Towards a Domain-Independent Function-Based Recognition System*

Melanie Sutton,[1] Haihua Liu,[1] Louise Stark,[2] Kevin Bowyer[1]

1. Department of Computer Science & Engineering
University of South Florida
Tampa, FL 33620-5399 USA
http://marathon.csee.usf.edu/
sutton, liuh, or kwb @csee.usf.edu

2. Department of Electrical & Computer Engineering
University of the Pacific
Stockton, CA 95211 USA
stark@napa.eng.uop.edu

Abstract

We present a system which takes as input a 3-D model, and produces as output a labeling indicating those portions of the shape which serve the functional requirements of some category. Presently, the reasoning is from shape alone, using concepts of physics and causation (e.g., to infer stability). The designed *form follows function* representation is powerful for a number of reasons. First, the requisite knowledge of the world is organized hierarchically by a series of specifications on the shape. The interpretation of this organization for a given shape facilitates the development and evaluation of alternative hypotheses to arrive at some categorical decision. Finally, the breadth and applicability of the approach is demonstrated by the system's extendability to multiple domains, where it has continued to function efficiently with very little additional knowledge.

1 MANUFACTURED FUNCTIONALITY

A number of objects that a robot may encounter in the world exhibit the presence of manufactured functionality. They were designed for and are used with a specific purpose in mind. The recognition of such artifacts (chairs, hammers, etc.) using function-based reasoning is therefore important because such objects are likely to be plentiful in a robot's environment. Within this context, we will demonstrate the feasibility of a *form follows function* approach to recognition which considers *design* (based on shape alone) first. The input to our system is the 3-D description of an object, such as those shown in Figure 1. The system identifies portions of the shape (e.g., a single face or group of faces) with specific uses (termed *functional elements*), and then offers one or more categorical labels for the object with corresponding measure(s) of goodness.

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

2 PREVIOUS WORK

Due to the complexity of object recognition, a good portion of the previous research has concentrated on the recognition of specific domains of objects, and no method has produced fool proof results. For example, many different CAD-based methods have been used for object recognition with only moderate success. It is clear that any approach which assumes that the recognition system is given an exact model for each object that it is to recognize is inadequate to support "general purpose" vision. Similarly, modeling approaches which attempt to capture a category of objects by generalizing the model of some particular prototype object are equally insufficient. For any reasonably complex environment, each of these approaches is ultimately limited by either the storage space required for the set of models or the processing speed at which matching can be performed. Even given sufficient computing resources, it is unlikely that all the parameterizations or component descriptions necessary to sufficiently generalize a prototype object model could readily be anticipated to capture entire categories of objects.

Popular early function-based approaches include Winston, *et al.*, where segmented, interpreted descriptions of objects are used, bypassing the inherent difficulty in deriving such information [17]. More recent methods have been reported which give a variety of implementation methodologies to achieve generic object recognition, including semantic nets [3, 5, 6], expert systems frameworks [7], direct shape analysis [9], and fuzzy set and possibility theory [16]. Bogoni and Bajcsy focus on the verification

