Language Defined Behavior for Humanoid Control

Steve McGill*, Seung-Joon Yi*, Qin He*, Heejin Jeong*, Dennis Hong† and Daniel D. Lee*

Abstract— **Humanoid robots are meant as an artificial surrogate for humans in the real world due to their anthropomorphic form and the ability to act like human agents in a world modified for humans. Current control interfaces for these robots do not reflect this philosophy, where robots wait for explicit commands. In this work, we suggest a universal natural language interface that guides humanoid robots via an accessible platform for untrained operators to to refine object detection and motion planners.**

Keywords: Human Robot Interaction, Natural Language Processing, Humanoid Robot

I. Introduction

Speech recognition has long been a frustrating affair for both researchers and users. However, the tides are changing, with notable commercial deployments having a huge success with everyday technology users. Apple's Siri is competing with Google's Now and Microsoft's Cortana for connecting users with information as quickly and correctly as possible with natural language interface. Many automobile manufacturers actively advertise their speech recognition for hands free operation of cars. These applications show that now we have reached the level where users are quite comfortable conversing to their devices to achieve tasks.

Robotics researchers are exploring natural language for human robot interaction. To command robots, the community is developing corpuses of speech commands, and their acquisition methodologies, for benchmarking speech recognition systems [1]. Natural language commands must then be grounded in the objects of the world, which can be done probabilistically to deal with perception noise [2]. Ambiguity in human commands can compound the noise, leading to task failures; prevention strategies for reducing this uncertainty has been explored using clarifying questions [3].

These methods define a roadmap for coupling a corpus of commands with inference of human command intent; at least in semi-structured environments. However, a command/confirm/execute process relies on the user to confirm object identification and path planning strategies of the robot before execution, unless the robot is quite certain. Operating in unstructured environments similar to the DARPA Robotics Challenge, certainty is not achievable. Unfortunately, having the robot wait for commands and confirmations increases the task completion time significantly; network degradation exacerbates this issue. A robot with more autonomy that is always acting, but listening for human feedback, can increase performance in this situation while moving towards cooperative human robot interaction [4].

* GRASP Laboratory, University of Pennsylvania, Philadelphia, PA 19104 {smcgill3,yiseung,heqin,heejinj,ddlee}@seas.upenn.edu
† RoMeLa Laboratory, UCLA dennishong@ucla.edu

Fig. 1. During the DRC Trials, the operator would command the robot to complete the valve task based on experience, disallowing the robot to chose its own policies.

II. Current Methodology

For the DARPA Robotics Challenge Trials, we use a general model driven approach, with a heavy focus on teleoperation [5]. The robot maintains a task description model which can be set by the operator or autonomously acquired by the robot using onboard perception modules and sensors such as RGBD camera or LIDAR. With the description model, the robot can generate each action for the task based.

For example, in the valve task, we use the three dimensional position and angle of the valve as the model, where the actions include approaching the valve, moving the arms to grasp the valve, rotating the valve and, finally, retracting the arms to the initial position. We have sliding autonomy for performing the task: the robot can complete the whole task autonomously given that the initial perception is good enough or the human operator can manually command and refine each action as needed.

III. Suggested Methodology

Our current methodology has a number of drawbacks, especially when the perception capability of the robot is less than ideal. Manually refining the task specific model, defined as a vector, needs some training and takes time; the robot has to frequently wait for the new commands, doing nothing.

We suggest a probabilistic belief model for the robot and natural language interface to handle these issues. Instead of a single model that describes the task target, the robot detects and maintains multiple models. It executes a behavior based on the most likely task model, which is updated through human intervention. To simplify the operator feedback, we use natural language cues instead of full model refinement. This guides – not commands – the robot to establish a task model that the human operator wants.

With this approach, we can avoid the many hours of trial and error practice sessions for the humans to understand appropriate model parameters and take advantage of autonomous detection and planning. The robot has its own belief of the surrounding environment, which is refined by natural language feedback from the human operator, and continuously takes actions to achieve a task.

