An Ontology-based Symbol Grounding System for Human-Robot Interaction

Patrick Beeson and **David Kortenkamp** and **R. Peter Bonasso**
TRACLabs Inc.

Andreas Persson and **Amy Loutfi**
Örebro University

Jonathan P. Bona
University at Buffalo SUNY

Introduction

One of the fundamental issues in HRI is to enable robust mappings of the robot's low-level perceptions of the world and the symbolic terms used by humans to describe the world. When a shared *symbol grounding* is achieved (when human and robot use the same terms to denote the same physical entities) a two way interaction is enabled. This interaction is an important step towards robots assisting humans in everyday tasks at home, as the human can easily understand the "intelligence" of the robot in a domain, and in turn the robot can query the human to bootstrap more knowledge to better assist in complex or novel situations.

A symbol grounding system must regularize the connections between the sensed physical world and language. Consider such a system running on a home care robot. If a box of pasta appears in front of the robot's sensors, a symbol identifying it ("pasta_box") should be generated with high consistency. The architecture should maintain the coherence of perceptually-grounded symbols over time, so knowledge of the location, permanence, and ubiquity of certain items is needed in order to track $pasta\_box_1$, for instance, and distinguish it from others. If something temporarily occludes $pasta\_box_1$ from the sensors, the architecture should not create a new symbol for the object when it reappears. If another box of pasta appears in a different place at the same time, then a second pasta symbol should be created, as one object cannot be in two places at once.

This paper presents a preliminary overview to the symbol grounding problem for HRI that relies on monocular vision processing and hierarchical *ontologies* to help define symbols. Our approach focuses on the use of a long-term memory model for a robot in a home environment that persists over a significant duration. The robot must learn and remember the properties, locations, and functions of hundreds of objects with which the homeowner interacts during normal activities. Starting with an a priori long-term memory stored in an ontology that will be updated throughout its operation, the robot then needs to link its perception of objects as well as actions (both its own and the homeowner's) to representations in its long-term memory. The long-term memory needs to be linked to both a working memory and a perceptual memory in order to support the following functions:

- Queries of the type:
  - Where is object X?
  - Why do you think it is there?
  - When was object X last used and where?
- Reasoning:
  - How likely is object X to have moved?
  - What sensor(s) are best used to find object X?
  - For what functions can object X be used?
- Generalization and specialization:
  - Bring me an object of type Y
  - Bring me something similar in function to X
- Learning
  - Remember this object and call it X
  - Object X appears to be a new type of object class Y
- Activity recognition
  - Why is the person using object X?
  - What objects will a person need to do an activity?

Large-scale Perceptual Anchoring

At the sensory level, the main challenge for this architecture is to create and maintain connections between percepts and symbols which refer to the same objects. This sensor-to-symbol problem has also been called the perceptual anchoring problem. *Perceptual anchoring* is a subset of symbol grounding that mainly concerns maintaining the appropriate symbol-percept link over time in robotic systems called an anchor. In a real home environment, a robot can encounter several hundreds or thousands of objects; thus, an important component is an efficient matching mechanism to match a newly acquired object to an existing anchor.

In this work, we follow an approach which utilizes the fact that both perceptual and symbolic information about anchors is available in order to enable a fast matching of anchors according to symbolic categories in a bottom-up manner (Loutfi, Coradeschi, and Saffiotti 2005). Persson and Loutfi (2013) presented a method to summarize a large database of existing anchors and enable matching of new objects. This method uses binary-valued visual features to anchor objects, where all binary-valued visual features of an

object are summarized through a weighted frequency count into 2D arrays and structured into computationally efficient hash tables. Hash tables are created according to symbolic categories such that a hash table encapsulates all objects with a joint symbolic category, e.g. a hash table is created for all objects associated with the symbol "pasta_box".

Currently, these hash tables and symbolic categories are built from a rich *reference space* of known objects photos and descriptions, which has been automatically extracted from online resources. As a result, an object match of a hash table in reference space will also result in a percept-symbol connection through the joint symbolic category for the hash table. For example, visual features on a new box of pasta placed in front of the robot will match the stored features of anchors in the hash table associated with the symbol "pasta_box". This will trigger the creation of a new anchor $pasta\_box_1$ that becomes an instance of the *BoxOfPasta* category in the ontology and thus has all the properties and relations associated with boxes of pasta.

Ontology

An ontology is a rigorous organization of knowledge of a domain, containing all relevant entities and their relations. In this work, the ontology describes the available objects, their capabilities, the tasks that can be performed, and the resulting states of those tasks (locations, temperatures, etc.). The ontology includes terms for entities that range from the general (e.g., *PhysicalEntity*) to specific (*Pasta*). Our ontology also has terms for particular products that are known objects recognizable by the vision system. These are described in the ontology itself using product information (name, dimensions, mass, etc.). The ontology is represented in OWL so that off-the-shelf OWL reasoners can be used to perform inference and maintain ontology consistency. Our domestic robot OWL ontology was built using TRACLabs' graphical software suite PRONTOE (Bell et al. 2013), which is used to define the classes, instances, and properties of the objects in the robot's environment and can be used by non-experts.

