Solving Ambiguities with Perspective Taking

Raquel Ros, Emrah Akin Sisbot, Rachid Alami, Jasmin Steinwender,
Katharina Hamann, Felix Warneken

Solving Ambiguities with Perspective Taking

Raquel Ros[1,2], E. Akin Sisbot[1,2], Rachid Alami[1,2], Jasmin Steinwender[3], Katharina Hamann[3], Felix Warneken[4]

[1]CNRS - LAAS, 7 avenue du Colonel Roche, F-31077 Toulouse, France
[2]Université de Toulouse, UPS, INSA, INP, ISAE, LAAS, F-31077 Toulouse, France
[3]Dep. Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[4]Dep. Psychology, Harvard University, Cambridge, MA, USA
Email: {rrosespi, easisbot, rachid}@laas.fr, {jasmin_steinwender,khamann}@eva.mpg.de, warneken@wjh.harvard.edu

Abstract—Humans constantly generate and solve ambiguities while interacting with each other in their every day activities. Hence, having a robot that is able to solve ambiguous situations is essential if we aim at achieving a fluent and acceptable human-robot interaction. We propose a strategy that combines three mechanisms to clarify ambiguous situations generated by the human partner. We implemented our approach and successfully performed validation tests in several different situations both, in simulation and with the HRP-2 robot.

Index Terms—Human-robot interaction, perspective taking

I. INTRODUCTION

Ambiguities are frequently generated among humans. In general, humans' first attempt to solve ambiguities is to solve them by themselves. If they fail, then they explicitly ask to the partner for more information that will help them clarify the confusion. A robot that interacts with humans must be prepared to handle possible ambiguities by itself, when possible, for two reasons. First, humans are not always aware of ambiguities they create and therefore, they will expect that the robot will be able to solve them internally. And second, a robot that is not able to solve ambiguities by itself would have to constantly inquire the human for clarification which would result in a tedious human-robot interaction.

In this work we propose a clarification strategy to solve possible ambiguities when referring to objects in a face-to-face interaction based on three mechanisms: visual perspective taking, spatial perspective taking and feature based descriptions.

II. CLARIFICATION STRATEGY

Given the scenarios depicted in Figure 1 and the human query "Can you give me the ball?", can the robot autonomously find out which object the human wants? In the first case (Figure 1a) it is clear which ball the human wants, since there is only one available. However, in the second scenario (Figure 1b) any of the two balls could be the one referred. Thus, the robot should ask which ball the human wants. In this work we identify two ways for doing so: using spatial perspective taking, or describing features of the objects [1]. In the former case, the robot could take the perspective of the human and ask whether she wants the ball on
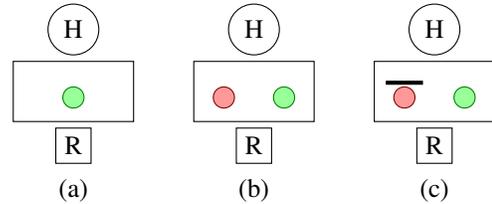
Fig. 1: Scenarios with different degree of ambiguity.

her left or on her right. In the latter case, the color description could be used. Hence, the robot would ask her partner whether the green one or the red one is the one referred. Finally, the third scenario (Figure 1c) depicts an ambiguous situation from the robot point of view, but unambiguous for the human. In this case, there is a visual perspective problem only solvable if the robot is able to infer that the human cannot ask for the red ball because it is occluded from her perspective (24-month-old children are able solve this type of ambiguities [2]). Thus, the only possible ball is the green one.

The most significant related work has been presented by [3] and [4]. However, they focused on visual perspective taking, while we extend clarification with two additional techniques. We briefly describe the implementation of the mechanisms in our system and then the overall strategy to combine them.

A. Visual Perspective Taking (VPT)

The complete model of the environment is known by the robot. To determine if an object is perceived by the human (or the robot) we use 2D perspective projections of the 3D environment (Figure 2a). We first obtain the projection of the isolated object (Figure 2b, the blue box), and we compare it with the "real" projection of the scene which considers occlusions of the evaluated object (Figure 2c, the teddy bear is partially occluding the blue box). We obtain the visibility ratio of the object comparing both images. An object is visible to an agent if the ratio is over a given threshold [5].

B. Spatial Perspective Taking (SPT)

Given the environment, the robot must be able to compute different spatial locations of the objects based on different reference frames (self and human). We divide the space around the referent into n regions, and subdivide them by a radius to model distance, i.e. near and far.
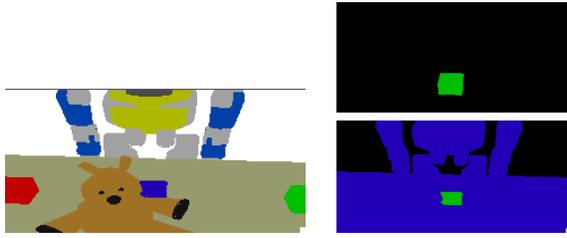
Fig. 2: Left, projections of the human's perspective, right top, object model projection, and right bottom, "real" projection.

