The 2016 AAAI Fall Symposium Series:
Artificial Intelligence for Human-Robot Interaction
Technical Report FS-16-01

Me and You Together: A Study on Collaboration in Manipulation Tasks

Miguel Faria[1,2] and **Rui Silva**[1,2,3] and **Francisco S. Melo**[1,2] and **Ana Paiva**[1,2]
{miguel.faria, rui.teixeira.silva}@tecnico.ulisboa.pt, {fmelo, ana.paiva}@inesc-id.pt
[1]INESC-ID, Lisboa, Portugal
[2]Instituto Superior Técnico, Universidade de Lisboa, Portugal
[3]Computer Science Department, Carnegie Mellon University, USA

Abstract

This paper presents an ongoing study in the area of Human-Robot Collaboration, more precisely collaborative manipulation tasks between one robot and multiple people. We study how different trajectories influence people's perception of the robot's goal. To achieve this, we propose an approach based on Probabilistic Motor Primitives and the notion of legibility and predictability of trajectories to create the movements the robot performs during task execution. In this approach we also propose combining legible and predictable trajectories depending on the state of the task in order to diminish the drawbacks associated with each type of trajectory.

Introduction

Robots are now evolving from being used as mere tools that help fulfill a task to entities that work and co-habit alongside humans. In order to achieve such milestone, robots' intentions need to be predictable so as not to confuse, surprise or even scare humans.

This paper presents an ongoing work that focuses on how different kinds of motions affect the transmission of intent during a collaborative manipulation task, between a robot and multiple people. Specifically, the paper studies the effects in both the fluency of the collaboration as well as in the efficiency of the task.

An example of a task we envision consists in having a robot serving as a bartender, with the task of filling a cup of water to one of several customers. In this case, if the robot performs a non-predictable motion, human users may become confused regarding as to whom the robot is serving. On the other hand, if the motion is clear, the chance of mistakes by the users decreases and the fluidity of the task increases.

Formally, the question that drives this work is "In a multiple user collaborative manipulation task, what is the best type of trajectory to perform, in order to minimize confusion between users regarding the robot's target?"

We present an approach that combines two types of trajectories, legible and predictable (Dragan et al. 2015), in order to exploit their strengths and minimize the consequences of their drawbacks. Moreover, the approach adopts a motion representation that is not only suitable for *learning from demonstration*, but also allows for efficient generalization of trajectories to new targets (Maeda et al. 2014).

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The remainder of this document is structured as follows. We start by reviewing related work. We then explore the system's architecture, detailing how the robot decides which cup to serve and the best movement to perform. We finish with some final comments on the work that has been done and what is left to do.

No experimental results are provided since, as of the writing of this paper, the studies are still under way. However, we expect that this work can contribute with more insights regarding how to build effective systems for interactions between people and robots in small spaces without hindering each other's performances.

Related Work

The use of motion learning in manipulation tasks is not new, and has been widely use in several tasks requiring a precise control of the robot's pose. For example, the work of (Pais et al. 2013), they propose learning from demonstration using kinesthetic teaching, in which they learn the model for the robot's movement in a specific task and the constraints it needs to respect in order to be successful.

In this work, the motion decision by the robot is performed using a framework called Probabilistic Movement Primitives (ProMP) (Maeda et al. 2014), which allows a robot to learn how to perform a manipulation task by observing a set of demonstrative trajectories for the same task and then extrapolate the correct way to move. ProMP are an alternative to Interaction Primitives (IP), defined in (Amor et al. 2014), that could not combine multiple movement primitives and as such could not adapt very well to changes during task execution. The ProMP allow the combination of multiple movement primitives since they create a probability distribution over all the recorded trajectories. Therefore, in the prediction step it is possible to combine information from various primitives, obtaining a better movement trajectory. In the same work Maeda et al. also show how this framework behaves during a collaborative task of assembling a box. Using ProMP the robot could easily adapt to changes in the conditions of the task, like the box flipping over.

