INTERACTING WITH INDUSTRIAL ROBOTS THROUGH A MULTI-MODAL LANGUAGE AND SENSOR SYSTEMS

Batu Akan, Baran Çürüklü, Lars Asplund

Intelligent Sensor Systems
Mälardalen University
Västerås, Sweden
{batu.akan,baran.curuklu,lars.asplund}@mdh.se

Abstract

Over the past few decades the use of industrial robots has increased a company's efficiency as well as strengthening their competitiveness in the market. Despite this fact, in many cases, robot automation investments are considered to be technically challenging as well as costly by small and medium sized enterprises (SME). We hypothesize that in order to make industrial robots more common within the SME sector, the robots should be reprogrammable by task experts rather than robot programming experts. Within this project we propose to develop a high level language for intelligent human robot interaction that relies on multi-sensor inputs providing an abstract instructional programming environment for the user. Eventually to bring robot programming to stage where it is as easy as working together with a colleague

1. Introduction

Robots have become more powerful and intelligent over the last decades. Companies producing mass market products such as car industries have been using industrial robots for machine tending, joining and welding metal sheets for several decades now. Thus, in many cases an investment in industrial robots is seen as a vital action that will strengthen a company's position in the market because such an investment will increase their production rate. However, in small medium enterprises (SME's) robots are not commonly found. Even though the hardware cost of industrial robots has decreased, the integration and programming costs of these industrial robots make them unfavorable for SME's. No matter how simple the production process might be, to integrate the robot, one has to rely on a robot programming expert. Either the company will have to setup a software department responsible for programming the robots or will have to out-source this need. In both cases these financial efforts do not pay up [11].

Also an industrial robot must be placed in a cell that will occupy valuable workspace and maybe operate only a couple of hours a day. It is, thus, very hard to motivate for an SME, which is constantly under pressure, to carry out a risky investment in robot automation. Obviously, these issues result in challenges with regard to high costs, limited flexibility, and reduced productivity.

In order to make industrial robots more favorable in the SME sector, the issues of flexibility has to be resolved. Typically for those SMEs, that have frequently changing applications, it is quite expensive to afford a professional programmer or technician, therefore an optimal human robot interaction solution is strongly demanded. Using a high-level language, which hides the low-level programming from the user, will enable a task expert who has knowledge in manufacturing process to easily program the robot and let the robot to switch between previously learned tasks.

For the past two decades, there has not been a significant leap forward in the world of industrial robot programming. Programming by coding has been the only way to manipulate or control the robot. However recently some higher level programming has been done in some industries. Recently automatic programming of robot has gained interest. Automatic programming systems can be grouped in three categories; Learning systems, programming by demonstration, and instructive systems. In learning systems, the robot learns from user provided examples and self exploration [1, 2, 14]. First the robot watches and observes the user through a range of sensors and then tries to imitate the user. Programming by demonstration is also a common way of tutoring robots for trajectory oriented tasks such as arc welding or gluing [11]. Myers *et al.* used programming by demonstration to teach the robot subtasks which are then grouped into sequential tasks by the programmer [9]. Calinon and Billard presents a probabilistic framework for robot programming by demonstration where the robot

This work described in this paper is supported by *Robotdalen* and *Sparbanksstiftelsen*.