Querying to Enable HRI

Given an ontology and an anchoring system that can match perception to ontology categories, human interaction with the robot can occur by querying the ontology. This querying should happen in a natural fashion. To do this, we are using the dynamic predictive memory architecture (DPMA). DPMA integrates the reactive execution system RAPs (Firby 1987) with a DMAP parser (Martin and Firby 1991). The RAPs system has reactive action packages to process queries and commands and to disambiguate parser results. When DPMA is started, it loads the ontology of the domain and starts the RAPs process, which monitors working memory for the appearance of a parsed query. When the user inputs a string like "How many tables are there?" in a UI, the statement is parsed and formatted into a description that is sent to the DPMA working memory. The RAPs system interprets that description and runs its own deductive query, which in our example returns a list of table instances. RAPs then packages up the list as a string and sends that back to the UI for display to the user. Eventually, we expect to handle more complicated expressions such as "What items in the kitchen used to be somewhere else?".

Related work

The integration of knowledge, representation, and reasoning (KRR) with embodied systems has been an increasingly interesting topic for cognitive robotics, including semantic mapping (Galindo et al. 2005; 2008), improving planning and control (Mozos et al. 2007), and HRI (Holzapfel, Neubig, and Waibel 2008; Kruijff et al. 2007). Pangercic et al. (2009) considers semantic knowledge, and in particular encyclopedic knowledge, in the context of household robotic environments. Another approach focuses on practical and grounded knowledge representation systems for autonomous household robots (Tenorth and Beetz 2008). Other HRI-oriented approaches focus on human-robot dialog. Zender et al. (2007) present an HRI architecture for human augmented mapping is used by a robot to improve its autonomously acquired metric map with qualitative information about locations and objects in the environment. Typically, such systems use small KRR systems, tailored to the specific application at hand.

The systems mentioned here, even if implicitly dealing with anchoring, lack a generic solution to the symbol grounding problem. They hard-code ad hoc solutions or use very small knowledge sets. Similarly, they often do not reason about multiple instances of the same type of object, either in the perceived scene or over time.

Conclusions and Future Work

This paper has discussed perceptual anchoring and its potential to enable HRI via symbolic representations of objects. The work presented here is a first step towards enabling perceptual anchoring to operate with larger symbolic (semantic) models such as ontologies, with a focus on large scale and long term anchoring in this context. Future work includes improving knowledge of common object uses and extending the representation of objects to characterize movement/change over time. These improvements may be facilitated by mining the Internet for new information about objects of interest and by aligning our ontology with one or more relevant external ontologies. Inferring additional properties of objects through reasoning (e.g., "Cooking pasta makes it hot") and incorporating additional sensory modalities (i.e. 3D data through RGB-D sensors) will also improve recognition and reasoning. The methods proposed here will be developed and validated on two smart home test platforms, one in Sweden and one in Houston, Texas.

Acknowledgments

We would very much like to acknowledge the late Silvia Coradeschi as part of our research team whose presence is greatly missed.

References

Bell, S.; Bonasso, R. P.; Boddy, M.; Kortenkamp, D.; and Schreckenghost, D. 2013. PRONTOE: A case study for de-

veloping ontologies for operations. In *Proceedings of the INSTICC International Conference on Knowledge Engineering and Ontology Development (KEOD)*.

Firby, R. J. 1987. An investigation into reactive planning in complex domains. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.

Galindo, C.; Saffiotti, A.; Coradeschi, S.; Buschka, P.; Fernandez-Madrigal, J.-A.; and González, J. 2005. Multi-hierarchical semantic maps for mobile robotics. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Galindo, C.; Fernandez-Madrigal, J.; González, J.; and Saffiotti, A. 2008. Robot task planning using semantic maps. *Robotics and Autonomous Systems* 56(11):955–966.

Holzapfel, H.; Neubig, D.; and Waibel, A. 2008. A dialogue approach to learning object descriptions and semantic categories. *Robotics and Autonomous Systems* 56(11):1004–1013.

Kruijff, G.-J.; Lison, P.; Benjamin, T.; Jacobsson, H.; and Hawes, N. 2007. Incremental, multi-level processing for comprehending situated dialogue in human-robot interaction. In *Proceedings from the Symposium on Language and Robots (LANGRO)*.

Loutfi, A.; Coradeschi, S.; and Saffiotti, A. 2005. Maintaining coherent perceptual information using anchoring. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

Martin, C. E., and Firby, R. J. 1991. Generating natural language expectations from a reactive execution system. In *Proceedings of the Cognitive Science Conference*.

Mozos, Ó. M.; Jensfelt, P.; Zender, H.; Kruijff, G.-J.; and Burgard, W. 2007. An integrated system for conceptual spatial representations of indoor environments for mobile robots. In *Proceedings of the IROS 2007 Workshop: From Sensors to Human Spatial Concepts (FS2HSC)*.

Pangercic, D.; Tavcar, R.; Tenorth, M.; and Beetz, M. 2009. Visual scene detection and interpretation using encyclopedic knowledge and formal description logic. In *Proceedings of the International Conference on Advanced Robotics (ICAR)*.

Persson, A., and Loutfi, A. 2013. A hash table approach for large scale perceptual anchoring. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics (SMC)*.

Tenorth, M., and Beetz, M. 2008. Towards practical and grounded knowledge representation systems for autonomous household robots. In *Proceedings of the International Workshop on Cognition for Technical Systems*.

Zender, H.; Jensfelt, P.; Mozos, Ó. M.; Kruijff, G.-J. M.; and Burgard, W. 2007. An integrated robotic system for spatial understanding and situated interaction in indoor environments. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.