Cognition-enabled Task Interpretation for Human-Robot Teams in a Simulation-based Search and Rescue Mission

(Extended Abstract)

Fereshta Yazdani
Institute for Artificial Intelligence
University of Bremen
Am Fallturm 1, 28359 Bremen, Germany
yazdani@cs.uni-bremen.de

Matthias Scheutz
HRI Laboratory
Tufts University
200 Boston Ave, Medford
MA 02155, USA
matthias.scheutz@tufts.edu

Michael Beetz
Institute for Artificial Intelligence
University of Bremen
Am Fallturm 1, 28359 Bremen, Germany
beetz@cs.uni-bremen.de

ABSTRACT

Due to humans and robots complementary capabilities, mixed human-robot teams are considerably deployed in real-world settings. A favored communication means is the natural language used by humans that is still a challenge for robotic teammates. They need to understand the environment from the viewpoint of their human teammates in order to be able to translate the instructions received in the natural language into plans. To deal with human-robot teaming, we report the results from an empirical study in a visual search and rescue task application for mixed human-robot teams. We show how those results and ideas can be intepreted into robot action plans by using a cognition-enabled task interpretation system.

1. INTRODUCTION

The interest in deploying mixed human-robot teams for complex search and rescue missions has been increasing as robots are becoming more capable in terms of their physical behaviors and mobility. An example shows the Sherpa project [5] which aims to investigate how a human operator could lead a team of robots in rescue missions after avalanche accidents.

While understanding natural language sentences or discourse is effortless for humans, interpreting such natural language expression especially understanding the meaning beyond the surface content of the utterance, is very difficult for robots. They require background knowledge in addition to knowledge about the words employed in the utterance. Ideally, communication between humans and robots should be as natural as between humans. However, instructions by humans must be restricted so that robots can understand them. The goal is to make the interactions as natural as possible. To be able to develop the necessary natural language understanding and reasoning mechanisms, we will briefly present few results of a conducted empirical study of a visual search and rescue task. We present some of the results by using our cognition-enabled interpretation system and show the interpretation from high-level natural language commands into low-level action plans comprehensible for robots. Those results can be used for dialogue systems in order to improve the understanding of natural language for robotic systems.

Appears in: *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), Das, Durfee, Larson, Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

2. RELATED WORK

Recent effort in HRI has focused on developing robots that work directly with humans as assistants or teammates [1, 9] to perform tasks or to achieve common goals. Much of the work in HRI is focused on making robots more “human-like”, especially many researchers have been developing robots that exhibit human traits [4] or can interact via natural language and gestures [7, 8].

In the context of natural language interactions, reference and object identity resolution received special attention. For example, Blodow et al. [3] introduce a framework that deals with references between observations and entities in a human living environment and enables the robot to detect a wide variety of objects of daily use.

There exist many studies in the field of HRI that are clear and well structured, however those are often prepared for indoor terrains where humans are able to interfere and correct robots during misinterpretations while in outdoor environments this is difficult.

In this paper we show how the instructions are formulated for an outdoor terain and how the robots have to process them in order to interpret and perform them correctly.

3. STUDY FOR SEARCH AND RESCUE ENVIRONMENTS

The experiment was conducted with humans to see how humans command other humans in a natural way in outdoor settings. We created several images of outdoor scene that described several tasks to be accomplished by the participants. These images include possible landmarks that present locations of interests where injured persons are detected (e.g., see Figure 1 for an example).

The figure shows a mountain environment with an avalanche between the mountains. In the right corner of this figure is a human team leader who is giving instructions

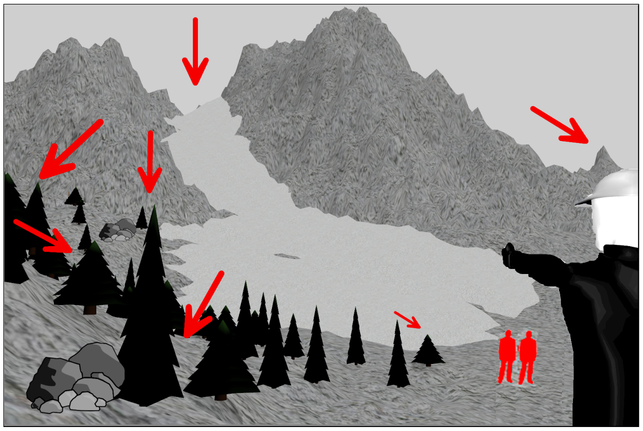

Figure 1: A virtual mountain terrain with a human team leader and his team, some landmarks, marked with red arrows, where possible injured persons have been detected.

based on the landmarks which have to be accomplished by the rescue team. Subjects were informed that their teams did not have any knowledge about the environment and its conditions, so the participants were compelled to consider these while giving instructions.

4. DATA EVALUATION AND DISCUSSION

The results of our study show that the several instructions were composed of referential and visual constructs. Participants applied those constructs in their commands in order to describe how they perceive the entity in the environment. For example the command "Go to the tallest tree" shows that the subject trys to measure the entity by using a description such as "the tallest".