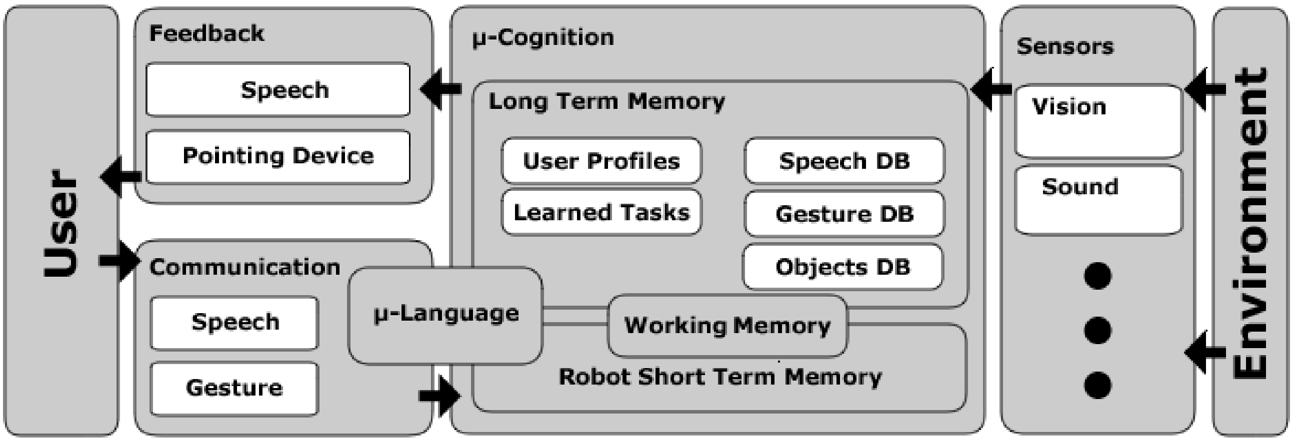

Figure 1. Block diagram of the proposed system.

learns new skills through human guidance [4].

In the case of instructive systems, series of instructions are given to the robot. This type of programming is best suited for executing series of tasks that the robot is already trained for. Using speech in order to instruct a robot provides a natural and intuitive way. Lauria *et al.* uses a speech based natural language input in order to navigate a mobile robot to move to different locations through specified routes [7]. Brick and Scheutz provides an incremental framework where the robot can act as the sufficient information to distinguish the intended referent from perceivable alternatives, even when this information occurs before the end of the syntactic constituent [3]. Hand gestures is also used as input; Voyles and Khosia integrated a gesture based set of commands into programing by demonstration framework [13]. Strobel *et al.* uses static gestures to direct the attention of the robot to a specific part of the scene [12]. Combining the two modalities can even provide improved robustness [6]. McGuire *et al.* makes use of speech and static gestures in order to draw the the robot's attention to the object of interest for grasping [8].

In this paper we propose a system for interacting with a robot. The proposed system consists of two integrated parts: a high-level language for human robot interaction and a cognitive robot architecture for robot knowledge representation. These two parts are highly integrated with each other, which makes this approach unique.

2. System architecture

The intelligent human robot interaction system (iHRI) proposed in this work addresses interaction between a human and a robot in a, for a human, natural way. The system consists of two parts: cognitive robot architecture (μ-cognition) and a high-level language (μ-language) as presented in Figure 1. The μ-language is a multi-modal language, which is used for interaction between a user and robot, whereas the μ-cognition is a cognitive robot architecture, thus a system for robot knowledge representation. These two systems are highly integrated with each other. A user that will interact with a robot will use the μ-language. This means that all information transfered between user and robot will go through this high level language. Note that the language will be designed in a way to allow two-way communication. This means that the robot can give feedback to a user when needed. This is highly important since all forms of interactions, e.g., handling ambiguities, resetting the whole system, and repeating a sequence, must be dealt within the same framework. μ-cognition is used for storing, representing and retrieval of information. The architecture of this system is similar to functional models of human memory function [10]. It consists of sub systems dedicated to specific functions: robot long term (rLTM), robot short term (rSTM), and robot working memory (rWM). These sub systems have different time windows as well as functions. The strength and uniqueness of the iHRI approach comes from that the knowledge representation part and the interaction language are strongly integrated with each other. We hypothesize that these two parts must be present from the early beginning of the system design.

