CHAPTER 7

Object Recognition Through Reasoning About Functionality: A Survey of Related Work

Kevin Bowyer, Melanie Sutton, and Louise Stark

7.1 Recognition Based on Functionality

Minsky (1991) is one of several well-known researchers who have argued for the necessity of representing knowledge about functionality:

> ...it is not enough to classify items of information simply in terms of the features or structures of those items themselves. This is because we rarely use a representation in an intentional vacuum, but we always have goals – and two objects may seem similar for one purpose but different for another purpose. Consequently, we must also take into account the functional aspects of what we know, and therefore we must classify things (and ideas) according to what they can be used for, or which goals they can help us achieve. Two armchairs of identical shape may seem equally comfortable as objects for sitting in, but those same chairs may seem very different for other purposes, for example, if they differ much in weight, fragility, cost, or appearance. . . . In each functional context we need to represent particularly well the heuristic connections between each object's internal features and relationships, and the possible functions of those objects.

The early part of this quote contrasts the approach of representing (only) features or structure of objects with the approach of representing knowledge about how an object functions to achieve a goal. Particularly in computer vision, objects have traditionally been represented by their shape or their appearance. Object recognition based on reasoning about functionality stands in contrast to these more traditional approaches, with the aim of achieving recognition at a more generic level. The middle part of this quote illustrates the ideas in the context of the object category `chair`. Figure 7.1 depicts a variety of typical chairs. Chair seems to be a favorite object category for research in this area and is the example that we started within the Generic Recognition Using Form and Function (GRUFF) project. The last part of this quote highlights the point that there is some relationship between the features and structure of an object and how that object may be used to achieve a particular function. From the computer vision perspective, it should be possible to analyze the shape or appearance of an object to

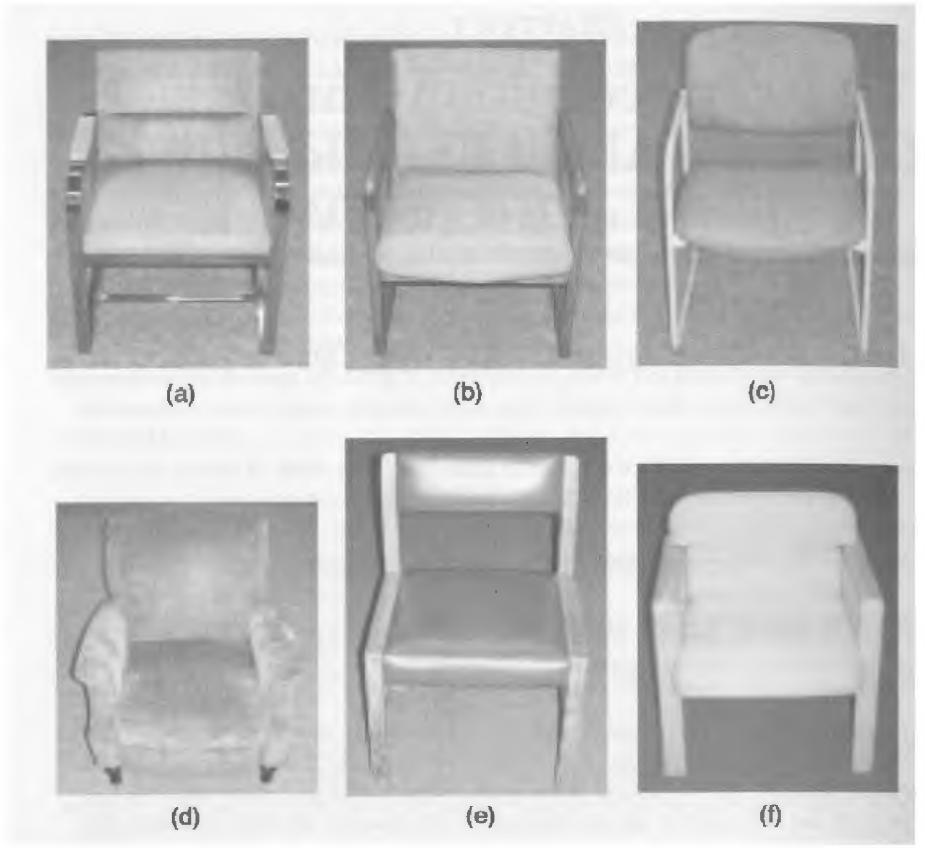

Figure 7.1. Typical chairs with varied shape and appearance. Object recognition based purely on representing shape and/or appearance would have great difficulty encompassing all of these objects in one model. The essence of a representation based on functionality is that each object would be recognized as a chair by reasoning that it has a sittable surface, back support, and arm support that could be used together.

find features and/or structures that make it possible to infer a potential function of the object.

The purpose of this chapter is to review work in this area over the last two decades. Our viewpoint is primarily that of a computer vision researcher wanting to enable object recognition at a generic, or category, level (see Rosch et al. 1976 for a discussion of object categories). However, sensing of shape and/or appearance is generally sufficient only to determine potential functionality. Actual functionality typically depends on material properties that are best determined by interacting with the object, and this level of reasoning about functionality involves both vision and robotics. Additionally, as object functionality becomes more complex, the representations used need to become correspondingly more systematic and powerful. Thus, object recognition based on functionality eventually incorporates elements of artificial intelligence, computer vision, and robotics. This chapter will touch on each of these areas, and we will attempt to show how they contribute to a greater whole.

The remainder of the chapter is organized as follows. Section 7.2 summarizes the development of the GRUFF approach pioneered by Stark and Bowyer (1991). Section 7.3 explores uses of function in artificial intelligence, computer vision, and robotics, as well as related research in cognitive psychology that impacts functionality-based object recognition research. Finally, Section 7.4 proposes research lines to be explored in the next generation of systems based on reasoning about the functionality of objects.

7.2 Development of the GRUFF Approach

The GRUFF project began in the late 1980s (Stark and Bowyer 1989), and work has continued now for approximately two decades. The initial work focused on demonstrating reasoning about the functionality of 3-D shapes as exemplars of the category `chair` (Stark and Bowyer 1990, 1991). This was quickly followed by generalizations to additional object categories whose exemplars could be described by rigid 3-D shapes (Stark and Bowyer 1994; Sutton et al. 1994) and by articulated 3-D shapes (Green et al. 1995). Another line of generalization moved away from analyzing ideal 3-D shape descriptions and into the segmentation and analysis of sensed range images of real objects (Stark et al. 1993; Hoover et al. 1995). Still another effort looked at using machine learning techniques to simplify the construction of functional representations (Woods et al. 1995). A later line of work, which still continues, began to explore robotic interaction with a sequence of sensed shape descriptions to confirm actual functionality and drive additional bottom-up processing (Stark et al. 1996; Sutton et al. 1998, 2002; Sutton and Stark 2008).