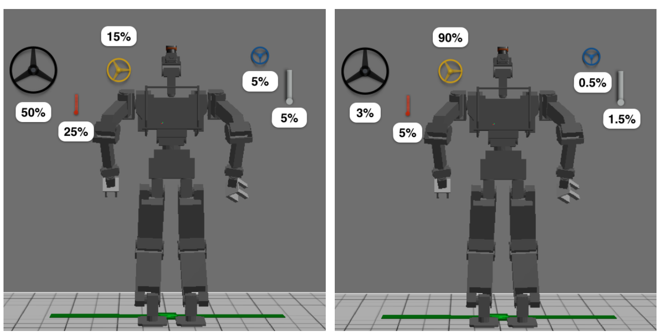

Fig. 2. A probabilistic scenario allows the robot to act autonomously, and receive updates from a human for replanning. Here, given a command to turn the yellow valve, the probabilities are updated.

A. Object Classifiers

The probabilistic approach requires a classifier to produce a set of detected valves, with associated parameters including size, position, color, type, etc. Additionally, a planner would take this information and assign a likelihood of approaching this valve in order to execute a behavior of approaching and turning the valve. However, during this routine, the robot may plan to use the wrong valve. The user can monitor the probabilities of each valve having the focus of the robot, and update these probabilities with simple commands: *"Grab the smallest valve"* or *"Grab the yellow valve."*

Here we can map the richness of language to the operator interface. While clicking on a particular valve can achieve a similar effect, the reasoning for *why* a particular valve is not understood by the robot. If more valves appear to the robot's perception, it will understand how to reassign the probabilities, as illustrated in Figure 2.

B. Path Planning

After approaching the valve, the robot needs to manipulate the valve with its gripper. Our planner generates trajectories for grabbing and turning the valve in either a single handed or bimanual manner, as shown in Figure 1. However, this planner is restricted to a very small set of known-to-work trajectories. By presenting the user a set of possible trajectories, with a likelihood of the robot performing them, we can increase the number of viable plans. With language cues, the user can specify to *"Use the one from the top"* or *"Turn with two hands."*

IV. Discussion

With humanoid robots providing a high degree of freedom system, the role of intent recognition becomes very important. Training operators to work with these systems can be time consuming with traditional user interfaces, and, additionally, a lack of bandwidth between operator and robot requires succinct high level instructions for controlling a humanoid. For these reasons, it makes sense to apply speech language processing systems to convey commands from human to humanoid.

A. Open Issues

With the proposed system, we find that there are still open questions in effectiveness. Dextrous maneuvers may require a cumbersome vocabulary to differentiate plans, where average users may still need some training.

Acknowledgments

We acknowledge the Defense Advanced Research Projects Agency (DARPA) through grant N65236-12-1- 1002. We also acknowledge the support of the ONR SAFFIR program under contract N00014-11-1-0074.

References

[1] E. Bastianelli, L. Iocchi, D. Nardi, G. Castellucci, D. Croce, and R. Basili, "RoboCup@Home Spoken Corpus: Using Robotic Competitions for Gathering Datasets," in *Proc. of RoboCup Symposium*, 2014. [Online]. Available: http://fei.edu.br/rcs/2014/SpecialTrackDev/robocupsymposium2014_submission_33.pdf

[2] S. Tellex, T. Kollar, S. Dickerson, M. Walter, A. Banerjee, S. Teller, and N. Roy, "Understanding natural language commands for robotic navigation and mobile manipulation," 2011. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI11/paper/view/3623

[3] S. Tellex, P. Thaker, R. Deits, D. Simeonov, T. Kollar, and N. Roy, "Toward information theoretic human-robot dialog," in *In Proceedings of Robotics: Science and Systems*, 2012.

[4] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, B. van Riemsdijk, and M. Sierhuis, "The fundamental principle of coactive design: Interdependence must shape autonomy," in *Proceedings of the 6th International Conference on Coordination, Organizations, Institutions, and Norms in Agent Systems*, ser. COIN@AAMAS'10. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 172–191. [Online]. Available: http://dl.acm.org/citation.cfm?id=2018118.2018130

[5] S.-J. Yi, S. McGill, L. Vadakedathu, I. He, Qin and Ha, M. Rouleau, D. Hong, and D. D. Lee, "Modular low-cost humanoid platform for disaster response," in *Intelligent Robots and Systems (IROS), 2014 IEEE/RSJ International Conference on*, 2014.