Towards Responsive Humanoids: Learning Interaction Models for Humanoid Robots

Heni Ben Amor AMOR@TU-FREIBERG.DE
Erik Berger BERGERE@STUDENT.TU-FREIBERG.DE
David Vogt VOGT3@STUDENT.TU-FREIBERG.DE
Bernhard Jung JUNG@TU-FREIBERG.DE
Institute of Computer Science, Technical University Bergakademie Freiberg, Germany

Abstract

This extended abstract presents our ongoing work on deriving interaction models for humanoid robots. The approach differs from earlier interaction learning approaches in that it learns a model from a recorded rapport between two human individuals. Extracted models are used to generate the behavior of humanoid robots.

1. Introduction

Human communication and interaction is based on highly interesting social and physical protocols. The discovery of the "chameleon effect" (Chartrand & Bargh, 1999), for example, showed that humans often unintentionally mimick another person's body posture during a rapport. Behavioral clues such as body language and alignment are vital for the success and the quality of human interactions. This is not limited to communication scenarios, but is also true in physical interaction and cooperation which involves close contact between the interaction partners, e.g. shaking hands or handing over an object. However, until now, most humanoid robot systems do not adhere even to basic social and physical protocols and, in particular, do not take the interaction partner's pose and body language into account. As a consequence, the interaction is often deemed unnatural. Imitation learning (Billard et al., 2008) could be a possible solution to this problem: the robot could learn to engage in natural interactions, by observing the way humans do this. Unfortunately, most imitation learning approaches have focused on learning a set of motor tasks such as walking (Chalodhorn et al., 2007), drumming (Schaal, 2003), or even pancake flipping (Kormushev et al., 2010). In all of these behaviors no interaction with human partners takes place.

To overcome this problem, we present ongoing work on an interaction learning approach that is based on the observation and reproduction of interpersonal communication. To this end, the interaction between two persons is recorded via a low-cost motion capture device. The recorded data is then used to automatically extract an interaction model using manifold learning techniques. The interaction model can then be used by a robot to interact in a natural way with a human partner. It is important to emphasize, that our approach differs from earlier interaction learning methods in that it extracts all necessary information (offline) from an observed rapport between two humans.

2. The Approach

The goal of our approach is to record and analyse a natural interaction setting between two (or possibly more) humans. By analysing the recorded data we hope to extract information which allows a humanoid robot to replace one of the human interaction partners. In other words, the robot tries to mimick the behavior of one of the two interaction partners. To achieve this, three different tasks need to accomplished. First, an model of the recorded interaction needs to be learned. At the current state of our research, the learned model is limited to body postures. Once it is learned, we can query the interaction model in the following way: "Given the current posture of the human, what is the most likely posture that the robot should take on?". In the opposite direction, we can also pose the query:"If the robot takes on this posture, what is the most likely posture that the human will take on?". A second important task is fitting the human motion to the robots body and dynamics, i.e. solving the correspondence

Appearing in *Proceedings of the ICML Workshop on New Developments in Imitation Learning*, Bellevue, WA, USA, 2011.

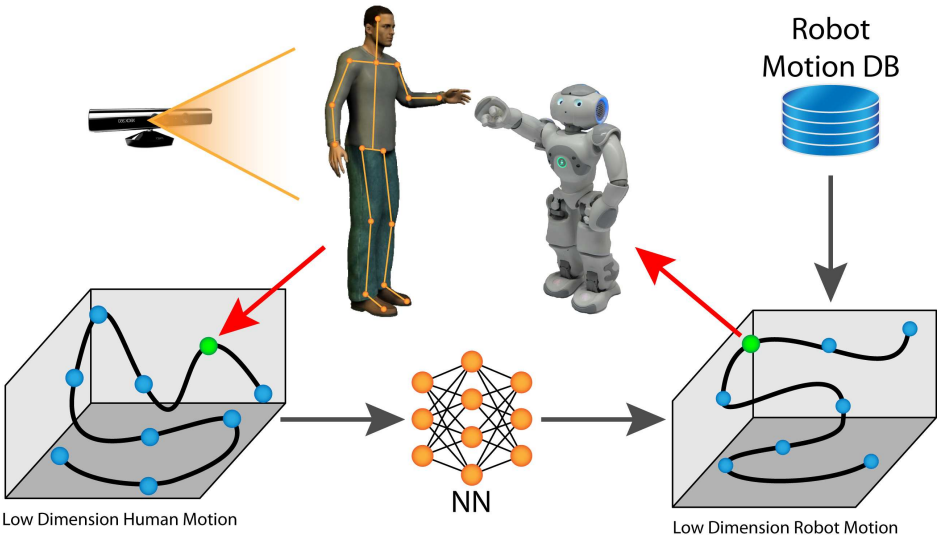

Figure 1. Applying a learned interaction model: First, the posture of the human ist captured using a Microsoft Kinect camera, leading to joint angle data. The data is projected into a low-dimensional space which was earlier learned. The projection yields a low-dimensional point. Using a neural network the corresponding point in the low-dimensional space of the robot is found. The point reflects a "response" posture that the robot should take on. The two low-dimensional spaces and the neural network mapping are learned from an earlier rapport between two humans.