7.2.1 Reasoning about Static 3-D Shapes

Early versions of GRUFF took an ideal 3-D shape as the starting point for reasoning about functionality. The shape models were simple boundary representations of polyhedral objects, of the type common in "CAD-based vision" efforts of that time. These shape descriptions were idealized in that they did not have any of the noise, artifacts, occlusion, or other problems that occur with 3-D sensing of real scenes. Also, the types of objects considered were ones that could reasonably be described by a rigid 3-D shape; for example, chairs and tables, or cups and plates.

The first GRUFF system reasoned about 3-D shapes to determine whether they satisfied the functionality of the object category `chair` (Stark and Bowyer 1991). The definition of the simplest category of chair could be given as:

```
Conventional Chair
  = Provides Sittable Surface + Provides Stable Support
```

The properties named in the definition are called *knowledge primitives*. Each is implemented as a procedure call that takes the 3-D shape description as input, possibly with features identified by previous knowledge primitives, and performs computations to determine if the named property can be satisfied by the shape. For example, the knowledge primitive `Provides Sittable Surface` looks for planar surfaces,

or approximately co-planar groups of surfaces, that cover an area of the appropriate size to be a seating surface. The execution of the knowledge primitive provides a list of the candidate surfaces found on the shape. The knowledge primitive `Provides Stable Support` then takes these candidates and performs computation to determine if the object will rest stably on a support plane when the seating surface is oriented up and force is applied to it. The result is a list of surfaces on the object that can be oriented upward and serve as stable seating surfaces. This first GRUFF system had multiple categories of chair, including `straight-back chair`, `arm chair`, `balans chair`, `lounge chair`, and `high chair`. All of these were elaborations or variations on the same basic idea of using functional primitives to process a 3-D shape to determine how it might serve a specific function.

Experimental results were shown for the first GRUFF system processing 101 different 3-D shapes created as 38 intended chair exemplars and 63 intended non-chair exemplars. The system declared that 37 of the 38 intended chair exemplars satisfied the functionality of a chair, and that 46 of the 63 intended non-chair exemplars could not satisfy the functionality of a chair. The disagreement on the intended non-chair exemplars came from the system determining that in fact they could serve as a chair if used in some novel orientation. The disagreement on the intended chair exemplar came from the fact that the exemplar only marginally met the requirements of individual knowledge primitives and slipped below the threshold for recognition. The system did not actually give a binary chair/non-chair result, but it did provide a measure that reflected how well the aggregate functionality was satisfied, with a threshold for a chair/non-chair decision.

The basic ideas used in the initial version of GRUFF proved sufficient to represent a much broader range of object classes than just `chair`. The system was relatively easily extended to various categories of furniture objects (Stark and Bowyer 1994), and also to various categories of dishes and cups (Sutton et al. 1994). Common themes in all these efforts are that the functionality involved rigid shape, and the models processed by the system were boundary representations of polyhedral shapes.

7.2.2 Functionality Requiring Articulated Shape

One generalization of GRUFF dealt with reasoning about objects whose functionality is realized by articulated motion of 3-D parts (Green et al. 1995). The input to this problem is a time sequence of shape models with no information about parts or articulation. This sequence is assumed to encompass the full range of motion of the underlying articulated object. GRUFF first infers a 3-D model of parts and their ranges of articulated motion, based on the sequence of observed 3-D shapes. GRUFF then uses knowledge primitives to determine the functionality of the recovered articulated model. The object category explored in this work is `Scissors`, and the high-level definition of its functionality is given as:

```
Scissors = Provides Opposing Finger Grasp
         + Provides Opposing Cutting Blades
         + Closing Grasp Causes Cutting
```


Experimental evaluation was done with time sequences of shape instances from twenty-four objects, including various "near miss" type objects. The evaluation confirms that the GRUFF approach can model the functionality of simple hand tools with articulated parts.

7.2.3 Shape Models from Sensed Range Images

An important line of generalization of GRUFF was to reason about functionality starting with sensed range images of real objects (Stark et al. 1993). This involves constructing "object plus unseen space" shape models from range images (Hoover et al. 1995). These models allow reasoning about both the surfaces of the shape that have been seen and the limits to the parts that have not yet been seen. The need for better segmentation of sensed range images spawned a separate project that focused on how to evaluate and compare segmentations of range images objectively (Hoover et al. 1996). The theme of dealing with real sensor input that provides partial and imperfect information about object shape and appearance is one that still confronts researchers today.

7.2.4 Robotic Interaction with Objects

Reasoning about partial object shape highlighted the fact that complete recognition often requires additional images from a selected viewpoint and/or interaction with the object to confirm possible functionality (Sutton et al. 1998). This version of GRUFF added an interaction-based reasoning subsystem to the shape-based reasoning capabilities of previous versions. The shape-based reasoning about functionality could, for an object category such as `cup`, suggest an approach to grasping the object and an enclosed volume that should hold liquid. The interaction-based reasoning would guide an attempt to use the object as a cup, sensing the scene to observe the results at appropriate points. For instance, after pouring a substance into the cup, a visual check is made for leaks. This theme of interaction with the object to confirm functionality suggested by shape or appearance is another that still confronts researchers today.

Major conclusions from the interaction-based version of GRUFF 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; and (3) interaction-based reasoning can be used to confirm the functionality of a categorized object without explicitly determining the object's material composition (Sutton et al. 1998). One noted limitation of this approach in 1998 was its dependence on isolated objects in a scene as well as a single parameter set utilized to minimize the average residual error between the true range data and resulting values after segmentation and model-building to construct 3-D models. In 2002, the use of context-based reasoning was applied to deal with more complex scenes involving multiple objects (Sutton et al. 2002). In a later work, Sutton and Stark also proposed recovering the 20 to 30 percent loss of usable data or models in each subsystem (model-building, shape-based reasoning, and interaction-based reasoning) by exploring parameter set selection driven by functional analysis of initial extracted surfaces from coarse and then refined segmentations (Sutton and Stark 2008).

