]. Brady et al. discussed the relation between geometric structure and functional significance in their design of the ‘Mechanic’s Mate’ system [13,14]. Kitahashi et al. brought the ‘functant’ into their representation to help define prototypical shapes [15]. In early work by Davis, analysis of hard rigid objects such as boxes, cages, buttons and rings suggested ideas about how function related to shape, and could be used to design geometric heuristics for analyzing the use of the object in a situation [16]. Within the context of device function, Goel has investigated a system which analyzes the design of physical devices such as electrical circuits and heat exchangers [17].

In a more general approach to modeling functional categories, a system designed by Stark et al. described the 3D functional requirements for objects to be members of the categories chair [18] and furniture [19]. This system was later expanded to include dishes [20] and handtools [21]. Rivlin and Rosenfeld have done extensive work in the area of reasoning about the functionality of an object’s parts, recovered from 2D or 3D image data [22]. Finally, Kim and Nevatia used functional descriptions to describe objects when creating a map for a mobile robot [23].

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 and Bajcsy investigated manipulatory interactions, such as piercing [24,25]. Duric et al. examined motion sequences of known objects as they were being used [26,27]. Kise et al. proposed a functional model that was capable of encoding dynamic functions such as force and motion common to objects in the category handtools [28]. Green et al. continued work on the system designed by Stark et al. and expanded it to include analysis of the 3D functional requirements of some nonrigid objects, whose function depended on parts joined by articulated connections [29,30].

Function from manipulation/navigation addresses recovering the function of an object by manipulating or navigating around it. Somewhat related to this area, Connell designed a system which considered the domain of manipulating brightly coloured, cylindrically shaped handtools [31]. Objects were modeled as 2D blob-like pixel regions in intensity images with expected characteristics such as size, position and orientation and signature-like features, such as texture, colour, gross shape and particular subpart configuration.

Work by Stansfield addressed the types of information a

gripper could extract just from exploring an arbitrary object placed in front of it [32,33]. This information was then used to perform recognition by matching the recovered characteristics to prototypical exemplars. Stark has explored using function-based reasoning with the GRUFF-3 system and analyzed simulations of object interaction using ‘Thing World’, a physically-based solid modeling system [34]. These simulations were carried out for the category chair and emphasized the use of shape-based reasoning to generate a plan for interaction. Finally, Rivlin et al. [35] 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.

3. Motivation

From this literature review, it should be evident that the ideas behind reasoning about function for the purpose of recognition are not new. In addition, there are many applications where there is still a clear need for autonomous agents to be able to interact with the environment. The work described in this paper is similar to Connell’s, in the sense that we also use versatile operators in the vision and robot components [31]. However, Connell’s system does not explicitly look for functionally relevant information but, rather, for shape-based features such as handles of a certain dimension, in pairs, etc. In addition, Connell maintains that 2D analysis is advantageous since then only a minimal ‘model’ description is needed for operation. Further, he believes it is difficult to acquire a geometric model of the target object and for a vision system to build metrically accurate representations of its surroundings. Alternatively, this paper demonstrates that accurate 3D representations of the surroundings are possible and more readily applicable to a greater variety of more complex domains of objects, compared to a 2D blob analysis approach to recognition.

One similarity we share with Stansfield’s work is the use of dimensional constraints to distinguish categories of objects [32,33]. However, the work described in this paper differs in a number of fundamental ways. First, Stansfield believes that each category of objects can be modeled by an *n*-sided prototype (the most representative shape in the category), and that a set of shape deformations can be specified a priori of that prototype to account for all variations inside that category. With this methodology, functionally similar but structurally different objects are organized as members of separate categories of objects. Therefore, each object in a given category is assumed to have the same set of parts. This means all ‘cups’ must have one handle. ‘Two handled cups’ would call for the formation of a new category, as would ‘no handled cups’.

A drawback of such part-based approaches is that they

require the definition of a finite and inclusive set of primitives or sufficient vocabulary of parts into which all objects can be decomposed. Alternatively, GRUFF-I is not part-based. Rather, shape-based analysis is used, allowing us to identify functionally significant areas, regardless of size or shape. These areas are determined based on applying general inference rules about shape, as opposed to rules for partitioning based on prespecified shape properties. Therefore, whereas the *n*-sided prototype representation could permit fewer faces for simpler objects, or may need more faces for manipulating complex objects, GRUFF-I has no such constraints. Once a functionally significant area is found, regardless of its location, the robot can be sent to approach that specific area, along any orientation, constrained only by manipulator positioning limits.

Finally, in Stansfield’s approach, exploration of the object is not model driven, and no a priori knowledge about the object is desired or assumed by the system. The GRUFF-I approach differs from this in that the output of the vision component is used to recognize the object, prior to any interaction. In this sense, the GRUFF-I approach resembles active vision systems which seek to interpret the current state of the scene and react in a meaningful manner. In other words, the system seeks to understand not only what and where the objects are, but how to interact with them efficiently. In this way, less energy is expended whenever objects are outside the domain of the system, or truly nonfunctional for a given category.

Finally, the work described here is a direct extension of the system designed by Stark [34]. This system used function-based category definitions, performed shape-based reasoning, and simulated interaction plans using the Thing World solid modeling package. The GRUFF-I system differs from this in that it contains physical visual and robotic components, operating on-line, with real objects. In addition, the GRUFF-I system has been designed to handle interaction with objects from both the furniture and dishes categories.

4. Experimental platform

An overview of the flow of control in the GRUFF-I system is provided in Fig. 1. The system is composed of three separate subsystems, as follows.

1. **Model Building** — the subsystem which uses the range image acquired with the vision component to build a 3D model of the object in the scene.
2. **Shape-based Reasoning** — the subsystem which performs static shape-based labeling of the 3D model.
3. **Interaction-based Reasoning** — the subsystem which uses the vision and robot components to perform dynamic confirmation of the 3D model, by interacting with the object in the scene and analyzing subsequent 2D intensity images.

As noted by the numbering in Fig. 1, the first step in the

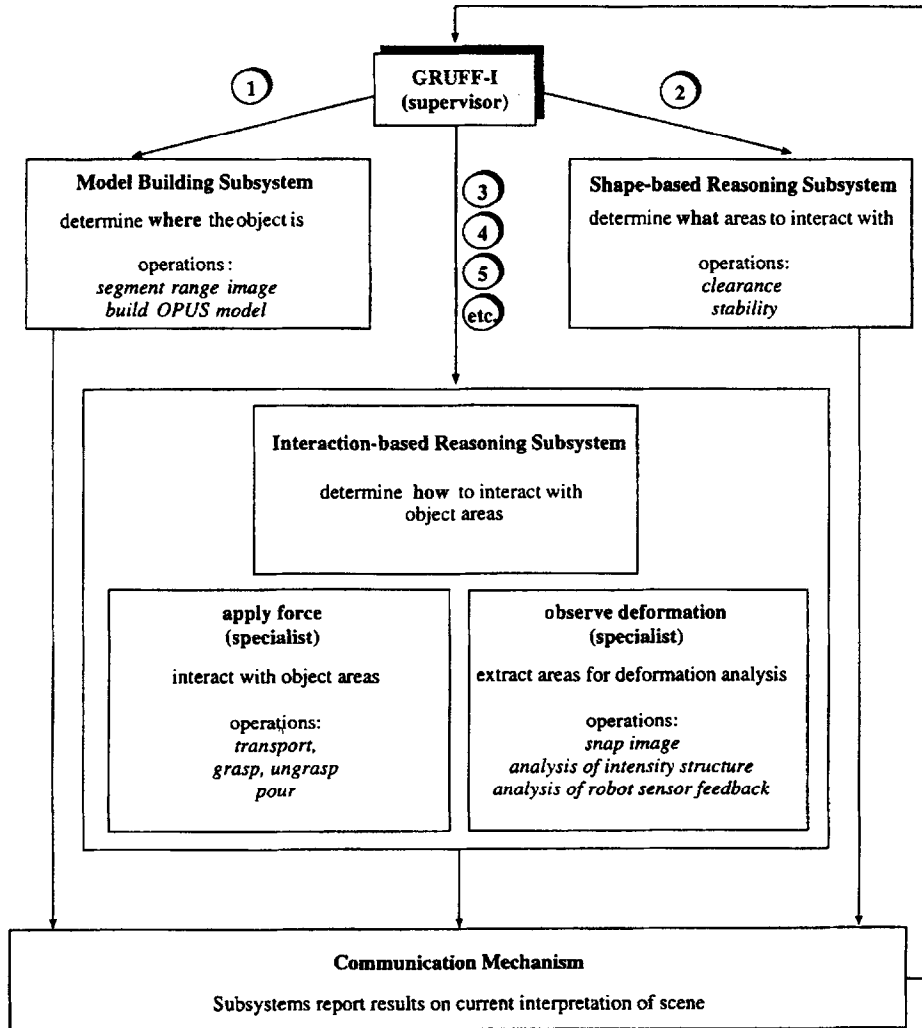


Fig. 1. The GRUFF-I subsystems communicate via shared files which contain information about how each subsystem interprets the scene.