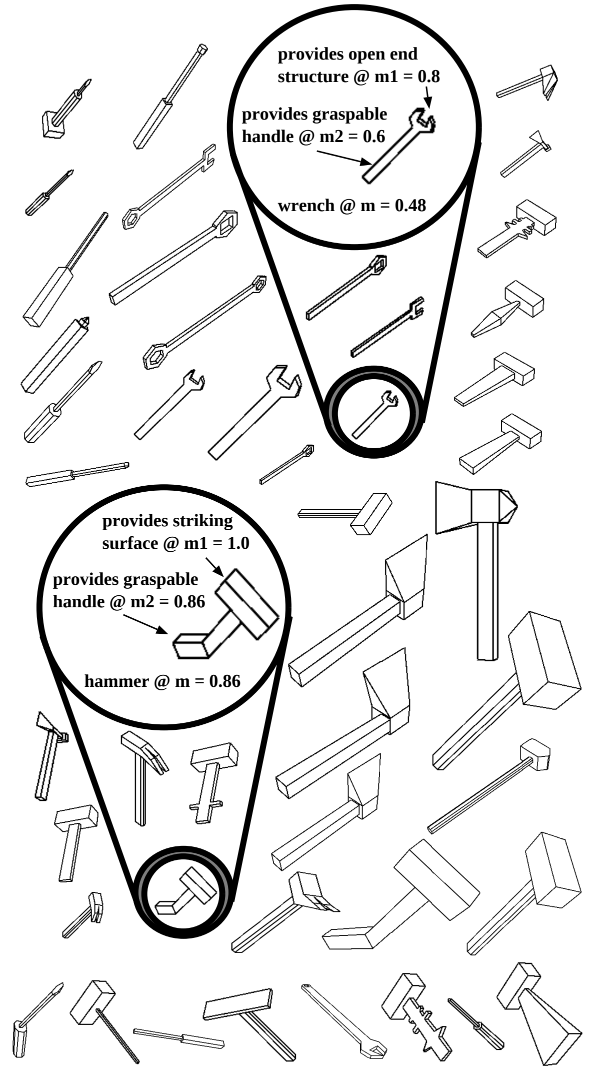

Figure 1: Shape-based distinctions across categories allow recognition (in terms of *categorization*) and labeling of important functional elements of the shape.

and recovery of the material properties of an object, using exploration techniques such as piercing [1]. Brand describes a system using causal and functional analysis to interactively build a model of relationships between parts in a scene [4]. Hodges has developed the EDISON system which uses computational models that can solve problems and reason creatively about mechanical devices [8]. Rivlin, *et al.* explore the issue of functionality in the context of object recognition at the functional part level for the category hammers [10]. Stansfield incorporates a feature-based approach, where objects are matched to prototypical exemplars [12]. Stark, *et al.* explore function-based recognition of objects using partially incomplete 3-D shape descriptions extracted from LRF images [14].

The input formats (including 2-D and 3-D data) and constraints on input orientations presented to the recognition processing vary considerably among these methods. The **GRUFF** system (**G**eneric **R**ecognition **U**sing **F**orm and **F**unction) is implemented to evaluate a "static" *uninterpreted* rigid 3-D shape, and the input orientations of the object are not restricted. We propose a method which is differentiated from other work in this area in that the system performs shape-based functional object recognition as a means to visually explore the object as much as possible, prior to any higher level top-down processing. The work described here expands on our earlier work in that we have applied these assumptions about generic object recognition to a new domain, hand tools.

3 THE REPRESENTATION OF FUNCTIONAL KNOWLEDGE

In order for the **GRUFF** system to be able to recognize objects in the category hand tools, functional knowledge of that superordinate category had to be added to the system.[1] This involves organizing the requisite knowledge of the new artifacts in terms of functional properties. Since we are dealing with manufactured objects, these properties can be described for each category of objects based on the expected use of the objects in everyday tasks. For example, a hammer is expected to be used for pounding tasks, allowing us to stipulate various requirements on the shape such as those shown in Figure 3. The *concept* of a set of hand tools is then understood by the system as a hierarchically organized tree of categories and functional requirements.

An initial investigation of the shape allows the system to first discern the relevance of the various categories, rather than parsing all possible categories. Orderings are based on domain-independent characteristics of the shape and its 3-D convex hull, such as volume or surface area. Once a category has been selected, important details are built up from further shape analysis. In this way, the system is able to reason, for example, that the shape is in fact a hammer, and also postulate the various functional "pieces" of the shape which allow it to be used in this manner.

This hierarchical processing also facilitates the development of alternative hypotheses to be tested. If the shape is ambiguous (able to satisfy the requirements of more than one category), the analysis will furnish alternative hypotheses to be systematically explored. These interpretations may subsequently provide different labels for the various portions of the shape. An interpretation of the shape as a member of one category may also be terminated (due to threshold considerations), while another interpretation is processed further.

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

[1] See [11] for research supporting the use of category classes in human perception.

Evaluations of the shape - each evaluation returns a measure of certainty that some constraint on the shape was met.

Integration - a series of shape evaluations is necessary to determine the compatibility of the shape with a target functional requirement or category. Rules for combining the compatibility values are such that the level of membership in a category can be no stronger than the weakest value from any one of the set of required evaluations.

Categorical decisions - when faced with a set of alternative interpretations of the shape, categorical decisions are determined by the highest final *association measure* to some category.

3.1 Evaluation: Applying Primitive Functional Operators

In order to use the properties shown in Figure 3 to recognize objects, the functional requirements must be converted into calls to appropriate operators. There are six operators (termed *knowledge primitives or KPs*) which can act on the shape to recover relevant information: *clearance, dimensions, enclosure, proximity, relative orientation* and *stability.* An example of how these might be used with the new artifacts, hand tools, is shown in Figure 2.