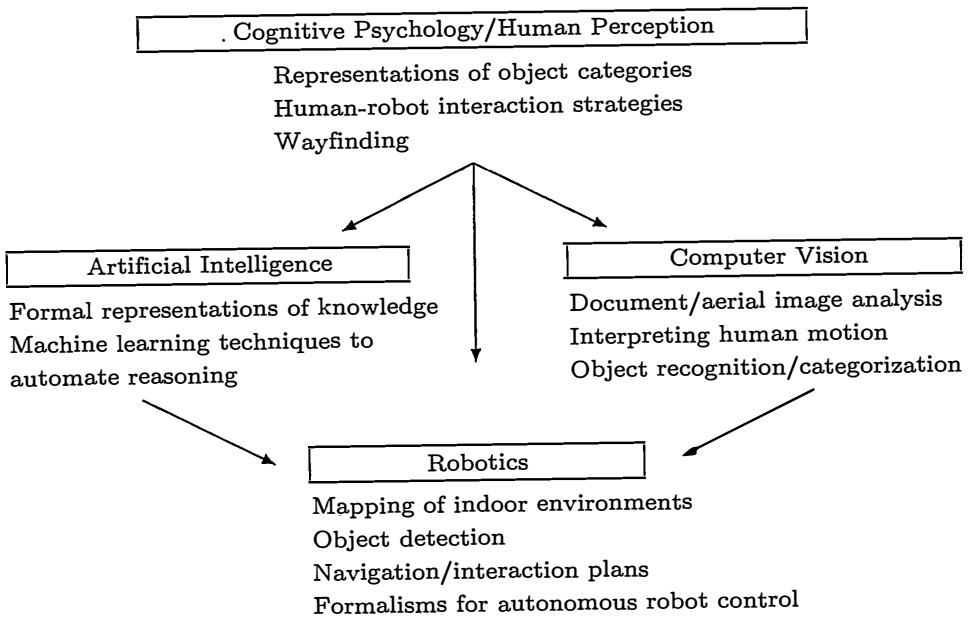

Figure 7.2. Research fields employing function-based approaches. Function-based approaches have been developed in various contexts across the disciplines, with robotics employing best practices from all fields as systems have increased in complexity.

7.3 Functionality in Related Disciplines

The following subsections summarize research on functionality within the fields of artificial intelligence (AI), computer vision, robotics, and other closely related disciplines. Figure 7.2 provides a brief overview of major research areas within each of these fields. Early in the development of functionality-based research, AI and computer vision often pulled directly from research in cognitive psychology and human perception. Examples include perception-inspired low-level image processing and representations for object categorization supporting generic object recognition. Robotics incorporated image analysis early on as part of its array of sensors, and benefited from advances in this area in early computer vision systems. As problem complexity has increased and software/hardware costs have decreased, multisensor robotics systems have continued to draw from computer vision, as well as lessons learned in related AI systems employing function-based approaches. In the following sections, we explore historical and recent trends in each of these areas and then summarize the open problems that remain.

7.3.1 Functionality-Related Work within AI

There are at least two different areas within AI that impact research on constructing systems that recognize objects by their functionality. One of these areas is the work on formal representations of knowledge about functionality. The other is the application

of machine learning techniques to automate the process of constructing systems that reason about object functionality.

As might be expected, research in AI concerned with reasoning about functionality of objects has developed greater formalism and depth than that in computer vision, and research in computer vision has been much more concerned with how to extract information about functionality from the sensed data of a scene. As efforts in robotics and vision tackle more complicated scenes and object functionalities, more complex representations will likely be needed.

Chandrasekaran (1994) reviews the development of his functional representation (FR) line of work, and the differences between it and related efforts in AI. Functional representation is an approach to describing the function, structure, and causal processes of a device. This allows for representing a more complicated object functionality than is present in simple objects such as chairs (Stark and Bowyer 1991), but that might be used by object categories such as tools with articulated parts (Green et al. 1995).

Hodges (1995) presents the functional ontology for naive mechanics (FONM) representation theory. The system takes advantage of causal relationships between the structure, behavior, and function of an object. Function is represented "in terms of its low-level structure and behavior and in terms of its use in problem-solving contexts" (Hodges 1995).

Chandrasekaran and Josephson (1997) make a useful distinction between three types of knowledge about functionality: "(i) what the device is intended to do (function), (ii) the causal process by which it does it, and (iii) how the process steps are enabled by the functions of the components of the device. . . . " They also make a distinction between the function of a device and the behavior of a device. This work is important for bringing out some of the complexity needed in formal systems for reasoning about device functionality.

Mukerjee and colleagues consider "conceptual descriptions" that use "linguistic" terms, as in the example "There is a hedge in front of the bench" (Mukerjee et al. (2000). The immediate context of the work is more in visualization of a conceptually described scene rather than recognition from image data. The idea of a continuum field is used as a means to represent the description given in linguistic terms. Experiments are performed with human subjects and placement of objects in a graphic scene in order to determine a preferred interpretation of linguistic terms. The connection of this work to reasoning about function is rather tenuous. In fact, Stark and Bowyer (1991) is cited as an example of "parameterized geometric models." However, the concepts in this work are potentially useful in reasoning about function, if the criteria for functionality are not crisply defined but instead involve what this work terms "linguistic" descriptions.

More recently, Gurevich et al. (2006) tackle the problem of learning how to recognize generic objects as seen in range images. They focus on how to generate negative examples automatically that are "near-misses" of positive examples of an object. Rules are introduced to transform a positive example into a near-miss negative example. This allows the user to, in effect, focus the learning algorithm on the important elements of an object definition. Experiments are performed with 200 range images of positive examples of each of the categories stool, chair, and fork. The experiments verify that the learning is more efficient with the near-miss negative examples than with general

negative examples. This work is important as an example of how to automate elements of the construction of a recognition system.

7.3.2 Functionality-Based Computer Vision

Computer vision researchers have pushed the concepts of functionality-driven object recognition in several directions. Perhaps the most surprising of these is the use of functionality in document image analysis. Another area that is very different from the GRUFF work is the recognition of roads and buildings in aerial images.