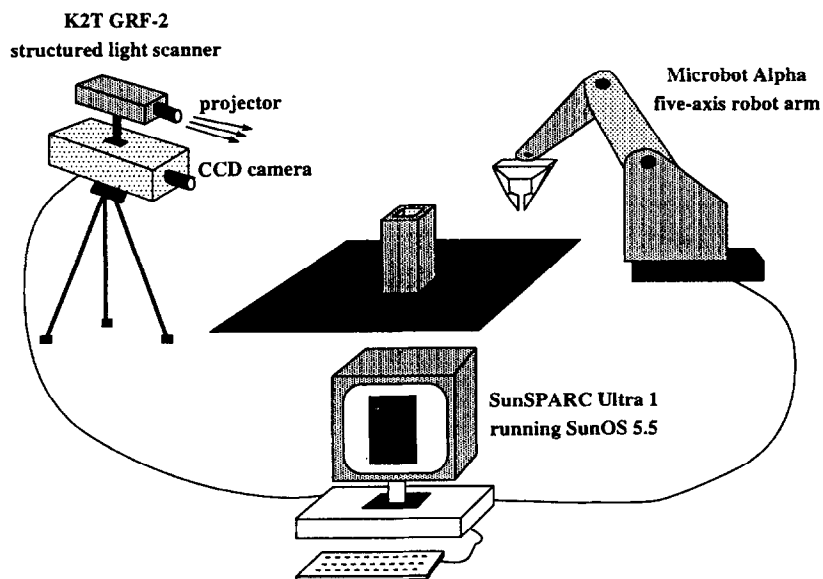


Fig. 2. The experimental set-up includes a Sun workstation, a Microbot robot arm, a K2T GRF-2 structured light scanner and circuitry to transmit/receive sensor signals.

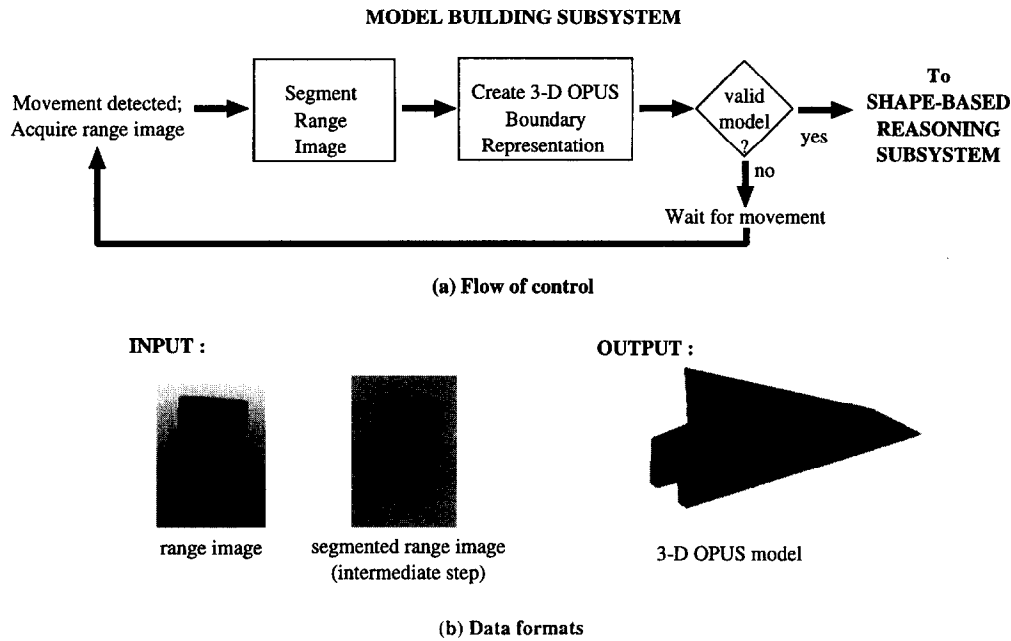


Fig. 3. In the Model Building Subsystem, the range image is segmented into approximate face regions from which an OPUS model is created.

process is to invoke the model building subsystem to build a 3D model and report information about the faces and vertices which compose the recovered model. The shape-based reasoning subsystem is then invoked to use this information and apply concepts of physics and causation to the 3D shape, using operations such as clearance and stability. The results of this analysis are reported in a text file. Finally, the interaction-based reasoning subsystem reads this text file, and is subsequently invoked multiple times, using operations to control the robot arm and the CCD camera, and to perform image processing, in order to confirm the shape-suggested functionality of the 3D model. The final results of this subsystem are also reported in a text file. The following sections describe the hardware components, the flow of control and the data formats for each of these subsystems.

4.1. System set-up and coordination

The GRUFF-I system runs on a SunSPARC Ultra 1 using the Solaris operating system (SunOS 5.5), with RS-232 connections to a robot arm and the vision component (see Fig. 2). The vision component can acquire both intensity and range images, and is composed of a K2T Gray-Code Range Finder and Sony CCD Monochrome Video Camera with a K2T V300 RGB Digitizer and Display Board (i.e. a structured light scanner). 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 1.0 m from the scene with the light pattern projector mounted just above it. The field-of-view limitations of this sensor system required some test objects to be scaled to roughly 1/10 of their normal size such that the object fit within a workspace centered by a 0.075 m cube.

The robot component uses a Microbot Alpha robot arm. This is a five-axis robotic arm with a cable-operated mechanical gripper. The arm has an open-loop coordinated point-to-point control system with the capability of automatic homing. Communication between the Sun workstation and the robot arm is done via an RS-232 serial link running at 9600 baud. The robot can be moved with a maximum speed of 0.508 m s^{-1} and is capable of lifting payloads up to 0.681 kg. The gripper can grasp objects up to a width of 0.0762 m with a programmable force of 4.3–30 Newtons. The overall positioning accuracy of the arm is roughly 0.003–0.005 m.

4.2. Overview of GRUFF-I subsystems

To initiate the GRUFF-I system, the vision component is invoked to snap intensity images (480×640) of the scene every 10 s. The workspace is then monitored for movement, such as when an object is placed in the area. Movement is detected whenever an intensity difference of greater than 10 is detected between 100 or more pixels taken from successive intensity images of the scene. If motion is detected, the vision component is invoked to acquire a range image. With a structured light scanner such as the one used in this set-up, this actually involves taking a set of intensity images as a sequence of light patterns is projected on the scene. Depth is then recovered via triangulation, with an accuracy of 0.001–0.002 m. Once the range image of the scene has been stored, control passes to the model building subsystem.

4.2.1. Model Building Subsystem

The Model Building Subsystem involves three stages, as shown in Fig. 3(a), and is based on two algorithms

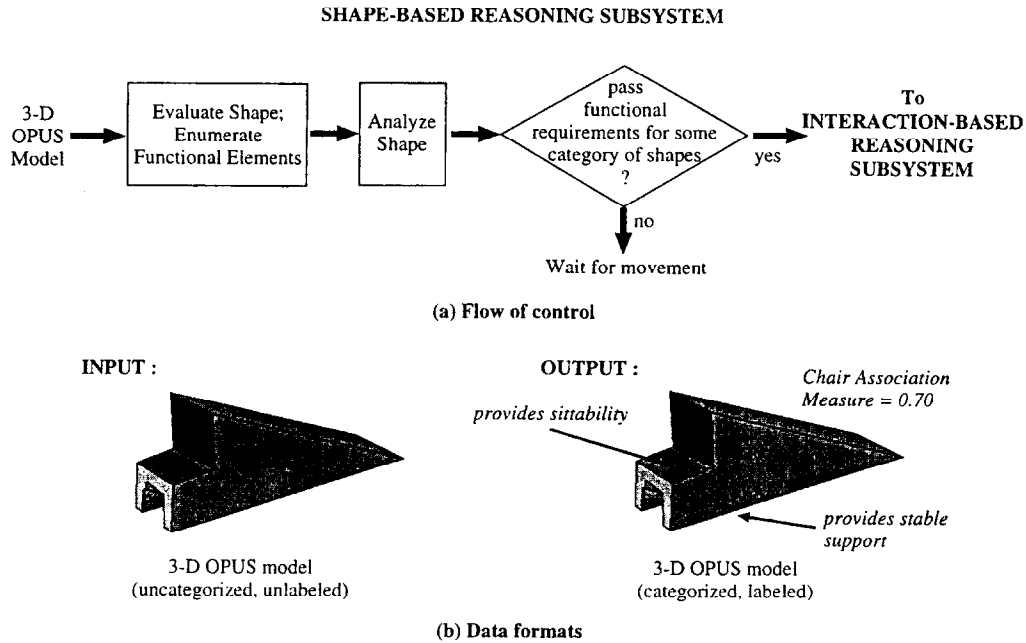


Fig. 4. The Shape-based Reasoning Subsystem performs static function-based reasoning about purely geometric properties.

developed by Hoover et al. [36,37]. The first stage involves segmenting the range image into approximate face regions. The second stage fits planar surfaces to these regions and connects them together to obtain a 3D polyhedral boundary representation model (defined in terms of faces, vertices and edges) of any object in the workspace. Finally, the third stage checks the validity of the recovered 3D model. An example of the input and output data formats used within this subsystem is shown in Fig. 3(b).

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, termed 'occlusion faces'. The resulting OPUS model of the object is a solid model enclosing the object and its surrounding occluded space. If a valid 3D model cannot be created, the system is re-initialized and the vision component is re-invoked to check for movement. If a valid model is formed, control passes to the shape-based reasoning subsystem.

4.2.2. Shape-based Reasoning Subsystem

The Shape-based Reasoning Subsystem takes as input the 3D boundary representation and applies concepts of physics and causation 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 or dishes. Each class of objects demands specific shape-based functional requirements be met, such as provides sittability or provides stable support, as shown in Fig. 4. 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 shape-based operators (termed knowl-

edge primitives (KPs)) which can be applied to determine functionality:

- clearance — determines the accessibility of some area within or around the shape;
- dimensions — determines the width, length, area, etc. of some surface;
- enclosure — determines locations of concavities in the shape;
- proximity — determines the relative location of two surfaces or characteristics of the shape;
- relative orientation — determines the orientation between two surfaces of the shape;
- stability — determines the stability of the shape in a given orientation.