It is important to note that these primitives are domain-independent and are not "model-based," in the sense of earlier CAD-based methods for object recognition. For example, there is no mapping to a prototypical representation of a handle, but rather a series of geometrical tests to confirm properties of "handleness." This set of primitives has therefore been sufficient to confirm the requirements of shapes in additional categories (i.e. furniture, dishes and articulated hand tools).

3.2 Integration: Combining Functional Evidence

The output of each primitive invocation is a measure indicating the suitability of some portion of the structure. These measurements are then combined to predict what categorical label is most appropriate. The final *association measure* for the shape is accumulated in a manner such that if a required shape analysis returns an evaluation measure lower than the current association measure, it lowers the overall evidence for that (sub)category. Representative instances of three families of aggregation calculi [2] were evaluated and compared for performance on a set of example shapes for the object category "chair" in the original **GRUFF** system [13]. The operation which provided the best performance at this category level is the T-norm or conjunctive operation:

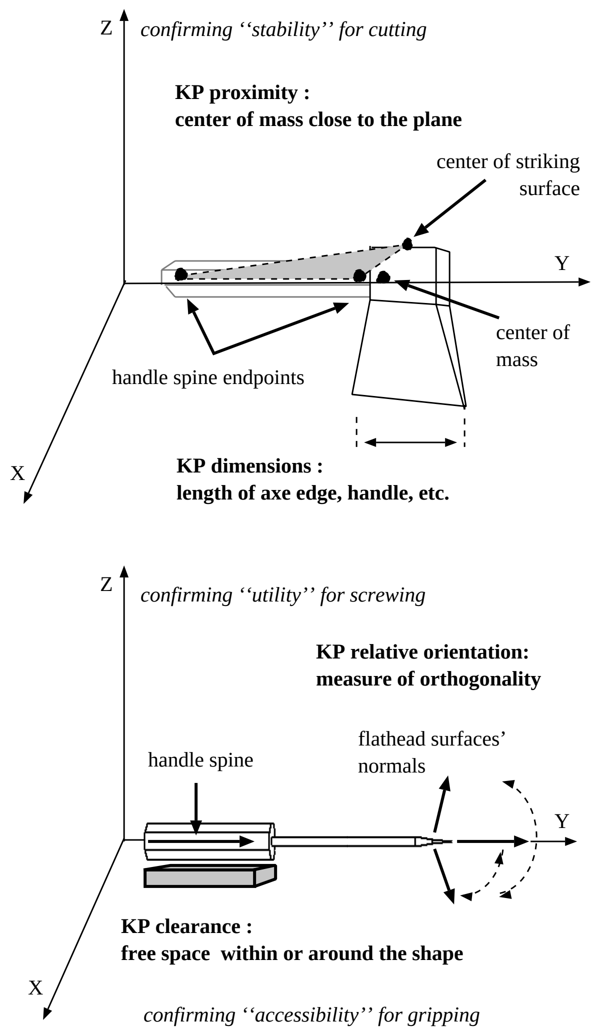

Figure 2: A series of knowledge primitives (KPs) are invoked to operate on selected portions of the shape to determine if requirements are met.

$$T(a, b) = a * b$$

(for arbitrary evaluation measures a and b). However, when the subcategory being investigated is a refinement of its parent subcategory (in the way that a claw hammer is a refinement of a regular hammer), the evidence gathered at the parent subcategory node is used in a way which raises the association measure calculated at the present subcategory node by some factor associated to the evidence gathered at the parent node, using a T-conorm operation:

$$S(a, b) = a + b - a * b$$

An example of these calculations is provided in Figure 3. In summary, after the first two stages, **GRUFF** will have analyzed an object in all possible orientations for each of the functional requirements of each category and subcategory in the hand tools tree.