One important aspect in any collaborative task, either among humans or between humans and robots, is the easy understanding of intention without the need to ask the other what they are doing (Strabala et al. 2012). This aspect led to the study of how to convey intention through the way we move during the execution of a task, the way we approach an object or our pose while we perform a task. There have been various works about how to make robots express the intent of a movement more clearly while they perform a task (Strabala et al. 2012; Breazeal et al. 2005; Jung et al. 2013; Gie 2013). Most of the results in these studies are based on animation and creating human-like movements based on animation principles (Gie 2013); and on the study of how people physically communicate intention prior to giving an object to another one or to a physical interaction (Strabala et al. 2012).

The most recent studies have shown that different ways to create a trajectory yield different results in the perception of motion intention by a human partner. They show that a movement that is purely efficient, where efficiency is concerned with energy usage and collision avoidance, is not always the best movement to convey intent and can sometimes scare people. A better way to transmit intention was shown to be movements that are not purely efficient, but that encompass a certain smoothness and also an earlier deviation towards the motion objective, allowing this way for the human partner to understand more easily what the robot is doing and why, and react more accordingly.With these results, Dragan et al. have defined the notions of legibility and predictability of a robot's motion and how a robot can plan trajectories that follow those notions (Dragan and Srinivasa 2013; Dragan, Lee, and Srinivasa 2013). In more recent work, Dragan et al. investigate the effects of predictable and legible motions in the perception of people about the objective of the robot's movement during collaborative tasks (Dragan et al. 2015). They show that when the human partner does not have previous knowledge about the robot's target, legible motions are better at conveying the movement's intent. On the other hand, when the human has an idea of what the robot's objective is, predictable motions are better.

System Design

In order to investigate the impact of different kinds of motions in the human perception of the robot's target in collaborative manipulation tasks, we use the notions of legible trajectories and predictable trajectories, defined by Dragan et al. in (Dragan, Lee, and Srinivasa 2013), in a system that would interact with multiple people at the same time. The system is integrated in a Baxter robot that will perform the collaboration task with the humans.

The system is composed by three main modules: one module is responsible for processing visual data in order to get information about the workspace; one module is responsible for social interactions with the human partners in the collaboration task and the last module is responsible for determining the current objective of the robot and generating the arm movement to fulfill the chosen objective. Figure 1 depicts the architecture of the system as a whole, as well as the data flow between the different modules in the system. The communications between different modules is done using Robot Operating System (ROS) topics since this allows the modules to alert the interest parts that there are updates,

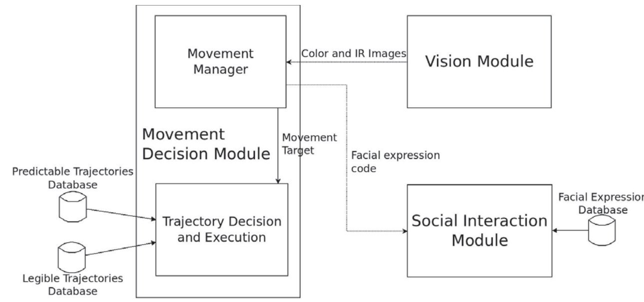

Figure 1: System's architecture and interactions between each of the system's modules (vision, movement decision and social interaction) and their sub-modules.

whilst simultaneously performing their tasks. Communications inside a module are done using ROS services because these typically involve requests for a sub-module to perform a specific task without which that module cannot continue its task.

The vision module receives data from a Kinect V2 camera, both color and infrared (IR) frames, and is responsible for identifying the possible objectives in the scene. In our case the objectives are cups that are carried by the human participants. These cups are color-coded and the vision module segments the received color images, looking for the colors of each cup. The segmentation process uses the HSV color space as it allows us to adjust the color parameters only by focusing in finding the smallest hue interval that only identifies the cups. This because the other two parameters of HSV, saturation and value, are related only with lighting conditions and by keeping those conditions the same, the segmentation process becomes a lot more straightforward.