Doermann et al. (1998) apply the concept of functionality to document image analysis. Document functional organization is defined as "organization in information transfer terms" with the analogy "between components of a document, which is a device for transferring information, and the parts of a tool, which is a device for transferring force" (Doermann et al. 1998). Documents can be defined by their layout structure, as in how the information is organized and presented, and their logical structure, as in what the document is trying to convey (i.e., semantic or conceptual organization). The functional level is an intermediate level "that relates to the efficiency with which the document transfers its information to the reader." As an example, a block of text might have a geometry description that includes its size and position on the page. Based on the geometry of a particular block in relation to other text, the functionality of the block may be classified as that of a "header." A further categorization of "title" would be a semantic labeling of the block. In one experiment, documents are classified according to use, as a reading document, a browsing document, or a searching document, based on the number and size of text blocks. Also, a functionality-driven approach is outlined for classifying the type of a document as journal article, newspaper article, or magazine article, and the pages of a journal article as title, body, or references. Beyond its contribution in the area of document analysis, this paper is interesting for how it shows that concepts of functionality can be applied to object classes very different from those studied in earlier research on functionality.

Mayer (1999) surveys work in the area of extracting buildings from aerial images and proposes a model and approach that incorporates elements of functionality. In this model, "the general parts of the model are: characteristic properties [that] are often the consequence of the function of objects. Importantly, they integrate knowledge about the 3D real world into the model. Typical examples for knowledge sources are, apart from constraints concerning the usefulness for humans (Stark and Bowyer 1991), construction instructions for different types of buildings and roads. For large parts of the knowledge about function it seems to be enough to take them into consideration for modeling . . . ".

Baumgartner et al. (1997) consider the perspective of functionality in developing a system to extract roads automatically from aerial images – "In the real world, the characteristics of roads can be described as consequences of their function for human beings." Texture analysis is used to describe context regions such as open-rural, forest, and suburban. Relations between background objects, such as buildings or trees, and parts of roads may vary with the type of context region. Image analysis for edges is done at both a coarse and a fine resolution. Road parts are aggregated into a road network.

Mirmehdi et al. (1999) outline a feedback-control strategy for processing images to recognize generic objects based on functionality. The essential idea is that initial processing would be computationally cheap and fast, serving to generate candidates that would then be explored further and, finally, leading to confirmed instances of objects. The initial, or lower-level, processing of the image is "very closely linked to the modeling and representation of the target objects." In the more interesting examples in this paper, which focus on detecting bridges in infrared images of large-scale outdoor scenes, the reasoning is based on edges detected in the image. The generic model of bridges used for interpretation is functionality-based "unlike normal model-based object recognition systems in which the knowledge of an object's appearance is provided by an explicit model of its shape, we adopt a signature and a stereotype of the object as image and model representations, respectively, with both being very loose and generic descriptions of the object based on their functionality after Stark and Bowyer (1991)." Examples are shown of various bridges that can be recognized in terms of sets of edges representing the roadway and the support columns.

Zerroug and Medioni (1995) and Medioni and Francois (2000) propose to approach the problem of generic object recognition by extracting 3-D volumetric primitives from 2-D images. The 3-D volumetric primitives are particular restricted subclasses of generalized cylinders. Objects are represented as a composition of volumetric primitives. Segmentation is proposed to use perceptual grouping and quasi-invariant properties of primitives/regions. This is a modern proposal related to older work by Marr (1982), Bierderman (1987), and others. This approach is related to reasoning about functionality in the sense that it is aimed at extracting generalized object models from images, which could potentially be followed by or accompanied by reasoning about functionality.

Peursum et al. (2003, 2004) analyze the pattern of activity in a video of a person interacting with a scene, in order to label parts of the scene according to their functionality. The emphasis on observing the pattern of human activity rather than reasoning directly about an object distinguishes their approach from that of Stark and Bowyer (1991): "Our premise is that interpreting human motion is much easier than recognizing arbitrary objects because the human body has constraints on its motion" (Peursum et al. 2004). An "interaction signature" between the person and a part of the scene is what characterizes a particular type of functionality. In experiments, pixels in an image are classified as representing an instance of either chair or floor in the scene. The accuracy of characterizing parts of the scene as floor is substantially higher than the accuracy for chair. In some ways, this work is like a realization of ideas explored in simulation in Stark et al. (1996), but with less emphasis on reasoning about object shape, and more of an emphasis on reasoning about human activity patterns in video.

In a similar vein, work by Duric and colleagues (1996) addresses the problem: "Given a model of an object, how can we use the motion of the object, while it is being used to perform a task, to determine its function?" The authors note that motion and form are required. Motion analysis is used to produce motion descriptors. A known set of motion-to-function mappings in the form of motion descriptors can be used to identify the action (function) taking place in the image sequence.

Finally, Wunstel and Moratz (2004) report on results of generic object recognition from range images using a functionality-based approach. They are interested in an

office-like context that contains instances of objects such as chair, door, table, and other such objects. They aim to deal with more general object shapes than were experimented with in GRUFF (Stark and Bowyer 1996). "Stark's original form and function algorithm [as described in section 7.2.1] contains a segmentation work step within the 21/2-D data, which is finally based on geometric limitations of the allowed objects. They only used objects composed of cuboids with sharp corners. In our environment, scenario freeform objects (preliminarily restricted to tables and chairs) have to be identified. A three-dimensional segmentation based approach is not suitable or necessary as we neither have nor need fully defined three dimensional object model descriptions" (Wunstel and Moratz 2004). By taking slices through the sensed 3-D model of the scene at particular heights and orientations, and processing these as 2-D images, they are able to recognize objects without creating an explicit segmentation into 3-D shapes or models. Example results are presented for an office scene. Froimovich et al. (2007) similarly have applied function-based object classification to raw range images. Data from 150 images from object instances from 10 categories are provided in this work, with examples provided for dealing with cluttered environments.

7.3.3 Reasoning about Functionality in Robotics

As highlighted in Figure 7.2, the evolution of robotics-based systems that employ function-based approaches has been buoyed by incorporating best practices gleaned from other fields. In the following sections we examine parts of this evolution, from research in service robots and generalized navigational systems to function-driven manipulation in various environments.

Moratz and Tenbrink (2007) define affordances as the "visually perceivable functional object aspects shared by the designer of the recognition module and the prospective robot user or instructor." Their model for an affordance-based recognition system is based on the need to build service robots that deal more effectively with coarse, underspecified knowledge. They propose that their system is more scalable than traditional systems, even when the robot encounters occlusion, sensor error, or detection difficulties, because their linguistic module will allow nonexpert users to be able to specify instructions relative to previously recognized function-based objects.