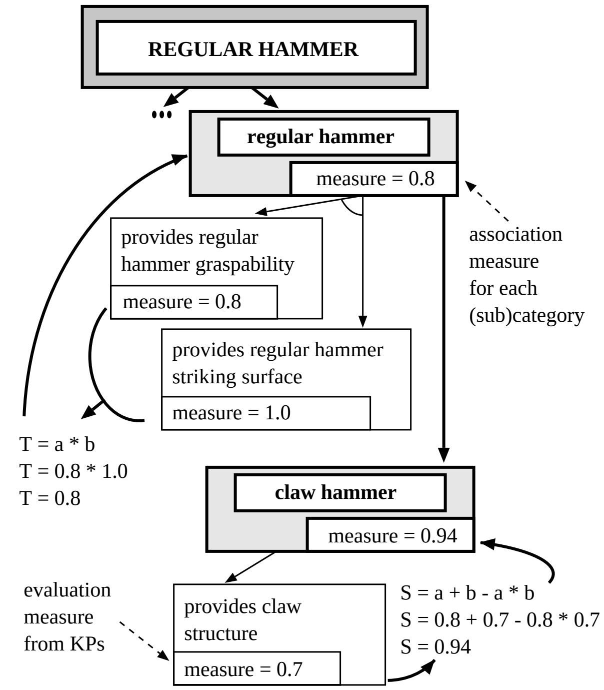

Figure 3: This figure shows a representative subset of the nodes within the category definition tree which organizes **GRUFF**'s knowledge of artifacts. Aggregation calculi are used to combine the *evaluation measures* of a series of individual primitive calls into a meaningful *association measure* representing the usefulness of the shape for a specific category.

3.3 Decision: Categorizing the Shape

GRUFF performs shape-based analysis and outputs two sources of information. The first is an association measure indicating the category with compatible functional requirements. With this measure, the system also supplies a list of functional elements and evaluation measures. These are specific portions of the object shape which have explicit functional uses consistent with the target category.

A total of 50 objects (displayed in Figures 1 and 4) were analyzed by the system for the categories screwdriver, wrench, and hammer.[2] The resulting association measures ranged from 0.415 to 1.0 for all the objects above the bottom row in Figure 1. Those examples in the bottom portion of this figure were determined by the system to be "dysfunctional" hand tools, unable to satisfy all of the requirements of any hand tool (sub)category. Further results are summarized in Figure 4.

[2] Object (a) is repeated in both figures.

Experimental Results for the Superordinate Category Hand Tools

Total # of Objects	50
# Intuitively Categorized as Hand Tools	42
# Categorized Correctly as Hand Tools by System	38
# Intuitively Categorized as Non-Hand Tools	8
# Categorized Correctly as Non-Hand Tools by System	8

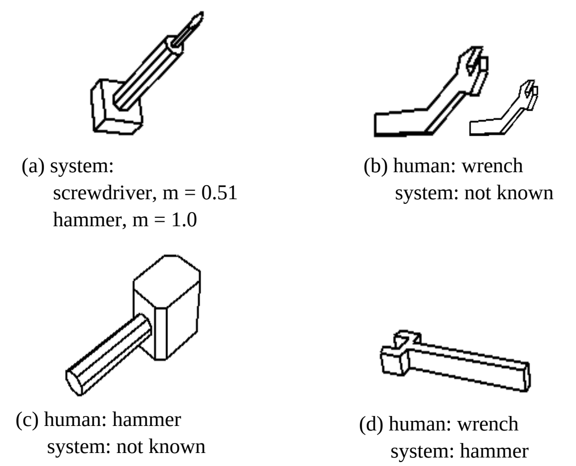

Figure 4: All final *association measures* above the system threshold of 0.4 are considered valid recognition results. As the examples indicate, (a) **GRUFF** may find multiple results for a single object, and (b)-(d) it is entirely possible that the designer of an object and the **GRUFF** system may disagree on intended functionalities.

4 IMPLEMENTATION DETAILS

The **GRUFF** system is written in C (executable program approximately 1.6 MB). The shape descriptions (defined in terms of faces and vertices) have average space requirements of about 10 KB per object. Total processing time on a SPARC station IPX is on the order of seconds to minutes (depending on the number of testable orientations or concavities). The models are available to interested researchers via anonymous ftp to figment.csee.usf.edu under the directory pub/errors_stuff/Objects. Further information can also be accessed via http://marathon.csee.usf.edu/.

5 SCALABILITY, COMPLEXITY, LIMITATIONS

The system we have presented is directly scalable to categories of manufactured objects whose functionality is mostly physical. The corresponding category definition tree grows slowly with the extension into new domains since the additional knowledge is based on functional descriptions for each new *category of objects*, rather than on defining new *object instance models*. For example, the **GRUFF** system executable grew from 1.4 MB to 1.6 MB

(200 KB of code added) when incorporating hand tools processing. With this extension we are able to process not only the 50 objects presented in this paper, but countless other hand tool-like objects which fall within the hand tool categories known to the system. All of this is in addition to the hundreds of object models that fall within the other domains previously known to the system (e.g., furniture and dishes).