The social module is responsible for making the interaction more natural and human-like. As shown in (Mumm and Mutlu 2011), when humans and robots need to interact in close proximity, the closer a human psychologically feels to a robot the better the communication between them is. This leads to a better collaboration between both parts. Because of this, the social component is really important for a collaboration task to be performed efficiently and in the system developed the social module takes cues from the movement decision module regarding changes in facial expressions and speech interactions that better adapt to the task progression. These social interactions occur by displaying happy expressions when a movement ends successfully, displaying a sad expression when a problem occurs and needs correcting or using speech interactions to correct the human partners if they are acting wrongly.

The movement decision module is the system's central module and is responsible for deciding when to select a next objective and which is the next objective. Besides this, the module is also responsible for communicating with the social module to signal when new interactions are necessary. In terms of the movement decision, the module chooses for next objective the closest one among all remaining and if there is a draw between two or more objectives then the system chooses one randomly. With the target selected, the system uses the Collaborative Probabilistic Movement Primitives (CoPMP) model, (Maeda et al. 2016) - created based on the trajectories demonstrated to the robot for the predictable trajectories or for the legible trajectories, depending on if he robot is performing predictable or legible trajectories - to extrapolate the best movement to achieve the given objective by combining the information available from those demonstrated trajectories. The resulting trajectory, created by the CoPMP model, is then executed using the Joint Trajectory Action Server available in the Baxter ROS SDK.

Our solution, as explained previously, has two modes of functioning: using only predictable trajectories or using only legible trajectories. This difference is important to evaluate whether a significant difference exists in the human perception of the robot's objective or not depending on the type of movement performed. It also allows us to investigate whether the results of Dragan et al (2015) also stand in settings involving collaborative manipulation tasks with multiple people.

The use of pure legible and predictable motions, as shown in (Dragan et al. 2015), works really well in conveying intentions during a collaborative task between one person and one robot. However using only one of these types of motions has some drawbacks: for example when using only legible motions and there is no ambiguity regarding the robot's objective, legible motions do not bring significant improvements over predictable trajectories. However it does bring extra expense of energy and takes more time performing the task than while using predictable trajectories. Predictable motions, as shown by Dragan et al., lack the explicitness that legible motions have when addressing cases of ambiguity between possible objectives for the robot. Due to those drawbacks we devised an approach that tries to combine the advantages of both trajectories, while reducing their drawbacks.

In this approach we use both predictable and legible trajectories during the collaborative task and it is the system's responsibility to decide if it should perform a predictable trajectory or a legible trajectory. In order for it to do that, we devised a set of rules that, although somewhat specific to our task, can easily be generalized to other collaborative tasks where the robot must manipulate an object. Based on the results from (Dragan et al. 2015) and observations of human-human interaction, we designed the following set of rules:

- if the selected objective is closely surrounded by two or more other of other objectives, a more direct movement (predictable movement) is preferable than a more open movement (legible movement);

- if the selected objective only has objects on one side then a movement that approaches the side the objective from the side with no objects is better than a more direct one, a legible movement is preferred to a predictable one;

- if there is only one remaining objective or if there is no ambiguity regarding possible objectives then a predictable movement is preferred to legible movements.

Experimental Setup

In this work we intend to answer the question *"In a multiple user collaborative manipulation task, what is the best type of trajectory to perform, in order to minimize confusion between users regarding the robot's target?"* and as such we created a system that will use the notions of legibility and predictability combined with a learning framework that allows a robot to better adapt to possible changes in the workspace configuration.

In order to test our system, we use the task of pouring water into the cup of three people. This task allows a comparison of the performance of different motion patterns (predictable, legible or a combination) in multi-user collaborative tasks.