problem. Finally, we want to be able to perform modifications on the recorded responses, e.g. exaggerating a motion, or adding secondary behaviors (blinking eyes). In the following sections we will give a brief overview of our approach which solves the aforementioned questions.

2.1. Learning Interaction Models

In order to create a model for human interactions we first record all body joint angles of two humans. This is done using a low-cost motion capture device, namely the Microsoft Kinect camera. The camera produces a depth image of the scene, in which we detect two people and calculate the configuration of the joints (24 joints). The recorded joint angles for each of the humans if then processed using dimensionality reduction separately. Our implementation supports a variety of dimensionality reduction techniques including Isomap (Tenenbaum et al., 2000), Locally Linear Embedding (Roweis & Saul, 2000), or Principal Component Analysis.

Applying dimensionality reduction transforms the high-dimensional joint angle data into a low-dimensional trajectory. The space in which this trajectory is embedded is called the low-dimensional posture space. As a result of the reduction step, we have for each timestep a point in the low-dimensional posture spaces of the first and the second interaction partner. In order to derive an interaction model, we learn a mapping which allows us to find for each point in the low-dimensional posture space of the first human an appropriate ("response") point in the low-dimensional space of the second human. Such a mapping can easily be learned using a neural network. The network is trained with the projected points of the first human as input and the points of the second human as output, and vice versa.

Later, when a human interacts with our robot (see Figure 1) the learned interaction model is used to synthesize the body postures of the latter. For example, when the human hands something over, the robot needs to time its motion very carefully in order to grasp the object exactly at the right moment. This depends highly on the posture of the human counterpart. To create the desired robot motion we acquire the current human joint angles and calculate the related low-dimensional point in the previously learned human posture model. This is done by deploying a kd-tree search algorithm. Once, the point is found, we employ the learned neural network mapping function, to find the corresponding response point in the low-dimensional space that originally was derived from the motions of the second human. The found point is then projected back, in order to generate the joint angles that the robot should take on.

While the execution of a human motion is in progress we continue to move along the projected trajectory in the first low dimensional model. This makes it possible to predetermine the human posture and create the robot pose in advance.

2.2. Optimization of Recorded Motions

As described earlier, recorded motions of humans can generally not directly be replayed by a humanoid robot platform. This is due to the difference in size, physiology, dynamics, and other parameters. In our particular case, the shoulder of a human has three and the elbow just one degree of freedom, while the employed NAO robot platform has two in both. Therefore, we need to find a mapping between the joint angles of the human and the robot, which ensures a stable and meaningful reproduction of the humans movements. Another problem ist that the motion radius of a joint differs between human and robot. This results in the inavailability of some postures.

To solve this correspondence problem, we use optimization algorithms, namely evolutionary algorithms (see Figure 2). For an efficient use of these algorithms we reduce the number of paramaters to restrict the size of the search space and thus the runtime of optimization. This is achieved using dimensionality reduction techniques. Then, we synthesize our motion by specifying a trajectory using just a few control points inside the low-dimensional space. Each trajectory is then tested within a simulator (called NaoSim), to determine if it produces a stable motion, i.e. the robot does not fall over and generally produces smooth movement. If the later condition is not met, then the low-dimensional trajectory is slightly varied using the evolutionary algorithm and, again, evaluated. At the end of the optimization process, we have a modification of the original human motion, which is suited for application on the real robot.

The fitness evaluation in the above optimization process can be performed in two ways: (1) For goal-directed motions, the fitness function can be specified using a mathematical expression (e.g. the smoothness of the movement). (2) For the other behaviors, such as gestures, it is possible for the user to give positive or negative feedback.

For more information on the optimization of motions inside a simulation environment, please refer to the following paper (Ben Amor et al., 2009).