The communication language has to be highly intuitive to use for task experts as well as other engineers involved in a production process. This is one of the main issues that the proposed system has to address. Instructing the robot using the μ-language will be carried out mainly through 4 modalities: (i) hand gestures, (ii) speech, (iii) vision based object recognition. In a latter phase of the project (iv) hand and body postures will be

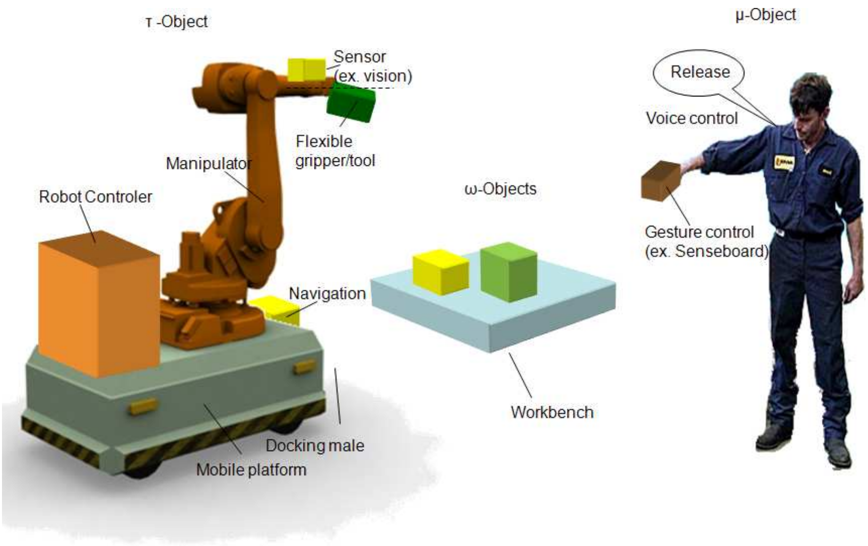

Figure 2. Sample cell environment demonstrating the μ-Object as the user giving commands to the τ-Object, in this case the robot. And a set of ω-Objects on the workbench.

introduced. Note that these modalities are common in human to human communication, thus, the goal is to make interaction to adapt to humans and not the opposite. It will be possible to extend this language with other modalities in the future.

The high level language proposed in this work is equipped with methods for representing and manipulating with objects. In the world defined by the μ-language there are three categories of objects. The μ-objects belong to the most important category, i.e., the instructor. This object can either be a human or a robot. This object is the object that gives instructions to other objects in the world, and hence in a certain time window, there is only one μ-object. Humans or robots that are (actively or passively) listening to the μ-object are defined as the τ-objects. The world itself consists of two conjunctive domains: (i) the robot cell (ii) everything that is outside the robot cell.

In the simplest way of communication, as in Figure 2, the world consists of a human instructing a robot. In this case the human is the μ-object, and the robot is a τ-object. Most likely the robot is in the cell and the human is outside the cell. When the robot gives feedback to the user, e.g. verifying that it understands a command it becomes the μ-object, consequently the user becomes a τ-object. The third category of objects consists of all other objects that can be found in the world. These objects are called the ω-objects, and are grouped under five sub categories: (i) manufacturing parts, (ii) movable tools, (iii) non-movable machines used by robots and/or humans, (iv) working benches, (v) miscellaneous objects.

A conversation using the μ-language can be divided into (or mapped to) semantical units all having a well-defined meaning in the μ-language as well as in a natural language, e.g. “the robot is waiting for additional instructions”, “robot moving closer to object X”, and “robot does not understand”. At a given time more than one such semantical unit can be active. Until now we have defined object types and semantical units of the μ-language. We hypothesize that these two concepts of the μ-language are strongly connected to the μ-cognition module, since this module deals with knowledge representation.