Our experiment thus reproduces a cafeteria-like scenario, in which people approach a bartender (the robot) and wait to be served before moving away. Like so, the participants will be instructed to approach the robot at the start of the experience as if they were to ask it to fill their cups and the robot will sequentially fill the cups. The experience will be repeated three times for each group of three participants, one for each of the approaches used to generate movement: predictable, legible and combination of both, with the order in which the robot starts filling the cups randomized so as not to create a pattern and deviate the results.

At the beginning of each interaction the robot will greet them, then it will wait until it recognizes the cups and fills them and when all three cups are filled, the robot will say goodbye. At the end of each interaction the participants will be asked to fill in a questionnaire, with which we ascertain what the participants felt about the robot regarding the intelligence of the robot and the likability and also if they noted any difference between the movements performed by the robot. Besides the questionnaires, the entire experience will be recorded and analyzed afterwards in order to discover how long each took participant, under each condition, to understand who the robot was directing its action towards, and if there were any that misunderstood the robot's objective.

Final Remarks

As stated in the beginning of this paper, this is an ongoing work and as such we do not have any results at the time.

Nevertheless, we are confident that with this work we will create a system that allows for better collaboration in tasks that require manipulation of objects in close proximity with robots, by making the robot's actions more understandable and its goals more explicit. Mostly, we expect that our combination of predictable with legible trajectories will reduce some of the shortcomings that each type of trajectories has by itself and that the use of learning instead of planning will allow the robot to better adapt to fast changes that occur during the task execution.

Acknowledgment

This work was partially supported by national funds through the Portuguese Fundação para a Ciência e a

Tecnologia under project UID/CEC/50021/2013. (INESC-ID multi annual funding) and the Carnegie Mellon Portugal Program and its Information and Communications Technologies Institute, under project CMUP-ERI/HCI/0051/2013. The second author acknowledges the PhD grant SFRH/BD/113695/2015.

References

Amor, H. B.; Neumann, G.; Kamthe, S.; Kroemer, O.; and Peters, J. 2014. Interaction Primitives for Human-Robot Cooperation Tasks. *International Conference on Robotics and Automation* 2831–2837.

Breazeal, C.; Kidd, C. D.; Thomaz, A. L.; Hoffman, G.; and Berlin, M. 2005. Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (October 2015):383–388.

Dragan, A. D., and Srinivasa, S. S. 2013. Generating Legible Motion. *Proceedings of Robotics: Science and Systems Conference (RSS 2013)* NP.

Dragan, A. D.; Bauman, S.; Forlizzi, J.; and Srinivasa, S. S. 2015. Effects of Robot Motion on Human-Robot Collaboration. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15* 1:51–58.

Dragan, A.; Lee, K.; and Srinivasa, S. 2013. Legibility and predictability of robot motion. *International Conference on Human-Robot Interaction* 1:301–308.

2013. Generating human-like motion for robots. *The International Journal of Robotics Research* 32(11):1275–1301.

Jung, M. F.; Lee, J. J.; DePalma, N.; Adalgeirsson, S. O.; Hinds, P. J.; and Breazeal, C. 2013. Engaging Robots: Easing Complex Human-Robot Teamwork with Backchanneling. *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13* 1555.

Maeda, G.; Ewerton, M.; Lioutikov, R.; Amor, H. B.; Peters, J.; and Neumann, G. 2014. Learning Interaction for Collaborative Tasks with Probabilistic Movement Primitives. In *International Conference on Humanoid Robots*, 527–534.

Maeda, G. J.; Neumann, G.; Ewerton, M.; Lioutikov, R.; Kroemer, O.; and Peters, J. 2016. Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks. *Autonomous Robots* 1–20.

Mumm, J., and Mutlu, B. 2011. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*, 331–338. ACM.

Pais, L.; Umezawa, K.; Nakamura, Y.; and Billard, A. 2013. Learning robot skills through motion segmentation and constraints extraction. In *HRI Workshop on Collaborative Manipulation*.

Strabala, K.; Lee, M. K.; Dragan, A.; Forlizzi, J.; and Srinivasa, S. S. 2012. Learning the communication of intent prior to physical collaboration. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* 968–973.