Other results include constructs like geometric spatial relations that are related towards entities in the outdoor environment. An example shows the command "Go to the big tree next to the pile of rocks", where a location is described by the relation of two entities in the world.

To process those results, we have defined and generated rules for syntactically and semantically interpretation of instructions into robotic action descriptions, e.g., by way of a low-level description for the Cognitive Robot Abstract Machine (CRAM) [2, 6]. This machine includes mechanisms for action execution and reasoning, and grounds symbolic expressions from the knowledge representation into the perception. An example to easily integrate those kinds of instructions into low-level descriptions is given by the instruction above *Go to the big tree next to the pile of rocks.*

It can be directly interpreted as a high-level specification and can be easily translated into an action description such as in Algorithm 1.

This description uses essential components of CRAM, called *designators* that describe various parameters and offer several forms of implementations for objects, locations and actions. This machine was extended by components for voice command interpretation to enable human-robot interaction in search and rescue missions.

Algorithm 1 Parametrized Action Description (1)

```
(an action
  (:type move)
  (:to (a region
         (visible
           (a region
             (:name tree)
             (:size big)
             (:next-to
               (a region
                 (:name pile-rocks))))))))
```

5. INTERPRETATION OF NATURAL LANGUAGE TASKS INTO ROBOTIC ACTION DESCRIPTIONS

By integrating those results into our cognition-enabled task interpretation system we are able to translate abstract and qualitative action descriptions into numeric action parameters which the robot's navigation subsystem can use. We use the effect-based action parameterization [6]. The basic idea, which is introduced in [10] is to interpret a qualitative description such as "to big tree next to pile of rocks" into a distribution of positions that satisfy the qualitative description. The execution time reasoning system is then sampling parameters from the distribution in order to find parameters that satisfy additional execution conditions such as the location to be observed not being occluded, the location being reachable with the available power resources, the wind conditions allowing safe flying, and so on. This generative model approach generates sampling-based solutions such as goal positions for moving the robot. The designators introduced in Algorithm 1 will be resolved and analyzed and evaluated by prolog queries such as in [10].

6. CONCLUSION AND FUTURE WORK

In this paper we introduced an experimental setup of an empirical study in a visual search and rescue task application for mixed human-robot teams and presented some of the empirical results. We briefly introduced how high-level instructions can be interpreted into robot plan parameterization actions that can be used for a robot in a simulation-based outdoor environment. Based on the evaluated results and ideas, in the future we will extend our system with natural language processing components in order to resolve complex linguistic and visual descriptions in the instructions by querying heterogeneous knowledge bases in order to clarify goals.

Acknowledgments

This work is supported in part by the EU FP7 project *SHERPA* (grant number #600958)

REFERENCES

[1] M. Beetz, G. Bartels, A. Albu-Schäffer, F. Bálint-Benczédi, R. Belder, D. Beßler, S. Haddadin, A. Maldonado, N. Mansfeld, T. Wiedemeyer, R. Weitschat, and J.-H. Worch. Robotic Agents Capable of Natural and Safe Physical Interaction with Human Co-workers. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, 2015.

[2] M. Beetz, L. Mösenlechner, M. Tenorth, and T. Rühr. Cram – a cognitive robot abstract machine. In *5th International Conference on Cognitive Systems (CogSys 2012)*, 2012.

[3] N. Blodow, D. Jain, Z. Marton, and M. Beetz. Perception and probabilistic anchoring for dynamic world state logging. In *10th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2010, Nashville, TN, USA, December 6-8, 2010*, pages 160–166, 2010.

[4] C. B. (Ferrell). Regulating human-robot interaction using "emotions", "drives" and facial expressions, 1998.

[5] L. Marconi, C. Melchiorri, M. Beetz, D. Pangercic, R. Siegwart, S. Leutenegger, R. Carloni, S. Stramigioli, H. Bruyninckx, P. Doherty, A. Kleiner, V. Lippiello, A. Finzi, B. Siciliano, A. Sala, and N. Tomatis. The sherpa project: smart collaboration between humans and ground-aerial robots for improving rescuing activities in alpine environments. In *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, College Station, Texas, USA, Nov. 5-8 2012.

[6] L. Mösenlechner and M. Beetz. Parameterizing Actions to have the Appropriate Effects. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, USA, September 25–30 2011.

[7] T. Rogers and M. Wilkes. The human agent: a work in progress toward human-humanoid interaction. In *In Proceedings of the International Conference on Systems, Man, and Cybernetics (Nashville, TN). Piscataway: IEEE*.

[8] J. Triesch and C. von der Malsburg. A gesture interface for human-robot interface. In *IEEE Conference on Face and Gesture*, Nara,Japan, 1998.

[9] C. Tsai, S. Hsieh, Y. Hsu, and Y. Wang. Human-robot interaction of an active mobile robotic assistant in intelligent space environments. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. IEEE, Oct. 11-14 2009.

[10] F. Yazdani, B. Brieber, and M. Beetz. Cognition-enabled robot control for mixed human-robot rescue teams. In *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13)*, 2014.