As an example, the functional requirement of 'provides cup/glass containment' includes the following KP tests:

1. stability — check for stability of shape with no external forces except gravity, assuming uniform density;
2. enclosure — check for a concavity in the current orientation;
3. dimensions — check for the appropriate volume of the concavity;
4. dimensions — check for the appropriate area of enclosing surface(s) for concavity;
5. dimensions — check for the appropriate width/depth of the concavity;
6. dimensions — check for the concavity being within the appropriate height range;
7. clearance — check for appropriate clearance above the enclosing area(s) of the concavity.

The final output of the shape-based reasoning subsystem

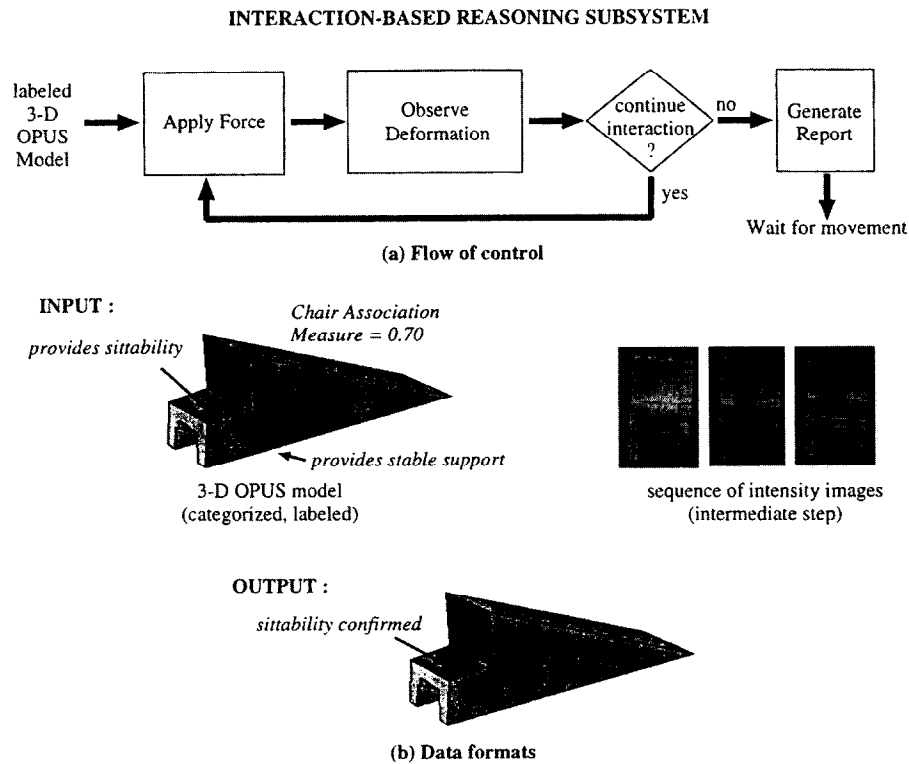


Fig. 5. The Interaction-based Reasoning Subsystem performs dynamic function-based reasoning about the sufficiency of the material properties of the object for a specified functional requirement.

is a numerical value in the range 0.0–1.0 for category membership, and the attachment of labels to various 3D surfaces, to indicate those areas of the model which are functionally significant for the specified category (termed ‘functional elements’). Provided the shape satisfies the requirements of some category, control passes to the interaction-based reasoning subsystem.

4.2.3. Interaction-based Reasoning Subsystem

The shape-based analysis in the previous subsystem assumes that the material properties of the object are sufficient. Since material property information is difficult to infer reliably from shape alone and is often necessary to confirm the usefulness of an object for a task, a plan for interacting with the object is needed. The interaction-based reasoning subsystem uses the labeled 3D model from the shape-based reasoning subsystem to produce a plan for interaction with the object, using a robot arm, as shown in Fig. 5. The robot arm is directed to absolute 3D coordinates defined in terms of the robot frame of reference. In addition, analysis of subsequent 2D intensity images taken during the interaction determines the success or failure of the interaction.

The creation of an interaction plan is accomplished by invoking a set of interaction-based primitive tests, as follows:

- Apply force — guide the robot arm during the interaction, using operations such as:

- GRASP/UNGRASP: grasp/ungrasp object at a given pair of points.
- TRANSPORT: push, pull, raise, or lower the object.
- POUR: pour a substance into the object at a specified point.

- Observe deformation — determines if the object is deforming during the interaction, in comparison to an initial reference instance of the shape, using analysis of consecutive 2D intensity images and robot sensor feedback.

For example, when an object has passed through the shape-based reasoning subsystem for the category chair, the following requirements have been met.

- Provides a sittable surface — some surface has been found which is accessible for sitting. Labeled functional elements include each surface which can provide sittability, along with an associated set of edges from these faces which have been determined to be clear for approaching to sit.
- Provides stable support — the shape can provide sittability with a force applied to points on the potential seat.

To confirm these requirements for the object in the scene, it is necessary to design the explicit interaction tests needed to confirm that the object can remain a member of the chair category. The interaction requirement ‘confirm sittability’ was designed for this purpose, and involves calls to apply force to position a weight on the potential seat and

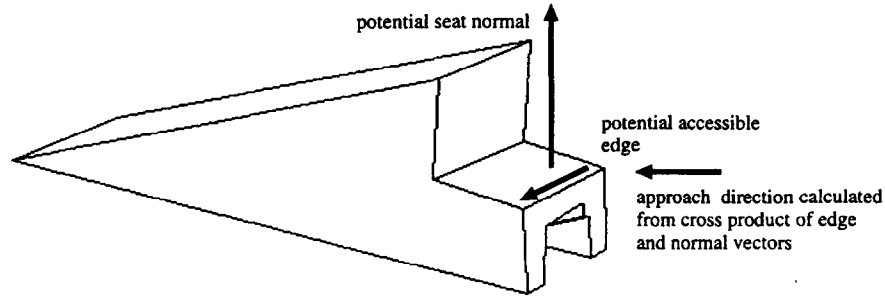


Fig. 6. The approach path for these objects is determined by the location of the sittable surface and the edges which have been determined to be accessible for sitting.

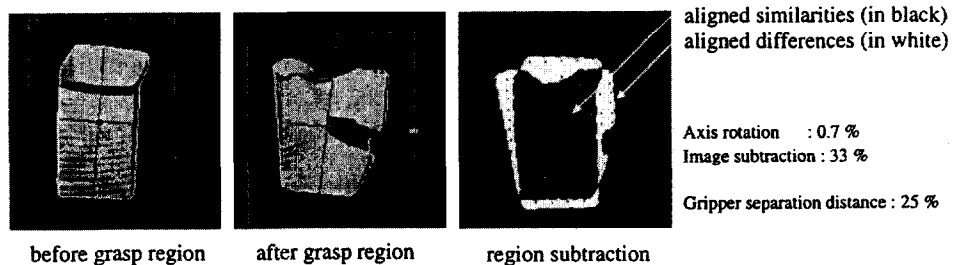
subsequent calls to observe deformation to determine if the interaction is successful.

The calls to the interaction-based primitives to perform this task are as follows:

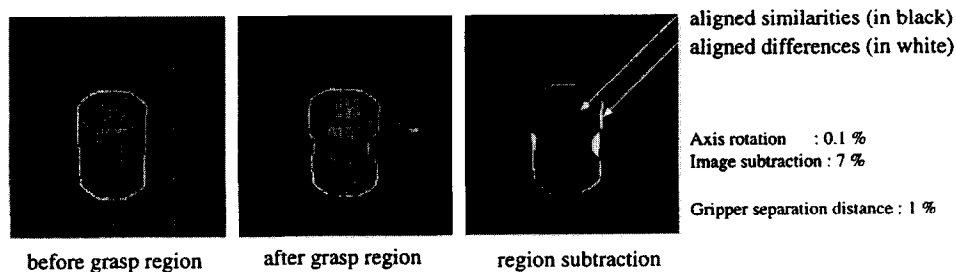
- apply force (GRASP, ...) — traverse a predetermined path to get a weight;
- apply force (TRANSPORT, ...) — position weight relative to the destination point;
- observe deformation (INTENSITY-STRUCTURE, ...) — compare images before and after weight positioning;
- apply force (UNGRASP, ...) — release the weight;
- observe deformation (INTENSITY-STRUCTURE, ...) — compare images before and after weight placement;
- apply force (GRASP, ...) — re-grasp the weight;
- observe deformation (POSITION, ...) — compare results of gripper location and separation distance before and after grasp;
- observe deformation (INTENSITY-STRUCTURE, ...)

- compare images before and after weight recovery;
- apply force (TRANSPORT, ...) — return weight to its starting position;
- apply force (UNGRASP, ...) — release the weight;
- observe deformation (INTENSITY-STRUCTURE, ...) — compare images before and after weight removed from scene.

For the chair category, the approach path is based on the path necessary to transport weight to the center of the seat. The Shape-based Reasoning Subsystem provides the information to approach the seat in a direction which is perpendicular to an edge which has been determined to be clear. The cross product of a candidate approach edge on the seat and the normal vector directed from the sittable surface is used for path planning, as shown in Fig. 6. After approaching along this direction, the robot arm places the weight in the center of the sittable surface, as estimated by the average of the vertices of this face.



(a) Results of deformation analysis for a deforming object



(b) Results of deformation analysis for a rigid object

Fig. 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 separation distance and position before and after grasping the object can be compared.

