Combining Visual Learning with a Generic Cognitive Model for Appliance Representation

Kanishka Ganguly, Konstantinos Zampogiannis, Cornelia Fermüller and Yiannis Aloimonos
Institute for Advanced Computer Studies, University of Maryland
College Park, MD 20740
Email: {*kganguly,kzampog,fer,yiannis*}@*umiacs.umd.edu*

***Abstract*—For robots to become ubiquitous in every home, it is imperative for them to be able to safely and intuitively interact with objects that humans use on a daily basis. To allow for such operation, we propose a cognitive model that will allow robots to recognize and work with common appliances, such as microwaves, refrigerators, dishwashers, and other similar equipment, found in everyday scenarios.**

I. MOTIVATION

Robots are rapidly becoming a part of our everyday life and they will need to intelligently interact with complex, unknown and changing environments. Automatic modeling of various aspects of the environment based on visual perception is a fundamental challenge to be addressed. We humans are extremely adept at manipulating equipment with very little experience and are able to generalize to other similar equipments easily.

There has been prior work done on visual detection of appliances [1], [2] which use visual sensors such as cameras or barcode scanners to recognize features on the appliances and then matching them to a pre-populated database to identify the appliance and how it is to be operated. These approaches lack generality, since every appliance will have its own unique features and it is possible that such matches might not exist in their proposed database.

II. GENERIC COGNITIVE MODEL

We propose a generic cognitive model for an appliance-agnostic visual learning procedure that will allow the system to identify, understand and ultimately operate an appliance, such as a refrigerators or a microwave oven. The cognitive model approach allows a system to describe an appliance at a high level of abstraction, focusing on a hierarchical definition of the appliance under observation, and provides a general interface for describing all the possible interactions with the appliance. This approach has the additional benefit of allowing development of modular and generic software packages that can be utilized by any robotic system for performing similar tasks.

A. Cognitive Model Description

Our proposed cognitive model is organized as a hierarchy of schemas, arranged in a top-down fashion based on the level of abstraction and generalization of description. The main idea behind our model is the common-sense observation that every appliance has a box-like geometry. All other operational aspects of this "box", such as the handle or the door, are positionally constrained to it in a very specific manner. This top-level box has a certain fixed size as a property and has "attributes" associated with it.

Every appliance, depicted as a box, has a set of "common" attributes associated with it, by virtue of its design and intended operation. These attributes include: *1)* an opening or a cavity, *2)* a door, *3)* a handle, and *4)* a control panel.

Each of these attributes have a fixed world location, specified relative to the box and has other task-specific properties associated with it. Correctly identifying these properties allow the system to generate a cognitive model that can specify, without explicitly requiring knowledge of the appliance itself, the possible tasks that can be performed with the appliance.

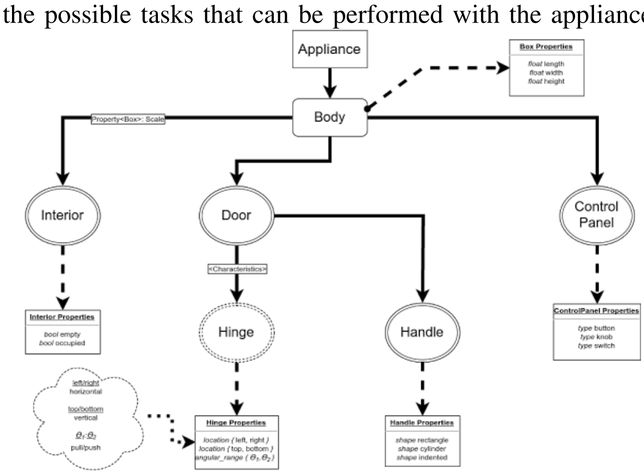

Fig. 1. Cognitive model hierarchy.

The cognitive model is generated according to the hierarchy specified in Fig. 1. From the figure, the topmost level of the hierarchy is populated by the box-like structure of the **Body** of the appliance itself. This level is characterized by the size of the appliance and its location in the world.

Following this, the hierarchy branches out into the **Interior**, also known as the internal cavity of the appliance. This interior cavity is also a box-like structure, usually having similar properties as the global box structure but having scaled-down property values. This **Interior** cavity is characterized by its relative location to the **Body** and has attributes of being `empty` or `occupied`.

Another child of the **Body**, we have the **Door**, having a relative positional constraint to the **Body**. The **Door** is characterized by the relative location of its sub-property, the **Hinge**. The **Door's** affordance, i.e. the property that allows the door to be manipulated, is decided by the properties of the **Hinge**, which has attributes of being located either at the `Top|Bottom` or `Left|Right` relative to the **Door**. This positional attribute dictates whether the **Door** will open horizontally or vertically and also determines the rotational/translational range of the **Door**.

The **Door** has another crucial sub-property, the **Handle**, which along with the **Hinge** also determines the affordance. The **Handle** is also relatively positioned to the **Door** and has attributes that describe the 'type' of handle present, namely a `rectangular` shape, a `cylindrical` shape or an `indented` handle.

