A Robotic World Model Framework Designed to Facilitate Human-robot Communication

Meghann Lomas, E. Vincent Cross II, Jonathan Darvill, R. Christopher Garrett, Michael Kopack, and Kenneth Whitebread
Lockheed Martin Advanced Technology Laboratories
3 Executive Campus, Suite 600, Cherry Hill, NJ 08002
1 856.792.9681
{mlomas, ecross, jdarvill, rgarrett, mkopack, kwhitebr}@atl.lmco.com

Abstract

We describe a novel world model framework designed to support situated human-robot communication through improved mutual knowledge about the physical world. This work focuses on enabling a robot to store and use semantic information from a human located in the same environment as the robot and respond using human-understandable terminology. This facilitates information sharing between a robot and a human and subsequently promotes team-based operations. Herein, we present motivation for our world model, an overview of the world model, a discussion of proof-of-concept simulations, and future work.

1 Introduction

As robots become more ubiquitous, their interactions with humans must become more natural and intuitive for humans. One of the main challenges to natural human-robot interaction is the "language barrier" between humans and robots. While a considerable amount of work has gone into making robot dialogue more human-like (Fong et al., 2005), the content of the conversation is frequently highly scripted.

An essential precondition to intuitive human-robot dialogue is the establishment of a common ground of understanding between humans and robots (Kiesler, 2005). Operators expect information to be presented in a way such that they can connect it with their own world information. This implies a need for robots to be capable of expressing information in human-understandable terms. By shifting some responsibility for establishing common ground to robots, interactions between humans and robots become considerably more natural for humans by reducing the need for humans to "translate" the robot's information.

Ultimately, the robot's world model is a key contributor to the "language barrier." Because humans and robots view and think about the world differently (having different "sensors" and "processing algorithms"), they subsequently have different world representations (Figure 1). Humans tend to think of the world as objects in space, while robotic representations vary based on sensors, but are typically coordinate-based representations of

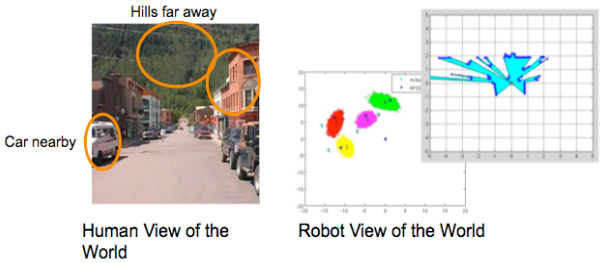

Figure 1. Humans and robots think and subsequently communicate about the world using different terminology.

Proceedings of the SIGDIAL 2011: the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pages 301–306,
Portland, Oregon, June 17-18, 2011.

free and occupied space. This presents a considerable challenge when humans want to communicate naturally with robots. For robots to become active partners for humans, they must be better able to share the information they have gathered about the world. To that end, we have begun to address the "language barrier" by focusing on how information is stored by the robot.

We have developed a novel world model representation that will enable a robot to merge information communicated by its human teammates with its own situational awareness data and use the resulting "operating picture" to drive planning and decision-making for navigation in unfamiliar environments. The ultimate aim of this research is to enable robots to communicate with humans *and* maintain an "actionable awareness" of the environment. This provides a number of benefits:

- *Increased robot situational awareness.* The robots will be able to learn about, store, and recall environmental information obtained from humans (or other robots). This can include information the robot would be incapable of getting on its own, either because it has not visited that region of the environment or because it is not capable of sensing that information.
- *Increased human situational awareness.* Humans will be able to receive information from robots in human-understandable terms.
- *Reduced workload and training for human-robot interaction*. Because robots will be able to communicate in human-understandable terms, people will be able to interact with robots in ways that are more natural to humans. As a result, people will need fewer specialized interfaces to interact with robots and subsequently less training.
- *Improved collaboration*. Because people and robots will be able to share information, the team will be able to operate more efficiently. Each team member will be able to contribute to team knowledge, which will allow for better planning.

2 World Model Overview

Our world model framework was designed using several key principles: that information must be stored in both human-understandable terms and in a format usable by the robot; that information must be capable of being added, deleted, or modified during operations; and that the world model framework should be capable of integrating with a wide variety of external systems including pre-existing perception and planning systems.