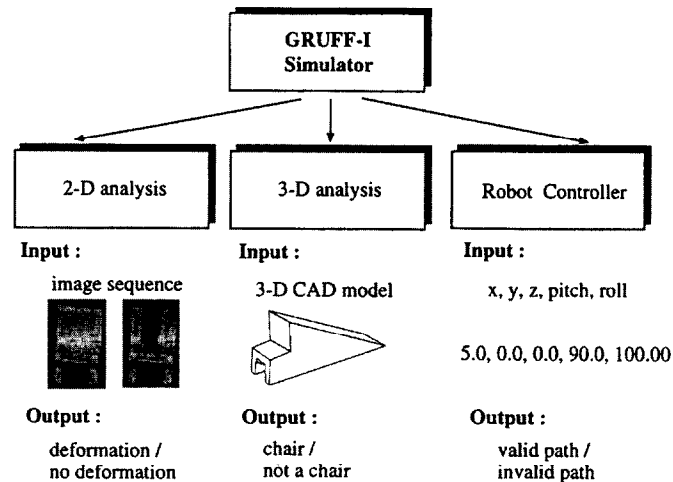


Fig. 8. The GRUFF-I system can operate in a simulation mode to test individual subsystems.

The goal of this interaction component of the system is therefore to determine the suitability or sufficiency of the material property requirements of the input shape, as opposed to the identification of its particular material composition. For example, based on interaction, the system would be able to determine that a paper maché cup is not functional due to its tendency to deform when weight is applied, as opposed to determining this by identifying that the object is composed of paper. Fig. 7 provides an additional example of the analysis of an 'apply force–observe deformation' sequence taken during interaction with a rigid wooden cup and a nonrigid paper cup.

Once this processing is completed, the final output of the system is a text report, which provides the object's numerical category membership value(s), and measures associated with how well it fulfilled each of the requirements for this category. The system then waits for movement, or the next object to be placed in the scene area.

4.3. Design of simulation mode

In the course of developing the overall system, a simulation mode was created, as shown in Fig. 8. This mode was originally designed, and continues to be used, to test and improve the algorithms used for 2D and 3D processing and the robot controller. In addition, designing and developing these modules as independent units has allowed for running and testing the system with only partial components, and allowed for better handling of equipment substitutions, when necessary.

For example, when testing 2D processing with the simulator, 3D analysis and the robot controller are disabled. A sequence of intensity images with associated sensor data feedback derived from a previous experiment or acquired using a different system is then provided as input to the 2D processing algorithm. The output indicates the success or failure of the interaction.

For testing 3D processing, the simulator disables the 2D analysis and the robot controller. It then performs shape-based reasoning on a given model, just as the original GRUFF-3 system would have. A user can use this option to test a new 3D model of an arbitrary object against the categories known to the system. Alternatively, a user could modify an existing model, by changing its position, scale, orientation, etc., and then use this option to determine how this affects categorization of the shape.

Finally, for testing the robot controller, the simulator disables 2D processing and tests the validity of any interaction paths generated by the 3D algorithm. In this option, the output indicates if any coordinates in the path fall outside the operating constraints of the robot arm.

4.4. Dimensions which affected system development

Using these hardware components and subsystems, three factors were considered important when designing, developing, and testing the GRUFF-I system, as shown in Fig. 9. The axis for the degree of interaction for the system is based on the number of robot arms employed during the interaction. For simpler categories or rigid objects, one robot arm is sufficient. However, for more complex categories of articulated objects, a cooperating pair of robot arms is necessary. The 'feedback from interaction' axis shows examples of various sensing modalities which could be attached to each arm or could communicate directly to the workstation. Finally, the 'complexity of interaction' axis shows the number of different categories of objects possible for recognition (i.e. furniture, dishes and handtools), and includes categories of objects which are rigid and articulated.

For the work described in this paper, one robot arm was used. Sensor data is derived from the vision component and the built-in force sensors of the robot arm. Finally, one subcategory from the furniture domain (i.e. chairs) and

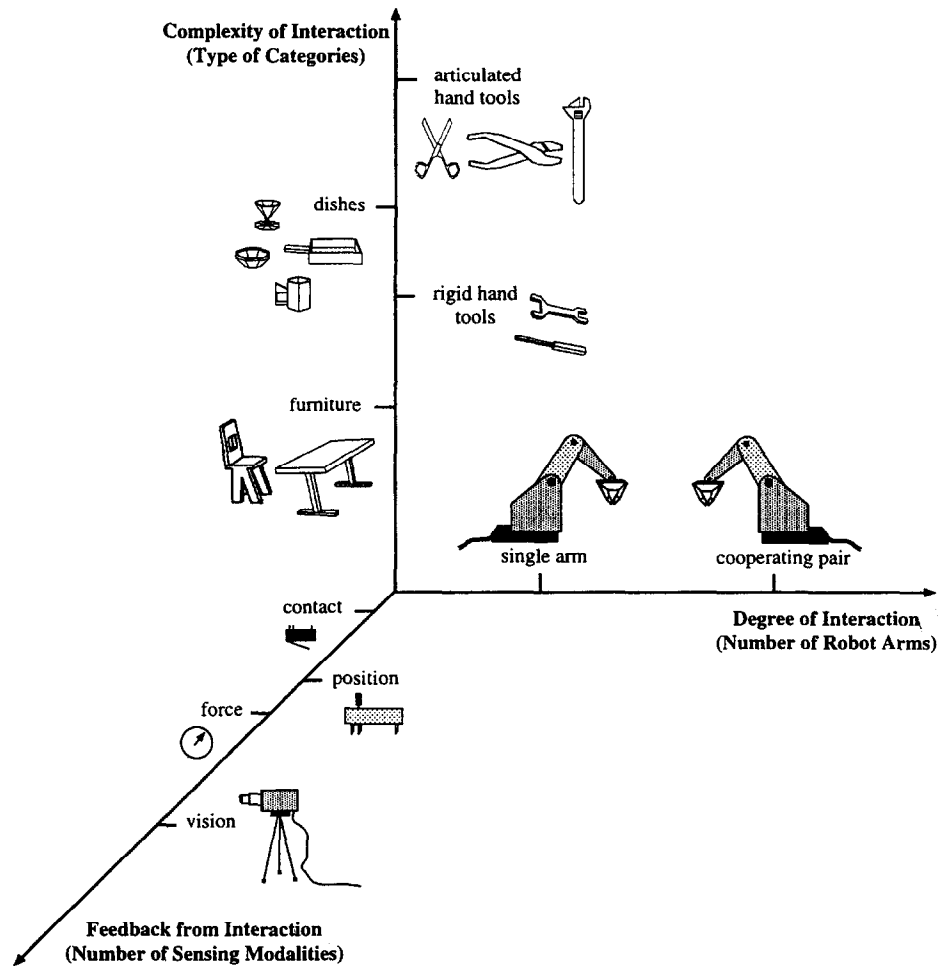


Fig. 9. Three factors influence the level of interaction possible and the overall complexity of the system.

one from the dishes domain (i.e. cups) were selected for experimentation. As this is just a subset of the possibilities presented in Fig. 9, the work in this paper can therefore be interpreted as a framework upon which more complex systems could eventually be built. The following sections examine each of the subsystems in detail and provide a set of experimental results from the integrated system.

5. Evaluation techniques

In order to evaluate the competence of the GRUFF-I system, three factors were considered significant. First, a set of test objects must be designed which meets the demands/constraints imposed by the experimental set-up and data acquisition components. This provides a means to test the robustness of each of the individual subsystems, as well as the overall system. Second, what constitutes a repeatable 'experiment' must be defined. Third, the methodology for analyzing the results of each experiment must be established. This section provides these details along with the results of testing the system with a set of physical objects.

5.1. Design of test objects

When designing the set of test objects, three characteristics were considered important, as follows.

1. Category/shape variation — the shape of the objects should be varied within a given category of objects. In addition, there should be objects from categories outside of the domain of objects known to the system.
2. Placement/orientation variation — the orientation of the object in the scene should be varied.
3. Material composition variation — the material composition of the objects should be varied to include different materials, such as paper, balsa wood, sponge and Styrofoam.

Each of these characteristics provides the opportunity to test different aspects of each of the subsystems. For example, category/shape variation provides a means to test the model building subsystem for its performance on creating both simple and complex models. More complex models would be those where there is a possibility of object self-occlusion, or those where shadow areas affect the heuristics used for surface fitting and/or joining. Shadow areas are an