The **Body** has another child, the **Control Panel**, which allows for operation of the appliance via its electronic control system. This control panel is characterized by the input it allows, such as via `knobs`, `switches` or `buttons`.

B. Using the Cognitive Model

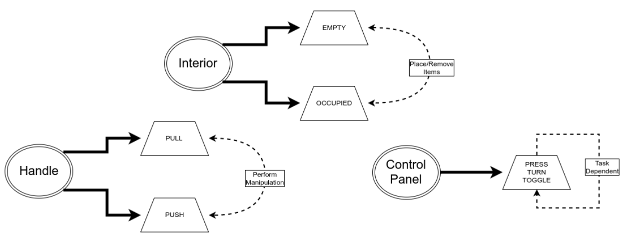

Fig. 2. Possible state transitions for each of the appliance parts.

Once a cognitive model for an appliance is generated, we can use the inherent properties of each sub-part of the appliance to allow the robot to operate the appliance on a task-by-task basis. Assuming a proper implementation of the manipulation mechanism, it becomes possible for the robot to start using the appliance based on the cognitive model. This is facilitated by using the cognitive model as a state machine, as illustrated in Fig. 2, where every sub-part of the appliance body has a set of states associated with it, depending on the situation. For instance, the **Interior** cavity of a microwave can either be `occupied` or `empty`, which will consequently dictate whether an object can be placed inside the cavity or be removed from it. A **Handle** can be either `pulled` or `pushed` and a **Control Panel** can be operated by `pressing|turning|toggling`, depending on the task.

III. Model Grounding by Observation

We implement a system that automatically grounds instances of the abstract cognitive model described above, given visual observations of humans operating the appliance under scrutiny. The input to our system is a recording, via a robot-mounted RGB-D camera, of a human demonstrating how to operate an appliance once. After several stages of offline processing, our system outputs a hierarchical structure that is a grounded instance of our proposed appliance cognitive model. Specifically, it both populates each individual part of the model with a suitable 3D geometric primitive in its correct position relative to the object reference frame and encodes affordances of movable parts by providing hinge connection specifications (e.g., a hinge for a door). Processing stages can be summarized as follows:

1) Fusion of the depth maps into a single point cloud while the entire observed scene remains rigid. Frames involving independently moving areas are flagged as *dynamic* and involve moving appliance parts. Every non-dynamic interval will result in a 3D point cloud *reconstruction* of a different (static) state of the appliance under observation.
2) Fitting of geometric primitives for all reconstructed point clouds. The first reconstruction corresponds to the initial appliance state (with all doors closed) and is used to fit the appliance exterior box. Subsequent reconstructions will be registered to the same coordinate frame and used to fit primitives to the (now visible) appliance cavities.
3) Dynamic segments are used to model moving parts and their joint types. Each segment is expected to capture the movement of a single appliance part. The part's motion is classified as either rotational (e.g., for a fridge door) or translational (e.g., for a drawer) and the estimated hinge parameters are stored in the appliance model.

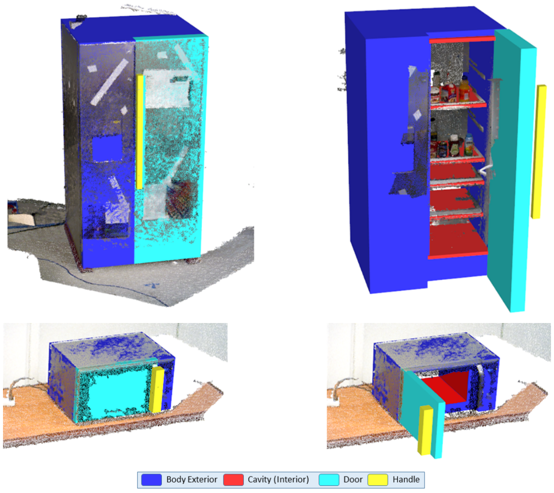

Fig. 3. Grounded model visualization for a fridge and a microwave oven.

IV. Results

In Fig. 3, we depict grounded instances of our cognitive model for two appliances: a fridge and a microwave oven. The left column shows models in 'closed' state, overlaid to the exterior reconstructed point cloud. The right shows each appliance model rendered as 'open' with interior boxes becoming visible, showing detected affordances of the moving parts.

References

[1] A. Hudnut and W. Gross, "Vision-enabled household appliances," Mar. 8 2011, uS Patent 7,903,838. [Online]. Available: https://www.google.com/patents/US7903838
[2] A. Enslin, "Vorrichtung und verfahren zum ansteuern eines haushaltsgeräts sowie ein haushaltsgerät apparatus and method for controlling a household appliance and a household appliance," Mar. 3 2016, dE Patent App. DE201,410,112,375. [Online]. Available: https://www.google.com/patents/DE102014112375A1?cl=en