Kim and Nevatia (1994, 1998) use a functionality-driven approach in having a mobile robot map out an indoor environment. They consider a functionality-driven approach because of its inherent ability to handle a broad variety of instances of a class of objects: "The most generic representation of an object is probably in terms of its functionality." Their system creates an "s-map" of an indoor environment, that marks the presence of objects such as doors and desks. Object categories such as these are defined in terms of surfaces that are important to their functionality, possibly with multiple surfaces at different levels of importance to the functionality. The implemented system uses planar surfaces in its modeling of functionality, and assumes that objects are encountered in a standard or typical pose. Experimental results show the ability to recognize substantially different instances of the object category "desk." This work illustrates how a mobile robot might recognize objects in its environment in terms of categories of relevant functionality.

Wixson (1992) studies the problem of having a mobile robot with a camera "head" explore a defined space for the presence of particular types of objects. In this context, he notes the limitations of simpler object modeling techniques and the potential advantage of a functionality-based approach. "Unfortunately, the great majority of objects that we might consider to be intermediate objects, such as desks, countertops, sofas, chairs, and bookshelves are ill-suited for recognition by traditional machine-vision algorithms. As noted in (Strat and Fischler 1990) and (Stark and Bowyer 1991) such algorithms assume that the object to be recognized is either: (a) definable by an explicit geometric model, or (b) has characteristic homogeneous and locally measurable features such as color or texture.... These assumptions are not valid for most of the large-scale 'generic' objects just mentioned, although we do believe that it should be possible to extend methods related to assumption 2. Further research must be performed on recognizing large-scale generic objects" Wixson (1992). This report does not present results of any object recognition or classification experiments.

Rivlin et al. (1995) blend elements of a GRUFF-like approach with part-whole definitions of what constitutes an object. "Comparing this approach to that of Stark and Bowyer for searching the image for a 'chair kind of support,' we would like to reason about a set of chair legs, a seat and a back, rather than a set of simple planar surfaces or 3D points." They develop the notion of "functional parts" and model simple functional parts by superquadrics. This approach to recognition is illustrated with a range image of a scene that has a mallet lying on a table. The range image can be segmented into regions, superquads fit to the regions, and the superquadric models then interpreted as functional parts. It is acknowledged that this approach is "appropriate for objects composed of simple volumetric parts" and that "we support only functionality that is defined in terms of an object's shape."

Rivlin and Rosenfeld (1995) discuss functionalities that objects in the environment can have for a navigating agent, or mobile robot. In the context of a hallway-cleaning task, objects in the environment of the mobile robot function as objects to be intercepted, objects to be avoided, or objects to be used as landmarks. These are very general categories of object functionality, defined relative to a particular task. Rivlin and Rosenfeld outline the image-processing operations needed for a mobile robot to categorize objects into these categories.

Cooper et al. (1995) look at knowledge representation meant to support robotic interaction with a scene: "What knowledge about a scene does a robot need in order to take intelligent action in that scene?" They seek a representation that embeds a causal understanding of the scene, because "understanding the world in causal terms is what permits intelligent interaction with that scene." Causal representations of scenes is based on naive or qualitative physics. One instance of the approach outlined by Cooper et al. (1995) generates causal explanations of stacks of blocks in an image, given a starting point and assuming blocks are represented as rectangular regions in the image. Another version generates explanations of a stereo pair of images of a scene of link-and-junction (tinker-toy) objects, including inference of extensions of occluded portions of links. A third instance of this approach is aimed at picking up coffee mugs with a robotic gripper. One important contribution of this work is showing how a causal explanation, or functional plan, of a scene can be generated based on image analysis and a set of qualitative physics rules.

Bogoni and Bajcsy (1995) make distinctions between intended, imposed, and intrinsic functionality, and construct a representation of functionality that combines elements of object structure, the context of the actor involved in the context of the function, and the application of the function. The formalism used for representing functionality in this work is based on discrete event systems. Conceptually, this work differs from that of Stark and Bowyer (1991) in that it emphasizes a representation of functionality that goes beyond reasoning about shape alone. It is more similar to that of Stark and colleagues (1996), except that Bogoni and Bajcsy's experimental work involves the use of robotic manipulators and vision. The experimental context is the general hand tools category of objects, and particular experiments involve the functionality of an object being used as a tool to pierce a material. This work is one of a small number to blend robotics and vision in its experiments.

Further work by Bogoni (1998) defines a representation of functionality that includes a "task description, shape description, force-shape maps (generated by interacting with the object), and histograms defining the behavior of the system with different materials and functionalities." Piercing and chopping are the two actions tested. The work "proposes a representation and a methodology for recovering functionality of objects for a robotic agent." Therefore, it is an extended version of the functional representation as defined by many others.

Wang et al. (2005) develop a generic ontology of objects. They focus on the context of an indoor environment that contains manufactured objects, and how a robot might interact with such objects. They see the GRUFF functionality-based approach as useful but as needing to be combined with geometric object models: "The Generic Recognition Using Form and Function GRUFF system Stark and Bowyer (1991) performs generic object recognition and context-based scene understanding. Objects are decomposed into functions based on the object's intended usage, where each function describes a minimum set of structural elements, and their geometric properties and relations. . . . GRUFF has instantiated function-based knowledge for everyday objects like furniture, kitchenware, and hand tools, and everyday scenes like office. However, it does not define any formal representation of generic shapes." They develop an ontology that defines functionality in terms of more specific shape and structural information than does GRUFF, and add information about associations between objects and typical interactions with objects. This paper does not present results of any object recognition or classification experiments.

Paletta et al. (2007) propose a multilayered framework for the developmental learning of affordances. This work includes real-world robot experiments to provide a proof of concept, with experiments demonstrating the ability of the system to successfully learn affordance-based cues. In a related line of work, Fritz et al. (2006) provide results from simulated scenarios utilizing this framework. With the thoroughness these researchers have demonstrated in laying the foundation for testing their framework in simulated and real-world scenarios, this line of research will be one to follow closely as new robot control architectures evolve in this area. Also contributing to emerging trends in this area will be new formalisms for affordances to facilitate autonomous robot control, such as those proposed by Sahin et al. (2007). This work is also supported by recent real-world experimental results to examine the strengths and limitations of the proposed formalisms and propose new directions. The "affordances" concepts used here come from the work of Gibson (1979). Gibson's theory of affordances is consistent with function-based vision. An object *affords* support by being flat, horizontal, and positioned at the proper height for sitting, all attributes that can be confirmed visually.