2.3. Animation Filter and Emotions

It is desirable for the robot not only to be able to reproduce a response, but also to be able to modulate it based on other parameters, such as it's current emotional state. In our approach this is realized using different filters that modify the low-dimensional trajectory. The basic ideas of these filters is derived from the "principles of animation" (Thomas & Johnston, 1995), which are widely used in the animation industry to increase the realism and attractiveness of a cartoon character.

A complete description of these filters is beyond the scope of this abstract, but the following example illustrates the basic principle: *exaggeration* is one of the key elements in animation as it can make the intention behind a particular movement easier understandable. We have discovered, that exaggeration can be achieved by scaling the low-dimensional trajectory corresponding that is used to generate the joint angles of the robot.
Also emotions can be embedded with different filters. When creating a so called sadness-filter, the eye color of the robot turns blue, the shoulders are more likely to hang down and the head is slightly facing downwards. In doing so any learned interaction model can be played back with different emotional additions.

3. Conclusion

In this extended abstract we presented ongoing work on learning models of interactions based on recorded motion capture data. The approach takes a recorded rapport between two people as input and produces an interaction model. The model can later be used to replace one of the human interaction partners by a humanoid robot. The goal of this research is to produce responsive robot behavior that can mimick the interaction style of a human person by observing his or her rapport with another person. So far, we have already finished the basic implementation of the learning system and developed a program that can apply a movement recorded with a Kinect camera onto the robot[1]. We hope to present the working system and other results at the ICML Workshop.

References

Ben Amor, H., Berger, E., Vogt, E., and Jung, B. Kinesthetic bootstrapping: Teaching motor skills to humanoid robots through physical interaction. In Mertsching, Bärbel, Hund, Marcus, and

[1] see http://www.youtube.com/watch?v=GM7v-hZCoHg

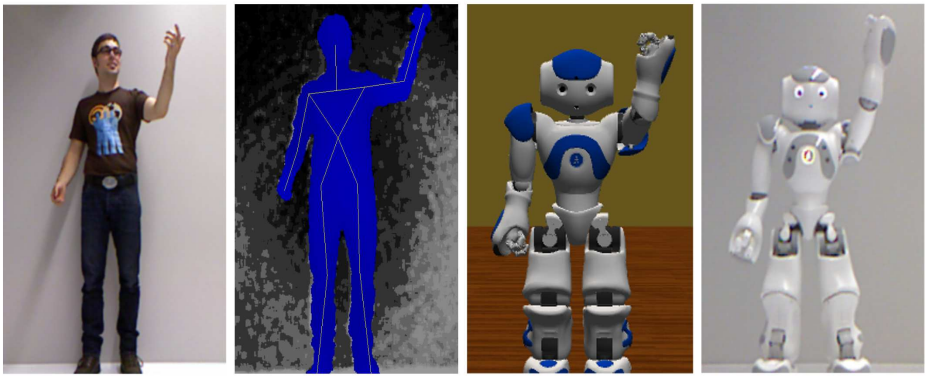

Figure 2. Solving the correspondence problem: (1) The RGB image of the Kinect camera. (2) The human skeleton as derived from the camera. (3) The ideal robot posture is optimized using evolutionary algorithms in a simulator. (4) The optimized posture as applied on the robot.

Aziz, Muhammad Zaheer (eds.), *KI*, volume 5803 of *Lecture Notes in Computer Science*, pp. 492–499. Springer, 2009. ISBN 978-3-642-04616-2.

Billard, A., Calinon, S., Dillmann, R., and Schaal, S. Survey: Robot Programming by Demonstration. In *Handbook of Robotics*, volume chapter 59. MIT Press, 2008.

Chalodhorn, R., Grimes, D. B., Grochow, K., and Rao, R. P. N. Learning to walk through imitation. In Veloso, Manuela M. (ed.), *IJCAI*, pp. 2084–2090, 2007.

Chartrand, T. L. and Bargh, J. A. The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893–910, 1999. doi: 10.1037/0022-3514.76.6.893.

Kormushev, P., Calinon, S., and Caldwell, D. G. Robot motor skill coordination with em-based reinforcement learning. In *Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, pp. 3232–3237, Taipei, Taiwan, October 2010.

Roweis, S. and Saul, L. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290: 2323–2326, 2000.

Schaal, S. Movement planning and imitation by shaping nonlinear attractors. In *Proceedings of the 12th yale workshop on adaptive and learning systems*, 2003.

Tenenbaum, Joshua B., de Silva, Vin, and Langford, John C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319 – 2323, 2000.

Thomas, F. and Johnston, O. *The Illusion of Life: Disney Animation*. Hyperion Books, New York, 1995.