To meet these principles, we have developed a layered framework that has internal functions for managing the world model and can integrate with external systems that use the world model, such as systems that populate it (perception systems) or use it to govern robotic actions (planning systems) (Figure 2).

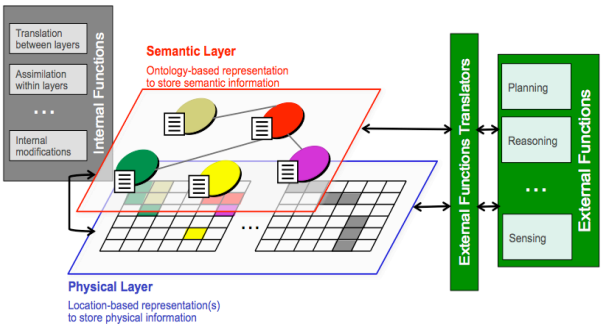

Figure 2. We have developed a two-layer world model that integrates with external functions via translation functions to support the use of a variety of robotic capabilities.

Layered world models have shown promise for both robot navigation (Kuipers and Byun, 1991; Mataric, 1990) and for communication with humans (Kennedy et al., 2007; Zender et al., 2008). Additionally, work in symbol grounding has supported robotic actions based on natural language interactions (Jacobsson et al., 2008, Hsiao et al., 2008). We leverage this research and extend it with the aim of supporting human-robot information sharing, robot navigation, and use by external systems.

The bottom layer stores a spatiotemporal description of the environment expressed in metrical terms. While there are several different possibilities for how this location-based information could be stored, we use a grid-based representation because it is commonly used by existing planners (e.g., a cost map-based planner) and it allows for flexibility of information storage. While our framework supports the inclusion of an arbitrary number of grids, our experimental prototype uses three: an occupancy grid that stores free and occupied space, an "object" grid, and a "terrain" grid. The object grid stores the types of objects in each

cell in ascending order of vertical position (e.g., "table, plate, apple"). The terrain grid stores terrain type in each cell and may also have multiple entries per cell (e.g., "sand, boulders" or "grass").

The top layer stores a relational description of the situation in semantic terms compatible with typical human descriptions of the physical environment. We use node-attribute structures in which objects (e.g., chairs, keys, trees, people, buildings) are represented as nodes that have a list of corresponding attributes (e.g., type, color, GPS coordinates, last time sensed, source of information, etc.). The nodes are connected by their relationships, which are human-understandable concepts (e.g., "near" or "above"). The graph form of the semantic layer supports the many, varied types of relationships between objects. There are many ways to express the physical relationships between objects, and humans often use ambiguous terms (Crangle et al., 1987). By establishing the semantic layer as a connected graph, we aim to support these ambiguous terms and ultimately provide a way for the robot to process their meaning.

In the top layer of the world model, we use an ontological representation to model the world, and include both an "upper ontology" that provides a template for what information can be included in the world as well as an instantiated world built from experience. In addition to providing a framework that stores the list of all objects that could be present in the world, their associated attributes, and the possible relationship between the objects, this upper layer includes other information such as the robot's goals and current high level plans and additional information the robot has about itself or the world (e.g., domain theory or object affordances). An additional benefit of an ontology-based representation is that it supports the inclusion of objects despite uncertainty. If a perception algorithm cannot confidently identify an object but can classify it, this class of object can be stored in the semantic layer of the world model and refined as more information is made available.

To support a consistent, complete view of the world, translation functions translate the information between the layers and assimilation functions merge information within layers. These translation functions support symbol grounding and enable the robot to use both semantically-described information along with sensed data. The translation functions are a set of functions, each of which translates an attribute, for example, a color translation function that translates between RGB values and a semantic label. More interesting are the location-based translation functions, for example "near A" translates to "within 2 meters of A's position." This introduces uncertainty into the position of the object and so we use a probabilistic approach for placing any unsensed (but described) object in the bottom layer. The location of the object is updated once the object is sensed by the robot.