An analysis of the complexity of each of the primitives and the overall system is presented in [15]. Our present implementation is based on polyhedral approximations of objects. Since we have been able to successfully recover such models from noisy LRF images [14], we do not plan to include curvature information in the system in the near future.

6 CONCLUSION

This version of **GRUFF** represents a system operating under what might be termed an *expectancy* paradigm. An initial hypothesis is formed for recognition using bottom-up visual and shape-based reasoning alone, expecting that material properties of an object are sufficient. However, the results of this preliminary analysis (yielding locations of important functional elements) can provide the guidance to instantiate further (more expensive) top-down exploratory modules to selectively attend to these areas.

Just as a diverse set of underlying mechanisms support human cognition and learning, we are designing **GRUFF** as a multi-stage recognition process encompassing overlapping and partially independent modules which operate on different types of information. Consequently, one current direction of research we are exploring is to interactively confirm functional *use* of such objects in a task employing robot arms. This will include further developing our approach to include registered LRF and intensity images, so that information about texture and color can be readily extracted and used. Of course, additional operators such as *apply force* or *observe deformation* will be necessary to determine if functional properties are still met. However, we expect these operators to also be applicable across multiple domains, allowing **GRUFF** to integrate additional sensory input with existing shape-based input for all presently known artifacts.

References

[1] L. Bogoni and R. Bajcsy. An active approach to characterization and recognition of functionality and functional properties. In *AAAI Workshop: Reasoning About Function*, pages 9–16, 1993.

[2] P.P. Bonissone and K.S. Decker. Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 217–247. North-Holland Publishing Company, 1985.

[3] M. Brady, P.E. Agre, D.J. Braunegg, and J.H. Connell. The mechanics mate. In T. O'Shea, editor, *Advances in Artificial Intelligence: European Conference of Artificial Intelligence*, pages 79–94. Elsevier Science Publishers, 1985.

[4] M. Brand. Vision systems that see in terms of function. In *Working Notes on Reasoning About Function*, pages 17–22, 1993.

[5] J.H. Connell and M. Brady. Generating and generalizing models of visual objects. *Artificial Intelligence*, 31:159–183, 1987.

[6] L. Dey, P.P. Das, and S. Chaudhury. Recognition and learning of unknown objects in a hierarchical knowledge-base. In *AAAI Fall Symposium Series - Technical Report FS-93-04*, pages 15–19, 1993.

[7] M. DiManzo, E. Trucco, F. Giunchiglia, and F. Ricci. FUR: Understanding FUnctional Reasoning. *International Journal of Intelligent Systems*, 4:431–457, 1989.

[8] J. Hodges. Naive mechanics–a computational model of device use and function in design improvisation. *IEEE Expert*, pages 14–27, February 1992.

[9] K. Kise, H. Hattori, T. Kitahashi, and K. Fukunaga. Representing and recognizing simple hand-tools based on their functions. In *Asian Conference on Computer Vision*, pages 656–659, November 1993.

[10] E. Rivlin, S.J. Dickinson, and A. Rosenfeld. Recognition by functional parts. In *Workshop on the Role of Functionality in Object Recognition*, Seattle, WA, 1994.

[11] E. Rosch, C.B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.

[12] S. A. Stansfield. Haptic perception with an articulated, sensate robot hand. *Robotica*, 10, 1992.

[13] L. Stark, L.O. Hall, and K.W. Bowyer. An investigation of methods of combining functional evidence for 3-D object recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(3):573–594, 1993.

[14] L. Stark, A. Hoover, D.B. Goldgof, and K.W. Bowyer. Function-based recognition from incomplete knowledge of shape. In *Proceedings of IEEE Workshop on Qualitative Vision*, pages 11–22, 1993.

[15] 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.

[16] L. Vaina and M. Jaulent. Object structure and action requirements: A compatibility model for functional recognition. *International Journal of Intelligent Systems*, 6:313–336, 1991.

[17] P. Winston, T. Binford, B. Katz, and M. Lowry. Learning physical descriptions from functional definitions, examples, and precedents. *AAAI*, pages 433–439, 1983.