7.3.4 Other Disciplines: Reasoning about Human Perception and Reasoning about Function

Finally, this chapter would not be complete without consideration of related research in human perception and recent implementations applying associated theories. Klatsky et al. (2005) perform experiments to investigate how children reason about whether an object can perform a given function. This area of work in psychology may help to inform the design of algorithms for robotic interaction with objects to reason about their function.

Recent work in the theory of how functional knowledge is represented and processed also includes the HIPE (history, intentional perspective, physical environment, and event sequences) theory of function developed by Barsalou et al. (2005). Although this multimodal system overall may be more complex in terms of the multiple representations of function that are encoded, efficiency is expected to be realized for given events whenever an agent's intentional perspective can be used to determine the functional knowledge that is retrieved.

Researchers such as Raubal and Moratz have expanded on this theory in their redesign of a robot architecture that incorporates the representation of affordance-based attributes associated with tasks (Raubal and Moratz 2007). Raubal and Moratz argue that a robot architecture that can support sensing and action based on functional compounds (rather than properties in isolation) can be more flexibly mapped to human affordance-driven tasks. Moratz (2006) has similarly argued for a function-based representation for tasks to make human-robot interaction and communication more natural and simplified as well as scalable to novel environments.

Along with Worboys, Raubal has also argued for the inclusion of affordances in models for wayfinding to more completely capture the processes of learning and problem-solving in these tasks (Raubal and Worboys 1999). Wayfinding is defined as "the purposeful, directed, and motivated movement from an origin to a specific distant destination that cannot be directly perceived by the traveler" (Raubal 2007). Emphasizing the relevance of ecological psychology (focusing on the information transactions between living systems and their environments), Raubal incorporates three realms of affordances in his model: physical, social-institutional, and mental (Raubal 2001).

7.4 Open Problem Areas

This survey has highlighted both historical and current efforts in using functionality for generic object recognition. Across the subfields of AI, computer vision, and robotics, it is clear that as systems scale up to more complex scenes, reasoning that is functionality-based (as opposed to category-based) holds great promise. The most successful systems described in this chapter were designed with architectures and

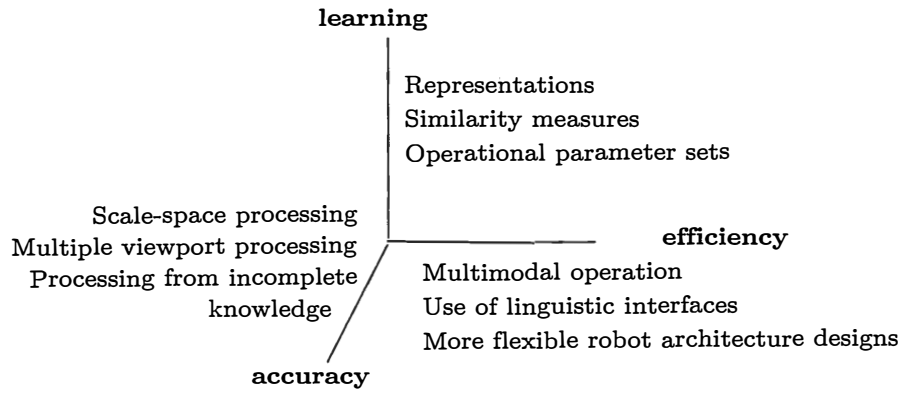

Figure 7.3. Open problem areas impacting future scalability of function-based reasoning approaches in object recognition.

representations supporting multiple sensors and object domains. However, much research is needed to ensure that the systems continue to evolve in ways that demonstrate continued scalability, efficiency, accuracy, and ability to learn, as summarized in Figure 7.3. As we look to the future of this field, in the following sections we examine the strengths and open research areas of GRUFF and a subset of representative systems that address each of these characteristics.

7.4.1 Successes and Limitations of GRUFF

The primary successes of the GRUFF project are as follows: (1) successful functionality-driven recognition of 3-D models of rigid shapes, (2) generalization of the approach to a variety of object categories involving both rigid and articulated shapes, (3) demonstration of functionality-driven analysis of partial object shapes derived from real range images of real objects, and (4) development of an approach for using robotic interaction to confirm actual functionality suggested by interpretation of 3-D shape.

One major limitation of GRUFF is that the shape descriptions that it analyzed were limited to polyhedra. Generalization to handle some broad class of curved shapes would be an important practical advance. This generalization presents some challenges in the implementation of the shape-based reasoning, but it probalily presents even greater challenges in segmenting range images into curved surface patches. Another limitation is that each generalization of GRUFF is essentially hand-crafted. Although some initial work has been done to explore the use of machine learning to speed the implementation of new category definitions, this is an important open area with great future potential.

7.4.2 Scalability and Efficiency

As noted by Raubal and Moratz (2007), the formal incorporation of affordance-based attributes in robot architectures is seen as an approach that will more readily and flexibly map to specified human affordance-based tasks that a robot is given to accomplish. This approach incorporates a hierarchical definition of affordances that is also

scale-dependent. Control mechanisms that act at the correct scale for a given task is one open problem in this research as well as in the GRUFF (Stark and Bowyer 1996; Sutton and Stark 2008) work.

Moratz has also noted the impact of perceptual granularity on system performance and observed that additional testing is required to resolve ambiguities (Moratz 2006). However, this is clearly the direction in which service robot scenarios are likely to be headed in order to increase flexibility and reduce training times compared to traditional methods of humans, such as forcing the specification of exact metric locations or use of predefined class IDs that the robot can understand for a subset of predefined tasks (Moratz and Tenbrink 2007). In addition, Moratz and Tenbrink (2007) note that scalable systems will allow multiple modes of spatiotemporal operation, including synchronous and asynchronous instruction that can later lead to problem solving and even learning in novel scenarios. Although their current system is synchronous in nature, the associated representation and model for linguistic interface is to be designed with this level of extensibility to intuitive instruction in mind.

7.4.3 Scalability and Accuracy

The incorporation of physical, social-institutional, and mental affordances in formal models of wayfinding is one approach in this task to create agents that behave like humans; more direct comparisons in this line of work should prove useful as these systems evolve to determine if this improves accuracy (Raubal 2001). The wayfinding model proposed by Raubal (2007) by default involves decisions impacted by multiple viewpoints and scale-space, because the goal cannot be perceived from a single viewpoint. As this model is applied to more complex tasks, it will be worth examining how readily the architecture handles these forms of data, and how accuracy is impacted by such incomplete knowledge.