The assimilation algorithms, which are also still in development, are built upon data fusion ideas because they merge data from multiple sources. Because a considerable amount of existing work has been done on integrating (assimilating) information at the sensor level, to date we have focused on assimilation in the semantic layer of our world model. We have developed heuristic-based algorithms that compare information stored in the world model with actively sensed information (essentially creating a temporary world model of the area currently being sensed by the robot). During operation, the robot's sensor detects an object and outputs a vector of possible object classifications. Each object classification has an associated confidence along with attributes of the object including size, color, etc. The assimilation component pulls all objects within a prescribed radius of the newly sensed object's location from the world model to compare them with the newly sensed object. The assimilation algorithm starts with the object closest in position to the newly sensed object and stops comparing objects if an object is determined to be "same as" the newly sensed object or if all objects with the prescribed radius are compared and none match.

To compare our newly sensed object with one of the objects already in the world model, the assimilation algorithm compares the object vectors, which contain the list and confidence in each object type and object attributes such as color, size, and location. Some attributes (like source of information) are ignored in this calculation. To compare two objects, we compute the distance between the object vectors. This distance is computed through a pairwise comparison of attributes in the vector lists. These distances are then weighted according to "importance" in assimilation process, for example objects with similar type should be more likely to be merged than objects

that only have similar color. We then sum the weighted distances; if sum is less than a prescribed threshold, we assume the objects are the same and then merge them. If not the same, the algorithm checks this object against the other objects within the radius and if none are found, adds the object as a new object. To merge objects, the algorithm merges the attribute vectors of the temporary object and the original object. Some parts of the vectors are averaged (e.g., color), some amalgamated (e.g., data source), and some pick one of the values (e.g., pick most recent time). Additionally, because it is stored in the world model, we can incorporate logic about the world to facilitate assimilation (e.g., "this object is immovable so it must not have changed position"). While this algorithm has served as an initial assimilation algorithm, we will continue researching and designing assimilation algorithms to better support the uncertainty present in the sensing outputs (e.g., false positives).

One of the key requirements of our world model is that it be able to integrate with external robotic systems. To accomplish this, the world model layers integrate with external functions that serve as translators to existing (or future) functions. These external translation functions pull relevant information from the world model and present it in a form usable by a planner. For example, we have created a planning translator that takes the grids from the physical layer and produces a cost map for a ground robot (with set parameters), which can then be used by any cost map-based planner.

3 Proof-of-Concept Simulations

To evaluate the feasibility of our world model framework, we performed several proof-of-concept simulations designed to both demonstrate and test the capabilities of our world model and subsequently to help the design process. We created different environments using Player/Stage and ran the robot through two scenarios. In both scenarios, humans needed robotic assistance to escape from a burning building and communicated with the robot using natural language. In the first scenario, a mobile robot was asked by a group of trapped people to unlock a door and alert them when the door was open. In the second scenario, two mobile robots were tasked with searching for trapped people and coordinating with first responders. Because the focus of the simulations was on evaluating the world model itself, we made the assumption that the robot had both camera and LIDAR sensors and had processing algorithms capable of outputting an object classification and a confusion matrix. We assumed the robot had both a speech processing and synthesis mechanism with which it could communicate verbally with people in the environment. We assumed the robot had a common A* planner that used a cost map representation for planning.

The first scenario highlighted the ability for the robot to understand and use human-communicated information by adding a human-described object to its world model and planning based on this assimilated information. At the beginning of the scenario, a human described the location of a key ("near the desk in the room with one table and one desk") and told the robot to open the locked east door. The human did not tell the robot to use the key to unlock the door, instead the robot used object affordances stored in its world model to establish a high-level plan of getting the key, then unlocking the door. When the human told the robot about the location of the key, the robot stored this location in the top layer and translated the object's position down to the bottom layer using a probabilistic translation algorithm that placed the key in the bottom layer at the most likely position within a certain region (whose size and position corresponded to "nearness"). The robot used a simple cost map-based planner to plan its movements and so the system created a cost map from all the relevant bottom layer information in a format used by a classic A* planner. As a result, this scenario showed that our world model enabled the robot to use information gathered by a human teammate and expressed in semantic terminology without a specially designed planner.

The second scenario illustrated the merits of our world model for *responding* to humans. In this scenario, once the robot had searched the environment, it was asked a series of questions by a first responder including: "How many people did you find?" and "How do I get to the fire extinguisher?" The latter question was particularly interesting because it forced the robot to describe a path in semantic terminology (as opposed to a list of waypoints). The robot used information from its top layer to describe the path from the first responder's

current position to the fire extinguisher. This scenario highlighted the ability for the robot to produce human-understandable and useful information despite having gathered the information using its low-level sensors and planner.