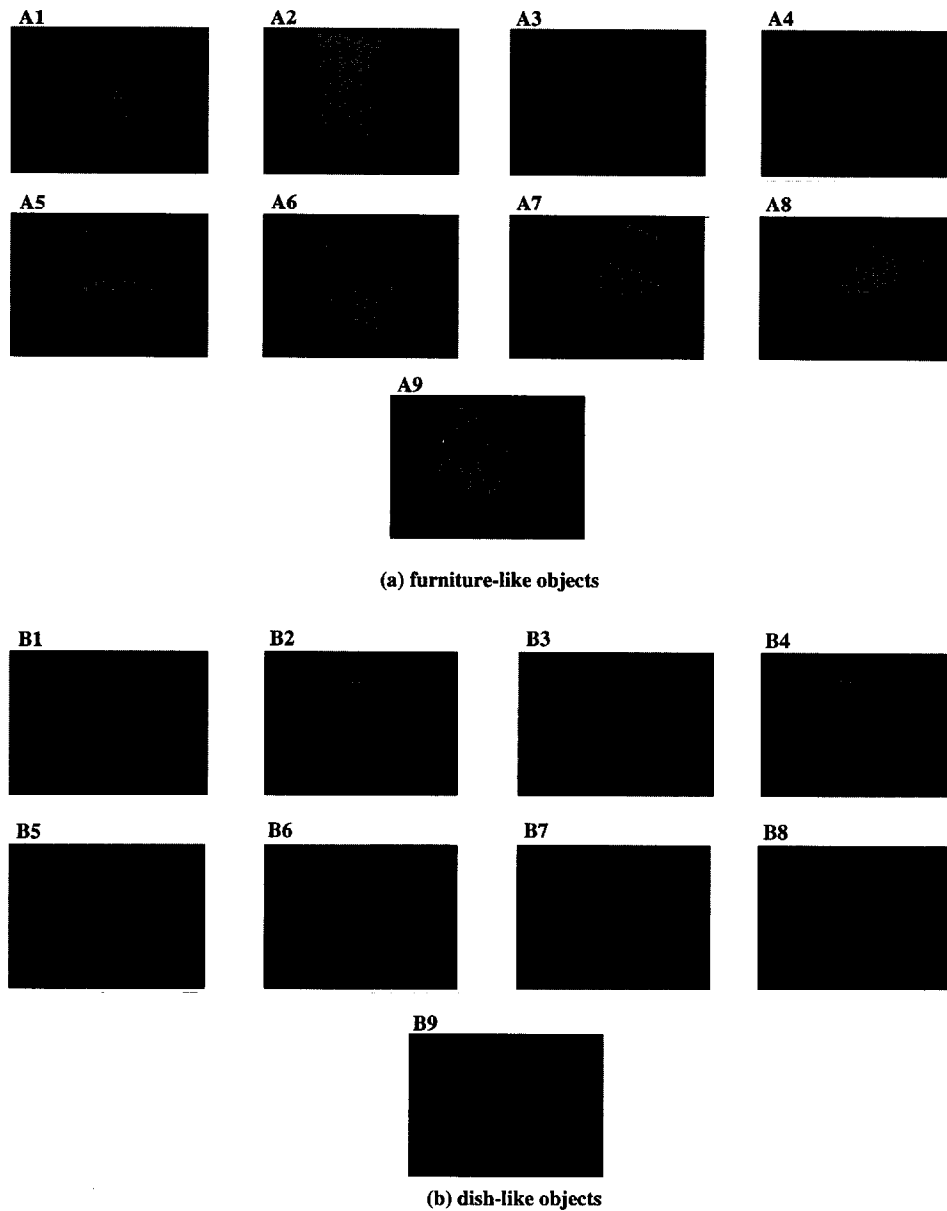


Fig. 10. Functional and non-functional object models.

inherent problem when using a structured light scanner. Whenever some surface of the object obstructs the view of the light projector, the resulting range pixels are marked as invalid. Such invalid pixels contribute to the overall complexity of the acquired model, and demonstrate the ability of the system to handle 'noise'.

Varying the placement and orientation of the object when it is positioned in the scene provides additional testing of the success of the model building subsystem and determines if placement affects heuristics used for range segmentation or model building. It also allows for testing the shape-based reasoning subsystem with objects showing varying levels of functionality for a given category. This is a result of the constraint that model recovery is performed from a single view, and if this view provides insufficient information

about functional surfaces, the object will not pass shape-based reasoning.

Variation of material composition allows the capability to test the effect of materials with different reflectivity properties on the range acquisition component of the Model Building Subsystem. This variation also provides the opportunity to test the dynamic knowledge primitives of the Interaction-based Reasoning Subsystem. Since the appearance of the different materials varies in the intensity image, varying the material composition of objects tests for any limitations of the deformation analysis used in the primitive observe deformation. For the primitive apply force, since different materials tend to have markedly different structural properties, this variation allows testing of the sufficiency of applications of forces to the object during interaction.

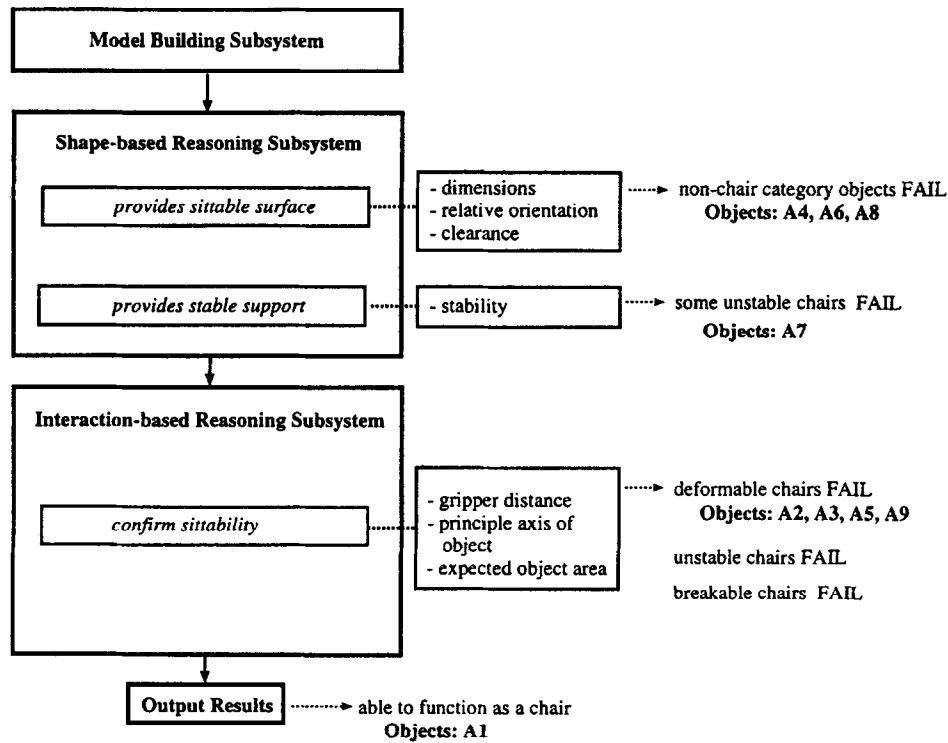


Fig. 11. Projected failure points for chair objects.

With these factors in mind, 18 object models were created from a variety of materials. An intensity image of each of these objects is provided in Fig. 10.

5.2. Recognition experiments

In the GRUFF-I system, an experiment or set of experiments is initiated after the system calibration procedures have been completed. For the results presented in this paper, two sets of experiments were run. For furniture-like objects, each object was placed in front of the scene in five different orientations, and a 0.09 kg weight composed of steel was used during the interaction. For dish-like objects, each object was placed in front of the scene in three different orientations. It is felt that this set of objects and orientations provide a broad range of test conditions to determine the feasibility of the GRUFF-I system approach. The total processing time per object is on the order of tens of minutes and is a combination of the CPU time to execute the various subsystems and the number of interaction sequences for the object category. For furniture-like objects, one interaction-based requirement must be met (confirm sittability), while for dish-like objects, three interaction-based requirements (confirm graspability, confirm containment and confirm stable containment) must be satisfied.

The first set of experiments examined the effect of interaction with the furniture-like objects in Fig. 10(a), using all three subsystems (model building, shape-based reasoning and interaction-based reasoning). The second set of

experiments examined the effect of interaction with the dish-like objects in Fig. 10(b), using the final two subsystems (shape-based reasoning and interaction-based reasoning). This latter set of experiments assumed a complete 3D model as input for the shape description, defined in terms of faces and vertices in the robot arm-centered coordinate system. This assumption was made as a result of problems with the added complexity of surface descriptions of concavities within the 3D shape in the Model Building Subsystem. In these cases, although OPUS models could be created, the topology of objects with concavities could not be handled properly by the post-processing routines used after model building. Systems which combine multiple views could form a close-to-complete model, as we are using here. However, combining views is a research topic in itself (see Refs. [38–40]) and is not in the scope of this paper.

5.3. Verification and validation methodology

Before initiating an experiment with a set of objects, the methodology for validating the results must be established. There are at least two levels at which this can occur. At the highest level, the user can enumerate a priori whether each object should or should not pass through each of the subsystems. The object can then be placed in the workspace, and the output of the system can be recorded. Alternatively, the user could enumerate the exact shape-based or interaction-based functional requirement the object should fail, if it should fail one. In this case, further details can be analyzed

Table 1
Summary of unpredicted subsystem failures

Category	Model building subsystem	Shape-based reasoning subsystem	Interaction-based reasoning subsystem
Chairs	13/45 (29%)	8/32 (25%)	3/18 (17%)
Cups	–	0/27 (0%)	7/27 (26%)

from the system to see if these more rigorous expectations are met. With these factors in mind, for each of the test objects shown in Fig. 10, the various points in the processing where each of these objects was expected to fail were enumerated, as summarized graphically in Fig. 11 for the chair category.

Inside the model building subsystem, an object may fail for legitimate reasons, such as when it is unable to satisfy all of the requirements for a category of objects. For example, noncup-like objects such as bowls and plates or objects without concavities are expected to fail the requirement ‘provides accessible glass containment’ due to the dimensional constraints imposed for concavities (including volume, height and depth). For the furniture category, objects such as tables, or chairs with nonfunctional sittable surfaces (i.e. those which have a nonparallel orientation or contain holes in the potential seat) are expected to fail the shape-based requirement ‘provides sittable surface’.

Expectations of failures in the interaction-based reasoning

subsystem are largely based on the material composition of the object and corresponding assumptions about how the structure will hold up during the interaction. Cup-like objects which are crushable, breakable or unstable are expected to fail the ‘confirm graspability’ requirement, since the gripper is expected to deform the object when grasping it. Objects with holes in the concavities, or blocked concavities, are expected to be unable to hold the containment material and therefore fail the requirement of ‘confirm containment’. In addition, somewhat unsteady cups should also fail this requirement, when the pouring material comes into contact with points on the object. Finally, bottomless cups are expected to fail the final requirement ‘confirm stable containment’, since the pouring material below the object may only be detected when the object is transported upwards. For furniture-like objects, those which are deformable when weight is applied to the seat are expected to fail the requirement ‘confirm sittability’.