7.4.4 Scalability and Learning

The systems just reviewed include sensors ranging from structured light scanners to laser range finders and stereo vision systems. Although stereo systems have some limitations for detecting planar surfaces owing to their lack of scene texture, all three systems will benefit from research on "learning" parameter sets for optimal navigation. Future work in this area includes automatically selecting the parameter sets based on whether the current task requires local (fine) or global (rough) details (Sutton and Stark 2008).

The design of similarity measurements for searches that yield multiple affordances is also a critical area, especially when the end goal is for the system to learn affordances from new environments. It will be worth noting future extensions to recent work with Janowicz, in which Raubal explores context-aware affordance-based similarity measures (Janowicz and Raubal 2007). Clearly future representation systems incorporating function must plan for how affordances can be compared. Janowicz and Raubal support a platform that incorporates weighting based on social-institutional constraints as well as outcomes, with control mechanisms that permit weights to be automatically determined or used as exclusion factors, depending on the task. This approach is important because the types of tasks encountered vary, and systems

need to incorporate similarity-based reasoning and planning to "learn" from each task.

As a further example, Maloof et al. (1996) describe a functionality-driven approach to creating a system to recognize blasting caps in x-ray images: "Although blasting caps are manufactured objects, there is enough variability in their manufacture that makes a CAD-based recognition system impractical. What is common to all blasting caps, however, is their functionality. Ultimately, blasting caps are defined by their functionality, not by their shapes." An interesting element of this work is that the AQ15c learning system was used to learn concepts of blasting caps and nonblasting caps from training examples. Although this is a fairly restricted object category to define from its appearance in x-ray images, it was still felt that higher-resolution images that show additional detail would be needed. Nevertheless, this work is important because it illustrates a relatively understudied topic. Very little work has been done on how to learn elements of the representation automatically from examples (Woods et al. 1995). Most research on functionality-driven object recognition has used hand-tailored representations of functionality. The ability to learn, or induce, a functionality-driven representation from a set of labeled examples would also seem to be an important topic for future research.

7.5 Conclusion

The preceding sections provide a window into some of the open areas that remain in function-based reasoning. Two decades ago, systems were limited by hardware and software constraints, but today the impacts of these choices are relatively insignificant. More significant are decisions that must be made early regarding the design of the overall system architecture to ensure scalability for exploring novel environments or completing novel tasks, supporting or deriving multiple function-based interpretations from a single viewpoint, combining function-based details from multiple viewpoints, or learning function-based characteristics from encountered objects or completed tasks. As researchers across the subfields of AI, computer vision, and robotics continue to blend their knowledge and incorporate related progress made in fields such as cognitive psychology, as well as biology and engineering, the systems arising from these efforts in future decades will benefit greatly. New systems evolving from these efforts should also prove to be even more robust, with potentially greater applicability and relevance to human-computer tasks and problem solving.

Bibliography

Barsalou L, Sloman S, Chaigneau S. 2005. The hipe theory of function. In *Representing functional features for language and space: insights from perception, categorization and development*, ed. L Carlson and E van der Zee, 131–147. New York: Oxford University Press.

Baumgartner A, Steger C, Mayer H, Eckstein W. 1997. Multi-resolution, semantic objects and context for road extraction. In *Semantic modeling for the acquisiton of topographic information from images and maps*, 140–156. Basel: Birkhauser Verlag.

Bierderman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94(2):115–147.

Bogoni L. 1998. More than just shape. *Artif Intell Engineering* 12(4):337–354.

Bogoni L, Bajcsy R. 1995. Interactive recognition and representation of functionality. *Comput Vis Image Und* 62(2):194–214.

Chandrasekaran B. 1994. Functional representation: a brief historical perspective. *Appl Artif Intell* 8(2):173–197.

Chandrasekaran B, Josephson JR. 1997. Representing function as effect. In Proceedings of functional modeling workshop, ed. M. Modarres.

Cooper PR, Birnbaum LA, Brand ME. 1995. Causal scene understanding. *Comput Vis Image Und* 62(2):215–231.

Doermann D, Rivlin E, Rosenfeld A. 1998. The function of documents. *Image Vision Comput* 16(11):799–814.

Duric Z, Fayman JA, Rivlin E. 1996. Function from motion. *IEEE Trans Pattern Anal Mach Intell* 18(6):579–591.

Fritz G, Paletta L, Kumar M, Dorffner G, Breithaupt R, Rome E. 2006. Visual learning of affordance based cues. In Proceedings of the 9th international conference on the simulation of adaptive behavior, SAB 2006, LNAI 4095, 52–64. Berlin: Springer-Verlag.

Froimovich G, Rivlin E, Shimshoni I, Soldea O. 2007. Efficient search and verification for function based classification from real range images. *Comput Vis Image Und* 105:200–217.

Gibson J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Green K, Eggert D, Stark L, Bowyer KW. 1995. Generic recognition of articulated objects through reasoning about potential function. *Comput Vis Image Und* 62(2):177–193.

Gurevich N, Markovitch S, Rivlin E. 2006. Active learning with near misses. In Proceedings of the twenty-first national conference on artificial intellgence.

Hodges J. 1995. Functional and physical object characteristics and object recognition in improvisation. *Comput Vision Image Und* 62(2):147–163.

Hoover AW, Goldgof DB, Bowyer KW. 1995. Extracting a valid boundary representation from a segmented range image. *IEEE Trans Pattern Anal Mach Intell* 17(9):920–924.

Hoover AW, Jean-Baptiste G, Jiang X, Flynn P, Bunke H, Goldgof D, Bowyer KW, Eggert D, Fitzgibbon A, Fisher R. 1996. An experimental comparison of range image segmentation algorithms. *IEEE Trans Pattern Anal Mach Intell* 18(7):673–689.

Janowicz K, Raubal M. 2007. Affordance-based similarity measurement for entity types. In Lecture notes in computer science: spatial information theory – conference on spatial information theory (COSIT'07), ed. S Winter, M Duckham, L Kulik, and B Kuipers. New York: Springer-Verlag.

Kim D, Nevatia R. 1994. A method for recognition and localization of generic objects for indoor navigation. In Proceedings of the DARPA image understanding workshop, II:1069–1076.

Kim D, Nevatia R. 1998. Recognition and localization of generic objects for indoor navigation using functionality. *Image Vision Comput* 16 (11):729–743.

Klatsky RL, Lederman SH, Mankinen JM. 2005. Visual and haptic exploratory procedures in children's judgments about tool function. *Infant Behav Develop* 28:240–249.