In the prototype that is currently under development, the μ-cognition module consists of a 3-component memory system: robot long term (rLTM), short term (rSTM), and working memory (or working attention) (rWM). The rLTM can be interpreted as a database holding information about the robot and the cell, objects that might be in the cell and their features, and relationships between them. In the rLTM information about previous tasks as well as all users is stored. Thus, during interaction, the user or the robot will be able to recall information from previous experiences. These experiences might be associated with the user or the task. The rLTM is the knowledge base of the robot. On the contrary the rSTM operates in a restricted time domain, and hence, is responsible for the sensor input to the robot and items retrieved from the rLTM. The rSTM operates in theory on 20-30 seconds. These two memory systems are coordinated by the rWM. The rWM is, thus, the master and almost all communication with the robot through the high level language will go through the rWM. In theory the time window of the rWM is restricted to few seconds, and its storage capacity is limited to few facts about objects, such as their relationships, and operations involving objects. In contrast to rSTM and rLTM modules the rWM will produce new facts about the environment. This makes the rWM unique in the proposed system.

We hypothesize that a memory system with limited capacity will play an important role in meeting real-time requirements of the system. As in humans the rWM module and the μ-language are strongly connected to each other. The robot always has to keep track of the semantical units as well as important objects, like the μ-object. This is done in the rWM. Since the rWM coordinated most of the information flow between different parts of the whole system it has to both rely on the other two memory models as well as equipped with methods for "forgetting" information. One of the main challenges of this project is to design algorithm, which is responsible for the function of the rWM module. Note further that the memory system proposed in this is work is not hierarchical.

3. Conclusion and Future Work

In this paper we presented an abstract language model for interacting with industrial robots. The main advantage of the architecture is that it integrates the language and cognitive memory model into one framework. We will determine a well defined grammar for our language and will work on a tight integration scheme with vision[5] and audio sensors. Eventually this high level language will not be only limited to industrial robots but can be used to program other service robots as well.

4. References

[1] G. Biggs and B. MacDonald. A survey of robot programming systems. Australasian Conference on Robotics and Automation, December 2003.

[2] A. Billard and S.Schaal. Robust learning of arm trajectories through human demonstration. pages 734–739. Intelligent Robots and Systems, 2001.

[3] T. Brick and M. Scheutz. Incremental natural language processing for hri. HRI'07, March 2007.

[4] S. Calinon and A. Billard. Active teaching in robot programming by demonstration. International Symposium on Robot and Human Interactive Communication, 2007.

[5] F. Ekstrand, J. Lidholm, and L. Asplund. Robotics for SMEs – 3D vision in real-time for navigation and object recognition. In *To be published in the Proceedings of International Symposium on Robotics 2008*, 2008.

[6] S. Iba, C. Paredis, and P. Khosla. Interactive multi-modal robot programming.

[7] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein. Mobile robot programming using natural languages. *robotics and Autonomous Systems*, 38(3):171–181, March 2002.

[8] P. McGuire, J. Fritsch, J. Steil, F. Roethling, G. Fink, S. Wachsmuth, G. Sagerer, and H. Ritter. Multi-modal human-machine communication for instructing robot grasping tasks. pages 1082–1089. nternational Conference on Intelligent Robots and Systems (IROS), 2002.

[9] D. Myers, M. Pritchard, and M. Brown. Automated programming of an industrial robot through teach-by showing. pages 4078–4083. International Conference on Robotics and Automation, 2001.

[10] D. Purves, G. Augustine, D. Fitzpatrick, W. Hall, A. Lamantia, J. McNamara, and S. Williams. *Neuroscience*. Sinauer Associates, fourth edition, June 2004.

[11] R. D. Schraft and C. Meyer. The need for an intuitive teaching method for small and medium enterprises. International Symposium on Robotics, May 2006.

[12] M. Strobel, J. Illmann, B. Kluge, and F. Marrone. Using spatial context knowledge in gesture recognition for commanding a domestic service robot. pages 468–473. Robot and Human Interactive Communication, 2002, 2002.

[13] R. M. Voyles and P. K. Khosia. Gesture-based programming: A preliminary approach.

[14] Y. Zhang and J. Weng. Action chaining by a developmental robot with value system. pages 53–60. International Conference on Development and Learning, 2002.