In both of the scenarios, the robot was given both instructions and information verbally from one or more of the people in the robot's environment. The robot stored this described information in the world model and merged it with the information the robot had gathered with its own sensors to form a cohesive view of the world. The robot then used both the described and sensed information to formulate a plan to accomplish its goals. At the end of the mission, the robot was asked questions about the environment and was able to answer using *human understandable* terminology.

In these simulations we were able to show the robot formulating a plan based on information it had not sensed by itself. Because the robot had only a simple cost map-based planner, it was essential that the semantic information be translated to the grid representations in the bottom layer. This allowed the planning translator to produce a cost map in the form expected by the planner.

We used these simulations to inform key design decisions including the need to have multiple grids in the bottom layer of the world model and to incorporate object affordances in the semantic layer. Another key insight was that uncertainty must be included in the semantic layer and that it is an important element in semantic layer assimilation.

4 Conclusions and Future Work

We have designed and developed a world model framework that supports situated information sharing between robots and humans. By integrating semantic and sensor-based terminology, we have enabled a robot to integrate information described in natural human terms with its own sensed information. In addition, we have shown how a robot with a standard A* planning algorithm can thereby plan and respond appropriately using information obtained in semantic terms.

Because this world model framework was designed to support a variety of robotic operations and capabilities, there are many areas of potential future work. These include facilitating robotic dialogue systems, developing reasoning systems that can use the semantic level information to predict certain aspects of the world model (such as how an event will affect the physical layout of the world or where an object will be in a certain amount of time), and enabling semantic-level planners that can perform high-level planning.

To further improve the functionality supported by this world model framework, there are a number of areas of future work within the framework itself. We are exploring the design changes needed to support modeling of dynamic objects and the types of assimilation algorithms that exist or need to be developed to truly integrate tracks generated by external perception systems into our world model. We are also looking into how to better reason about spatial relationships, particularly those that are only true when described from a specific vantage point. Additionally, we would like to improve the translation algorithms by exploring additional scenarios and determining what mechanisms are needed. In the area of multi-robot coordination, we want to explore physical layer assimilation, which includes the ability to align reference frame for heterogeneous robots. Finally, we would also like to apply our world model on multiple real robots with speech systems and evaluate the world model in a series of real-world operations.

References

Terrence W. Fong, Illah Nourbakhsh, Robert Ambrose, Reid Simmons, Alan Schultz, and Jean Scholtz. The peer-to-peer human-robot interaction project. AIAA Space, 2005.

S. Kiesler. Fostering common ground in human-robot interaction. Robot and Human Interactive Communication Proceedings. ROMAN 2005. The 14th IEEE International Workshop. Nashville, TE. Aug 2005.

Benjamin Kuipers and Yung-Tai Byun. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representation. Journal of Robotics and Autonomous Systems, 8:47–63, 1991.

Maja Mataric. A distributed model for mobile robot environment-learning and navigation. Technical Report, MIT Artificial Intelligence Laboratory, 1990.

William G. Kennedy, Magdalena D. Bugajska, Matthew Marge, William Adams, Benjamin R. Fransen, Dennis Perzanowski, Alan C. Schultz, and J. Gregory Trafton. Spatial representation and reasoning for human-robot collaboration. In Proceedings of the

Twenty-Second Conference on Artificial Intelligence, 2007.

C. Crangle, P. Suppes, and S. Michalowski. Types of verbal interaction with instructable robots. In Proceedings of the Workshop on Space Telerobotics, Vol 2, 1987.

H. Zender, O. Martinez Mozos, P. Jenselt, G.-J. M. Kruijff, and W. Burgard. Conceptual Spatial Representations for Indoor Mobile Robots. Robotics and Autonomous Systems, Special Issue "From Sensors to Human Spatial Concepts." Vol. 56, Issue 6. pp. 493-502. Elsevier. June 2008.

H. Jacobsson, N. Hawes, G-J. Kruijff, J. Wyatt, Cross-modal Content Binding in Information-Processing Architectures. Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI). March 2008. Amsterdam, The Netherlands.

Kai-yuh Hsiao, Soroush Vosoughi, Stefanie Tellex, Rony Kubat, Deb Roy. (2008). Object Schemas for Responsive Robotic Language Use. Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction, pages 233-240.