C. Feature Based Description (FBD)

Objects have features that allow us to differentiate them from one to another. For instance, color, size, shape, dimensions, etc. Besides, we can also categorize objects in different classes and refer to their class as a descriptor. We have defined a flat database that contains the descriptions of the available objects in the environment. Given a set of objects, we look for a feature where each object has a different value, i.e. we search for a discriminative feature. If found, the robot proposes the different values so the human can indicate the one that describes the referred object.

D. Clarification Process

A decision tree combines the above techniques the following way. Given the human request, first we obtain a list of the potential referred objects. Next, we apply VPT to eliminate the candidates that are occluded from the human perspective. Finally, based on descriptors of spatial locations (SPT) and features (FDB) we search for a discriminant (a feature that has a unique value for each object) and ask the human about its value. If there are still several candidates, the human provides more clues to the robot and the process starts over again. At any moment the process may end if either only one object is available, i.e. the ambiguity has been solved, or when there are no candidates, i.e. the robot cannot solve the ambiguity.

III. VALIDATION TESTS

The whole system is integrated into HRP-2 robotics platform as components of the LAAS architecture [6]. In order to acquire and keep a coherent model of the environment, three main modules are used: the *Object Recognition Module*, detects and localizes objects through markers; the *Human Detection Module*, localizes and tracks the human looking orientation through motion capture cameras; and the *Robot Manager Module*, provides the robot's current configuration. The *Perspective Reasoner* constantly updates environment information and answers to the queries of the *Clarification Module*. The interaction is done through a keyword and a screen. Figure 3 illustrates the overall architecture.

In order to test our system we have designed six different scenarios where ambiguities may easily arise during an interaction. In all cases there is a face-to-face interaction between a human and the robot and a table with objects on it. The layout of the scenarios is not random, and in fact it tries to
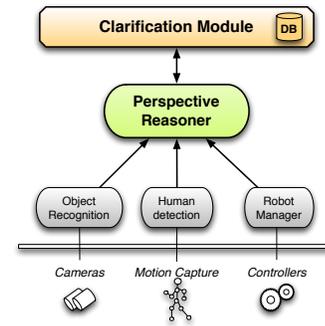

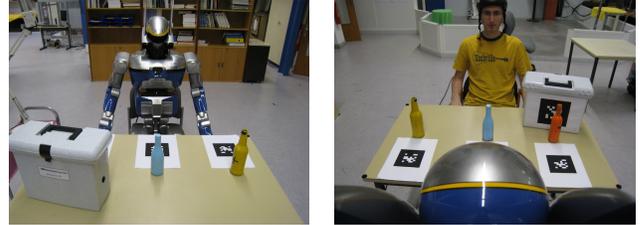Fig. 3: General architecture of the system.

Fig. 4: Scenario with HRP-2, both perspectives.

cover a wide variety of situations where clarification should be performed in different ways based on the source of ambiguity.

We have tested our approach in simulation all scenarios, and the clarification algorithm successfully solved all situations as expected. Additionally, we tested two of the most significant scenarios (the ones with more sources of ambiguities) with HRP-2 in a real environment (Figure 4). The different components of the system responded accurately when running the tests and the ambiguities presented to the robot were successfully solved.

IV. CONCLUSION

Solving ambiguities is essential for a robot to interact with humans that constantly generate and solve them in their daily lives. We have identified three techniques and a strategy to combine them that may allow a robot to clarify ambiguous queries from a human partner. To test the reasoning engine we have defined different scenarios and successfully performed the tests both in simulation and with a real robot (HRP-2).

REFERENCES

[1] M. M. Louwerse and A. Bangerter, "Focusing attention with deictic gestures and linguistic expressions," in *Proceedings of the Cognitive Science Society*, 2005, pp. 1331–1336.
[2] H. Moll and M. Tomasello, "Level 1 perspective-taking at 24 months of age," *British Journal of Developmental Psychology*, vol. 24, 2006.
[3] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schultz, "Enabling effective humanrobot interaction using perspective-taking in robots," *IEEE Transactions on systems, man and cybernetics - Part A: Systems and Humans*, vol. 35, no. 4, pp. 460–470, 2005.
[4] M. Berlin, J. Gray, A. L. Thomaz, and C. Breazeal, "Perspective taking: An organizing principle for learning in human-robot interaction," in *In Proceedings of AAAI*, 2006.
[5] L. F. Marin-Urias, "Reasoning about space for human-robot interaction," Ph.D. dissertation, Universite Paul Sabatier, 2009.
[6] R. Alami, R. Chatila, S. Fleury, M. Ghallab, and F. Ingrand, "An architecture for autonomy," *International Journal of Robotics Research*, vol. 17, pp. 315–337, 1998.