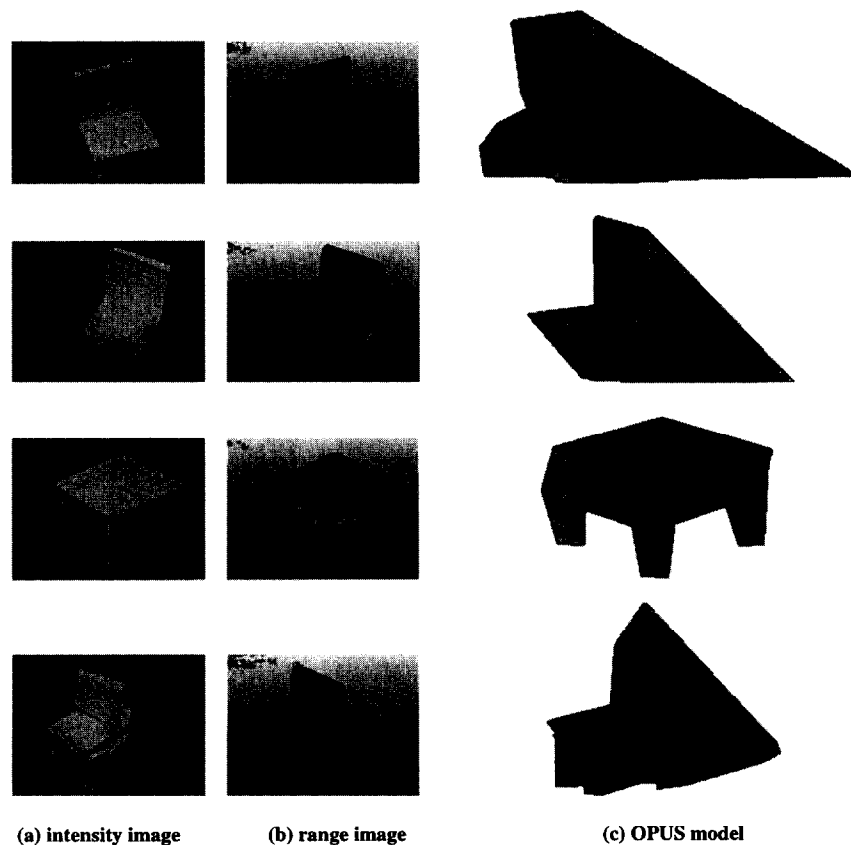


Fig. 12. Representative OPUS models for chair objects A6–A9.

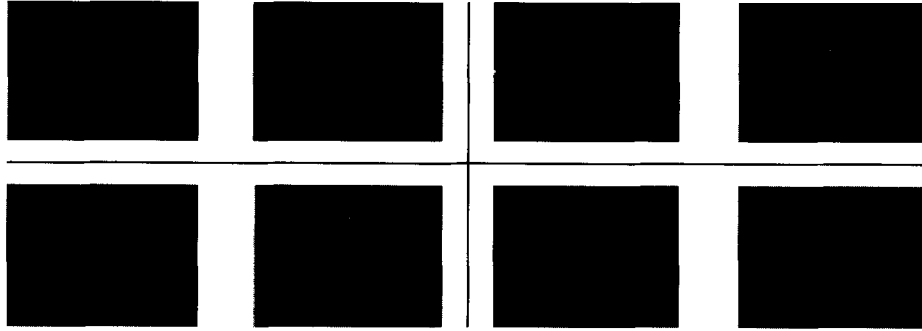


Fig. 13. Representative image sequences for cup-like objects B1–B4.

6. Demonstration of system competence

Table 1 provides a summary of the results of recognition processing, in terms of where each of the physical objects unexpectedly failed during the trials. As noted previously, the model building subsystem is disabled for cup-like objects (denoted by ‘–’ in Table 1) due to difficulties in post-processing objects with concavities. For these orientations, a complete 3D model defined in terms of the robot arm-centered coordinate system is used. Fig. 12 shows a subset of the example OPUS models acquired when testing furniture-like objects. In addition, Fig. 13 provides a subset of the intensity images acquired during testing dish-like objects. The following sections summarize specific difficulties or discrepancies within each of these subsystems.

6.1. Model Building

For this subsystem, for the furniture category of objects, the system was unable to create a valid 3D model for 13/45 orientations (29%). In 10/45 orientations, an OPUS could not be created from the segmented range image. An example of these types of errors is shown in Fig. 14. The range segmentation stage of the model building subsystem tended

to be sensitive to noise and the reflective properties of the surface on which the objects sit, which contributed to these problems. The final column of Fig. 14 demonstrates how, in many cases, pieces of the objects may not be recovered during segmentation due to being joined to the background region.

In 3/45 orientations, an OPUS model could be created, but was determined to be an invalid boundary representation, due to the constraint that two faces may only intersect each other at a common edge in the model. Fig. 15 demonstrates the three orientations in which such erroneous face intersections were encountered in the OPUS models.

6.2. Shape-based Reasoning

Unexpected results for shape-based analysis occurred for 8/32 orientations (25%) for furniture-like objects and 0% of the time for dish-like objects. In 4/32 orientations for the furniture category, the object failed shape-based reasoning when it was expected to pass. These failures were attributable to the dimensional constraints imposed for the category chair. For example, for the object shown in Fig. 16 (object A3), the dimensions of the sittable surface exceed the system thresholds for this surface. Recall that in such cases, when shape-based reasoning fails, no interaction plan

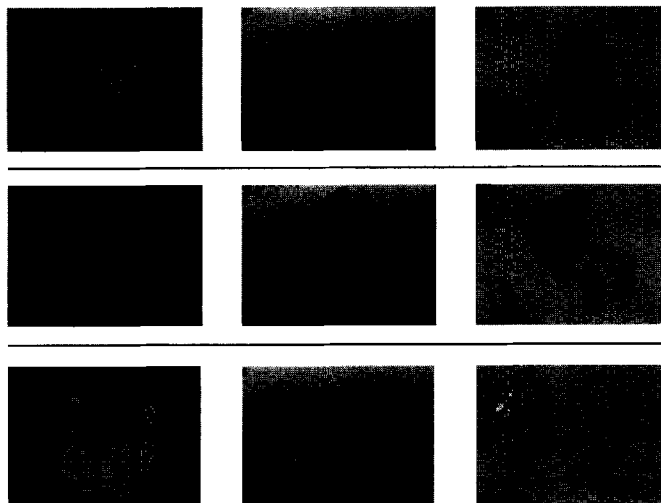


Fig. 14. Object orientations in which no OPUS model could be created.

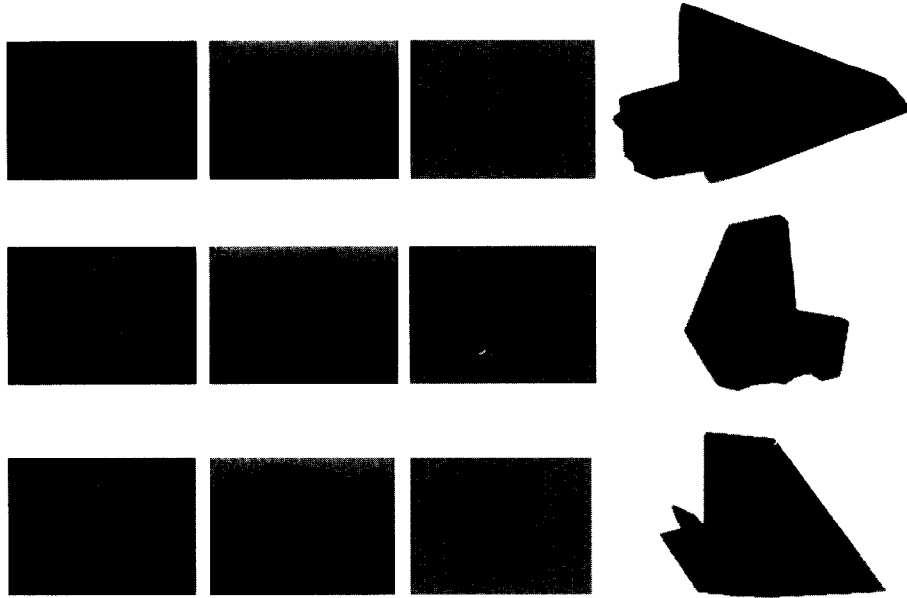


Fig. 15. Object orientations in which resulting OPUS model was invalid. The boundary representations were invalid due to face intersections in the recovered model.

is created or executed. This is one of the advantages of the GRUFF-I approach, in terms of the efficient use of system resources. There is no point to interact with an object which does not meet the minimal shape-based requirements to function as a potential member of a category of objects.

In 4/32 orientations for the furniture category, the object passed shape-based reasoning when it was expected to fail. This occurred for objects A6 (three orientations) and A7 (one orientation). For object A6, the problem is attributable to the fact that the sittable surface orientation was not significantly off from being parallel with the ground plane. For object A7, the physical model appeared unstable. However, the recovered 3D model did not give this impression. This chair, which has only one leg (see Fig. 10), has a corresponding OPUS model (shown in Fig. 12) which gives an impression of stability, allowing it to pass successfully through the 3D shape-based analysis.

6.3. Interaction-based Reasoning

For the furniture category of objects, five different objects (A1, A2, A3, A5 and A9) were expected to pass shape-based reasoning (25 object orientations). From this set, 18 orientations actually passed through the model building

and shape-based reasoning subsystems to make it to the interaction-based reasoning subsystem. Using a 0.09 kg weight, 3/18 (17%) of the orientations which were predicted to fail actually passed interaction-based analysis for the requirement 'confirm sittability'. These orientations occurred for object A2, indicating that the system considered a chair composed of thick Styrofoam to be as functional as one composed of balsa wood.