Maloof M, Duric Z, Michalski R, Rosenfeld A. 1996. Recognizing blasting caps in x-ray images. In Proceedings of the DARPA image understanding workshop, 1257–1261.

Marr D. 1982. *Vision: a computational investigation into the human representation and processing of visual information*. New York: Holt and Company.

Mayer H. 1999. Automatic object extraction from aerial imagery – a survey focusing on buildings. *Comput Vis Image Und* 74(2):132–149.

Medioni GG, Francois AR. 2000. 3-d structures for generic object recognition. In International conference on pattern recognition, 1030–1037.

Minsky M. 1991. Logical vs. analogical, or symbolic vs. connectionist, or neat vs. scruffy. *AI Magazine* 12(2):34–51.
Mirmehdi M, Palmer PL, Kittler J, Dabis H. 1999. Feedback control strategies for object recognition. *IEEE Trans Image Proc* 8(8):1084–1101.
Moratz R. 2006. Intuitive linguistic joint object reference in human-robot interaction. In Proceedings of the twenty-first national conference on artificial intelligence (AAAI), 1483–1488.
Moratz R, Tenbrink T. 2007. Affordance-based human-robot interaction. In Proceedings of the Dagstuhl seminar 06231 "towards affordance-based robot control," eds. E Rome and J Hertzberg.
Mukerjee A, Gupta K, Nautiyal S et al. 2000. Conceptual description of visual scenes from linguistic models. *Image Vision Comput* 18(2):173–187.
Paletta L, Fritz G, Kintzler F, Irran J, Dorffner G. 2007. Learning to perceive affordances in a framework of developmental embodied cognition. In Proceedings of the 6th IEEE international conference on development and learning (ICDL), 2007.
Peursum P, Venkatesh S, West GA, Bui HH. 2003. Object labelling from human action recognition. In Proceedings of the first IEEE international conference on pervasive computing and communications.
Peursum P, Venkatesh S, West GAW, Bui HH. 2004. Using interaction signatures to find and label chairs and floors. *Pervasive Comput* 3(4):58–65.
Raubal M. 2001. Ontology and epistemology for agent-based wayfinding simulation. *Int J Geog Inf Sci* 15(7):653–665.
Raubal M. 2007. Wayfinding: affordances and agent simulation. In Encyclopedia of GIS, ed. S Shekhar and H Xiong. New York: Springer-Verlag.
Raubal M, Moratz R. 2007. A functional model for affordance-based agents. In Lecture notes in artificial intelligence: affordance-based robot control, ed. J Hertzberg and E Rome. Berlin: Springer-Verlag.
Raubal M, Worboys M. 1999. A formal model of the process of wayfinding in built environments. In Spatial information theory – cognitive and computational foundations of geographic information science. International conference (COSIT), Stade, Germany, 381–399.
Rivlin E, Dickinson SJ, Rosenfeld A. 1995. Recognition by functional parts. *Comput Vis Image Und* 62(2):164–176.
Rivlin E, Rosenfeld A. 1995. Navigational functionalities. *Comput Vis Image Und* 62(2):232–244.
Rosch E, Mervis C, Gray W, Johnson D, Boyes-Braem P. 1976. Basic objects in natural categories. *Cogn Psychol* 8:382–439.
Sahin E, Cakmak M, Dogar M, Ugur E, Ucoluk G. 2007. To afford or not to afford: a new formalization of affordances towards affordance-based robot control. *Adapt Behav* 15(4).
Stark L, Bowyer K. 1996. *Generic object recognition using form and function*. New York: World Scientific.
Stark L, Bowyer KW. 1989. Functional description as a a knowledge representation of 3-d objects. In IASTED international symposium on expert systems theory and applications, 49–54.
Stark L, Bowyer KW. July 1990. Achieving generalized object recognition through reasoning about association of function to structure. In AAAI-90 workshop on qualitative vision, 137–141.
Stark L, Bowyer KW. 1991. Achieving generalized object recognition through reasoning about association of function to structure. *IEEE Trans Pattern Anal Mach Intell* 13(10):1097–1104.
Stark L, Bowyer KW. 1994. Function-based generic recognition for multiple object categories. *CVGIP: Image Understanding* 59(1):1–21.
Stark L, Bowyer KW, Hoover AW, Goldgof DB. 1996. Recognizing object function through reasoning about partial shape descriptions and dynamic physical properties. *Proc IEEE* 84(11):1640–1656.
Stark L, Hoover AW, Goldgof DB, Bowyer KW. July 1993. Function-based object recognition from incomplete knowledge of object shape. In AAAI workshop on reasoning about function, July, 141–148.

Strat TM, Fischler MA. 1990. A context-based recognition system for natural scenes and complex domains. In Proceedings of the DARPA image understanding workshop, September, 456–472.
Sutton M, Stark L, Hughes K. 2002. Exploiting context in function-based reasoning. In Lecture notes in computer science, ed. G Hager, H Christensen, H Bunke, and R Klein, 357–373, Springer.
Sutton MA, Stark L. 2008. Function-based reasoning for goal-oriented image segmentation. In Affordance-based robot control, lectures notes in artificial intelligence 4760, ed. E Rome, 159–172. Berlin: Springer-Verlag.
Sutton MA, Stark L, Bowyer KW. 1994. Gruff-3: generalizing the domain of a function-based recognition system. *Pattern Recog* 27(12):1743–1766.
Sutton MA, Stark L, Bowyer KW. 1998. Function from visual analysis and physical interaction: a methodology for recognition of generic classes of objects. *Image Vision Comput* 16(11):745–764.
Wang E, Kim YS, Kim SA. 2005. An object ontology using form-function reasoning to support robot context understanding. *Comput Aided Des Appl* 2(6):815–824.
Wixson LE. 1992. Exploiting world structure to efficiently search for objects. Technical Report 434, Computer Science Department, University of Rochester.
Woods KS, Cook D, Hall L, Stark L, Bowyer KW. 1995. Learning membership functions in a function-based object recognition system. *J Artif Intell Res* 3:187–222.
Wunstel M, Moratz R. 2004. Automatic object recognition within an office environment. In Proceedings of the first canadian conference on computer and robot vision (CRV), IEEE.
Zerroug M, Medioni G. 1995. The challenge of generic object recognition. In Object representation in computer vision: international NSF-ARPA workshop, 217–232. Springer-Verlag LNCS 994.