For the cup category of objects, the objects failed to be processed as expected in the interaction-based reasoning subsystem 26% of the time. Similar to the results for furniture-like objects, in 6/27 orientations the object passed interaction when it was expected to fail. This occurred for objects B6 and B7. For this category, the error could be attributed to the method used for the interaction-based requirement 'confirm containment'. For these orientations, the containment material fell outside the workspace area and could therefore not be detected. In 1/27 orientations (object B1), the object failed interaction when it was expected to pass. This occurred as a result of a positioning error in the final location of the pouring material beaker before release. In this case, the pouring substance was detected outside the object area, when none was expected.



Fig. 16. Object orientation which failed Shape-based Reasoning.

6.4. Discussion of results

It is difficult to compare this system quantitatively to other systems, since there is no ground truth set of objects from which a number of systems have been tested. However, we can comment on the results of the system in terms of sensitivity to sensor noise and in terms of the repeatability of the experiments. In this case, the major trouble spots occurred in the model building and interaction-based reasoning subsystems.

Although it is possible to run the model building subsystem with various sets of parameters, for the GRUFF-I system, one parameter set was selected for all test objects, with the primary objective of minimizing the average residual value in the constructed valid model (see Refs. [41,42] for details). A smaller residual value was sought to ensure the models would more closely match the actual range data. However, due to the number of model building errors observed, it may be appropriate to consider the effect of using scale space processing instead of a single set of parameters [41,43]. This would mean that different sets of parameters would be applied to the range image until a valid model was formed, leading to a larger set of object orientations which could ultimately be tested with the final two subsystems.

In addition, Hoover outlines a number of strategies for handling shadow areas, including allowing pixels on the border of shadow regions to bleed across foreground and background boundaries [41]. Experimenting with these strategies for both categories of objects also has the potential to increase the success of the model building subsystem. However, the cost of incorporating such bleed-through strategies would be a potentially decreased accuracy of the constructed model. For example, such heuristics would allow real object faces to extend where none exist, which could have detrimental results in the interaction-based reasoning subsystem, when path planning for specific points on the object.

Upon reaching the interaction-based reasoning subsystem, the accumulated error of all the systems became a factor to the success or failure of the robot arm's path execution, since the scale of the objects had to be reduced to meet the workspace constraints. Sources of error which contribute to the overall error include the following:

- Calibration Procedure — approximations are used for the parameters for the relationship between the coordinate systems for the projector, the CCD camera, the calibration cube, the workspace and the robot arm.
- Model Building Subsystem — various heuristics are used in both the segmentation and surface fitting stages.
- Interaction-based Reasoning Subsystem — the robot control routines use an approximation for the kinematics of the arm and are also affected by the mechanical problems associated with the arm, such as motor slippage.

Since the GRUFF-I system is an endpoint open loop system (vision is not used for correcting path planning), this accumulated error can cause the gripper to be positioned incorrectly for a specific interaction test. The positioning discrepancy generally affected the final results only for smaller objects.

Finally, the errors within the Interaction-based Reasoning Subsystem can be attributed to the limited amount of sensor feedback when testing the requirements 'confirm sittability' and 'confirm containment'. For example, the experiments with furniture indicated the dependence of the 'confirm sittability' requirement on a reasonable amount of applied weight. A more rigorous test for confirm sittability could test objects for degrees of functionality, using a range of weights. Alternatively, the incorporation of a force sensor at the end of the weight could provide feedback about the object's compliance. Along these lines, to handle the cases for the requirement 'confirm containment', when the pouring material was not detected within the camera's view, a weight sensor could be added to provide feedback on the weight of the object before and after testing this requirement. In either case, the added observational powers of the system would be at the cost of increased circuitry, hardware and processing time to acquire and record the measurements.

7. Conclusion/future work

Previous versions of the GRUFF system could only indicate the potential sufficiency of a given 3D shape for the functional requirements of some category. The work described in this paper incorporates actual visual and robotic components, operating on-line, to perform interaction with physical objects. The GRUFF-I system can distinguish between functional and nonfunctional objects, based either on shape alone or on the results of interactions which are suggested by successful shape analysis. This system therefore operates 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 the material properties of an object are sufficient. The results of this preliminary analysis (yielding locations of functionally significant areas on the object) provides the guidance to instantiate further (more expensive) top-down exploratory modules to confirm the shape-suggested functionality.

The original contribution of this work is that it is the first nonpart-based approach to use function-based reasoning to symbolically label and plan for interaction with objects. The existing research in generic object recognition includes a variety of input formats (including 2D and 3D data) and constraints on input orientations. Alternatively, the GRUFF-I system is implemented to evaluate a 'static' uninterpreted rigid 3D shape obtained from real image data. A set of physical artifacts was created and used to evaluate the prototype system, by placing each object in the workspace

in a number of orientations. Additional contributions of this work include the following:

Metrically accurate representations of the world can be built and used for higher level reasoning. Valid OPUS models were obtained approximately 70% of the time for furniture-like objects. Based on a set of transformation equations determined during calibration, these models were ultimately defined in terms of the robot arm-centered coordinate system. Path planning with a Microbot robot arm using 3D points in this coordinate system resulted in an overall accuracy of approximately 0.01 m.

Shape-based reasoning prior to interaction-based reasoning provides an efficient methodology for object recognition, in terms of the judicious use of system resources. In the GRUFF-I system, shape-based functional object recognition provides the means to visually explore (pre-process) the object as much as possible to develop a list of intelligent locations for subsequent higher level top-down processing. If the object does not have the minimal shape-based characteristics to function as a member of a class of objects, no interaction plan is created. In addition, objects in nonfunctional orientations will not yield interaction plans. Several researchers, such as Bogoni and Bajcsy [44] and Krotkov [45] have focused on the recognition of the material properties of objects, without considering shape. However, robotic control can be expensive in terms of positioning, re-calibration and time. Pre-recognition allows us to visually explore the object's 3D model as much as possible, to develop a plan of minimal work for confirming object functionality. The cost of such shape-based analysis is directly proportional to the complexity of the model. This analysis is therefore not computationally cheap in an absolute sense but appears cheaper relative to the costs involved with interacting, independent of object shape.

Interaction-based reasoning can be used to confirm the functionality of a categorized object without explicitly determining the object's material composition. It may be entirely possible to code algorithms to distinguish between the image properties of various materials such as wood, sponge and paper, using reflective properties, texture, etc. However, such systems can ultimately be fooled if the object is covered with an external surface having favorable 'rigid' properties. In the GRUFF-I system, shape-based analysis operates under the assumption that the material properties of the object are sufficient. Interaction-based analysis confirms or disproves this assumption by interacting with the object in a task consistent with its shape-predicted functionality. This determines if, regardless of composition, the object will retain its functionality under use.

The goal of this work was not to solve the whole problem of machine perception, nor to develop a perfect system for a special pre-defined visual mini-world. Rather, the objective was to develop a set of algorithms for three subsystems, integrate the results, and demonstrate the competence of the resulting prototype system. Although there were a

number of constraints on the workspace, researchers can make use of the results of this work in richer visual environments in the following ways:

In terms of the *sufficiency of the representation/organization* within the GRUFF-I system, the required knowledge of the world is organized hierarchically by a series of shape-based and interaction-based requirements for category membership. Research in human perception by Rosch et al. supports the use of such hierarchical category classes [3]. In terms of *breadth*, the system is extendible to include both multiple domains of objects and multiple sensors. In addition, the system demonstrates that a small knowledge base of shape-based and interaction-based knowledge primitives is sufficient to test and confirm a variety of functional requirements. The GRUFF-I system is also *generative*, in the sense of [46], in that it is capable of recognizing artifacts which have never been encountered before.

In terms of *practicality*, the GRUFF-I system approach can be differentiated from the work of Stansfield, who used a six-sided spatial polyhedral for designing interaction plans [32]. The determination of the locations of functionally significant areas by the GRUFF-I system does not limit the approach or grasp strategies. Furthermore, analyzing the entire shape for functionally significant areas is more useful than basing recognition on matching to prototypical part descriptions or color-based blob characteristics, in the sense of Rivlin et al. [47] and Connell [31].

In addition to the changes suggested in the discussion of results, such as adding force and weight sensors to provide additional feedback when testing requirements such as 'provides sittability' and 'confirm containment', there are a number of future directions for this work. One possibility would be the incorporation of building and analyzing 3D models during each stage of the interaction and the design of a corresponding set of 3D deformation analysis routines. This suggests the possibility of integrating and weighing different sources of information from different subsystems and the possibility of exploiting parallel programming to increase efficiency. It may also be possible to design heuristics guided by information from registered range and intensity images for object detection and deformation analysis.

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 functional information' for model building. This suggests the need for the development of another interaction-based primitive which could be used to re-orient the object to acquire these additional views.

Finally, in the current implementation, isolated objects are assumed. However, more interesting applications are possible with scenes of objects. Toward this end, Hoover has designed a system for acquiring what he calls a space envelope representation of the world [41]. An example of

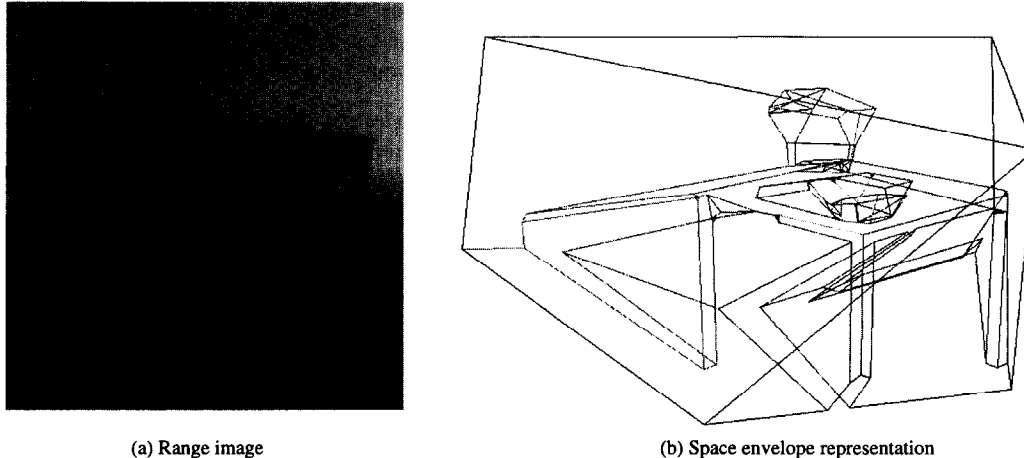


Fig. 17. Functionality in the 'large'.

this representation is shown in Fig. 17. This offers an exciting opportunity to consider the 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.

The development of representations which will support generic object recognition is clearly of fundamental importance if the goal of autonomous real-world systems is to be achieved. Just as clearly, this area of research is still in its infancy. Object representations which are eventually developed to support this goal are likely to incorporate elements of all the types of models and sensor data described here, as well as information from additional modalities.

Acknowledgements

This research was supported by AFOSR grant F49620-92-J-0223, NSF grant IRI-9120895 and a NASA Florida Space Grant Consortium graduate fellowship.

References

- [1] S. Palmer, E. Rosch, P. Chase, Canonical perspective and the perception of objects, in: J. Long, A. Baddeley (Eds.), *Attention and Performance IX*, Lawrence Erlbaum Associates, 1981, pp. 135–151.
- [2] E. Rosch, Human categorization, in: Neil Warren (Ed.), *Studies in Cross-cultural Psychology*, Academic Press, New York, 1977, pp. 1–49.
- [3] E. Rosch, C.B. Mervis, W. Gray, D. Johnson, P. Boyes-Braem, Basic objects in natural categories, *Cognitive Psychology* 8 (1976) 382–439.
- [4] L. Dey, P.P. Das, S. Chaudhury, Recognition and learning of unknown objects in a hierarchical knowledge-base, in *AAAI Fall Symposium Series — Technical Report FS-93-04*, 1993, pp. 15–19.
- [5] P. Winston, T. Binford, B. Katz, M. Lowry, Learning physical descriptions from functional definitions, examples, and precedents, *AAAI*, 1983, pp. 433–439.
- [6] M. Minsky, *The Society of Mind*, Simon and Schuster, New York, 1985.
- [7] J. Hodges, Naive mechanics — a computational model of device use and function indesign improvisation, *IEEE Expert* Feb (1992) 14–27.
- [8] J. Hodges, Functional and physical object characteristics and object recognition in improvisation, *Computer Vision and Image Understanding* 62 (2) (1995) 147–163.
- [9] M. Brand, Vision systems that see in terms of function, in *AAAI-93 Workshop on Reasoning about Function*, Washington, DC, July 1993, pp. 17–22.
- [10] P.R. Cooper, L.A. Birnbaum, M.E. Brand, Causal scene understanding, *Computer Vision and Image Understanding* 62 (2) (1995) 215–231.
- [11] R.A. Brooks, *Model-Based Computer Vision*, Vol. 85, UMI Research Press, Ann Arbor, MI, 1984.
- [12] R.A. Brooks, Model-based three-dimensional interpretations of two-dimensional images, in: M.A. Fischler, O. Firschein (Eds.), *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, Morgan Kaufmann Publishers, Inc., 1987, pp. 360–370.
- [13] M. Brady, P.E. Agre, D.J. Braunegg, J.H. Connell, The mechanics mate, in: T. O'Shea (Ed.), *Advances in Artificial Intelligence: European Conference of Artificial Intelligence*, Elsevier Science Publishers, Amsterdam, 1985, pp. 79–94.
- [14] J.H. Connell, M. Brady, Generating and generalizing models of visual objects, *Artificial Intelligence* 31 (1987) 159–183.
- [15] T. Kitahashi, N. Abe, S. Dan, K. Kanda, H. Ogawa, A function-based model of an object for image understanding, in: H. Jaakkola, H. Kanassalo, S. Ohsuga (Eds.), *Advances in Information Modelling and Knowledge Bases*, IOS Press, 1991, pp. 91–97.
- [16] E. Davis, Shape and function of solid objects: some examples, *Technical Report 137*, New York University, October 1984.
- [17] A.K. Goel, Integrating case-based and model-based reasoning, *AI Magazine* (1992) 50–54.
- [18] L. Stark, 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*, 1991, pp. 251–256.
- [19] L. Stark, K.W. Bowyer, Function-based recognition for multiple object categories, *CVGIP: Image Understanding* 59 (1) (1994) 1–21.
- [20] M.A. Sutton, L. Stark, K.W. Bowyer, GRUFF-3: generalizing the

- domain of function-based recognition system, *Pattern Recognition* 27 (12) (1994) 1743–1766.
- [21] M.A. Sutton, H. Liu, L. Stark, K.W. Bowyer, Towards a domain-independent function-based recognition system, in *Proceedings of Florida Artificial Intelligence Research Symposium*, 1995, pp. 294–298.
- [22] E. Rivlin, S.J. Dickinson, A. Rosenfeld, Recognition by functional parts, *Computer Vision and Image Understanding* 62 (2) (1995) 164–176.
- [23] D. Kim, R. Nevatia, A method for recognition and localiation of generic objects for indoor navigation, in *IEEE Workshop on Applications of Computer Vision*, 1994, pp. 280–288.
- [24] L. Bogoni, Identification of functional features through observations and interactions, Ph.D. thesis, University of Pennsylvania, 1995.
- [25] L. Bogoni, R. Bajcsy, Interactive recognition and representation of functionality, *Computer Vision and Image Understanding* 62 (2) (1995) 194–214.
- [26] Z. Dric, J. Fayman, E. Rivlin, Recognizing functionality, in *Proceedings of International Symposium on Computer Vision*, 1995, pp. 247–252.
- [27] Z. Dric, J.A. Fayman, and E. Rivlin, Function from Motion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (6) (1996) 579–591.
- [28] K. Kise, H. Hattori, T. Kitahashi, K. Fukunaga, Representing and recognizing simple hand-tools based on their functions, in *Asian Conference on Computer Vision*, 1993, pp. 656–659.
- [29] K. Green, Generic Recognition of Articulated Objects Through Reasoning About Potential Function, Ph.D. thesis, University of South Florida, 1996.
- [30] K. Green, D. Eggert, L. Stark, K. Bowyer, Generic recognition of articulated objects through reasoning about potential function, *Computer Vision and Image Understanding* 62 (2) (1995) 177–193.
- [31] J.H. Connell, Get me that screwdriver! Developing a sensory-action vocabulary for fetch-and-carry tasks, *IBM CyberJournal Research Report RC 19473*, April 1994.
- [32] S.A. Stansfield, Haptic perception with an articulated, sensate robot hand, *Robotica* 10 (1992).
- [33] S.A. Stansfield, Representing generic objects for exploration and recognition, in *Proceedings of IEEE International Conference on Robotics and Automation*, 1988, pp. 1090–1096.
- [34] L. Stark, Recognizing object function through reasoning about 3-D shape and dynamic physical properties, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 251–256.
- [35] E. Rivlin, A. Rosenfeld, Navigational functionalities, *Computer Vision and Image Understanding* 62 (2) (1995) 232–244.
- [36] A. Hoover, D. Goldgof, K.W. Bowyer, Extracting a valid boundary representation from a segmented range image, *IEEE transactions on Pattern Analysis and Machine Intelligence* 17 (9) (1995) 920–924.
- [37] A. Hoover, D. Goldgof, K.W. Bowyer, Building a b-rep from a segmented range image, in *Proceedings of IEEE Second CAD-Based Vision Workshop*, 1994, pp. 74–81.
- [38] Y. Chen, G. Medioni, Surface descriptions of complex objects from multiple range images, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, pp. 153–158.
- [39] Z. Chen, S.Y. Ho, Incremental model building of polyhedral objects using structured light, *Pattern recognition* 36 (1) (1993) 33–46.
- [40] J. Maver, R. Bajcsy, Occlusions as a guide to planning the next view, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15 (5) (1993) 417–433.
- [41] A. Hoover, The Space Envelope Representation for 3D Scenes, Ph.D. thesis, University of South Florida, 1996.
- [42] M.A. Sutton, Function from visual analysis and physical interaction: a methodology for recognition of generic classes of objects, Ph.D. thesis, University of South Florida, 1997.
- [43] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.
- [44] L. Bogoni, R. Bajcsy, An active approach to characterization and recognition of functionality and functional properties, in *AAAI Workshop: Reasoning About Function*, 1993, pp. 9–16.
- [45] E. Krotkov, Perception of material properties by robotic probing: preliminary investigations, in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1995, pp. 88–94.
- [46] M. Brand, An eye for design: why, where and how to look for causal structure in visual scenes, in *Proceedings of SPIE Workshop on Intelligent Robots and Computer Vision XI: Algorithms, Techniques and Active Vision*, 1992, pp. 180–188.
- [47] E. Rivlin, A. Rosenfeld, D. Perlis, Recognition of object functionality in goal-directed robotics, in *AIII-93 Workshop on Reasoning about Function*, Washington, DC, July 1993, pp. 126–130.