High five! Towards the development of reaching with CNNs on a soft robot arm

Brice Denoun[1], Massimiliano Patacchiola[2], Martin F. Stoelen[2] and Angelo Cangelosi[2]

Manipulation and locomotion are the first modalities that a child uses to learn the structure of her own body [1]. Reaching and grasping are used by the child for several purposes, including learning the structure of her body, and to better understand the environment. To achieve these tasks children use mainly two senses, sight and touch. It has been shown that learning to reach occurs in different stages during development [2]. One characterizing feature of this development is the *motor babbling* phenomenon. It is similar to speech babbling but it happens with gestures, where the child attempts exploratory and approximate movements many times, and learns from the resulting change to body and environment state [3]. Here we attempt to enable a soft robot arm, the GummiArm, to learn a 'high five' behavior, based on motor babbling and an emerging body knowledge with self/other distinction. The GummiArm is a bio-inspired soft robot arm which can be 3D printed [4]. It consists of a basic rotating head, which is a Kinect device, and an arm with 7+1 degrees of freedom, 5 of which are variable-stiffness (Fig 1). Since the GummiArm can be highly passively compliant, this platform is quite suitable for spatial exploration. Young children can for example incorrectly approximate the distance between their body and their goal, resulting in an impact with the target. We wanted to have the same type of behavior as a child who is trying to reach and touch the hand of her parent. When the arm reaches its goal, it makes a 'high five'. We are therefore applying the motor babbling appearing in the third stage of reaching learning.

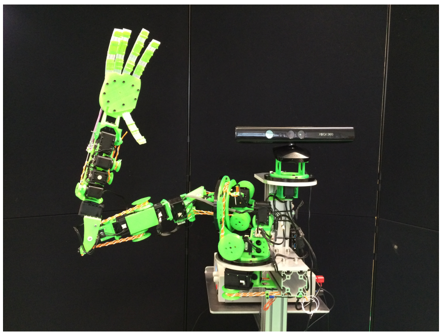

Fig. 1. *GummiArm ready to High Five*

With children, the wide variety of movements drives a reinforcement-type learning. Here, the movements will be used to generate a mapping from the visual hand pose to joint pose. While the arm is moving, the Kinect device will take pictures of the arm, and we will then try to match each position with joints positions and joint forces. A child's brain is able to differentiate and recognize properly hands from faces, which are two main developmental targets. Therefore, we decided to integrate both face and hand detection in our model. We implemented a Convolutional Neural Network (CNN) [5] similar to LeNet5 [6]. We used TensorFlow under Python and an NVIDIA GeForce GTX980 graphics card. The CNN had 3 convolutions layers followed by a pooling layer and a normalization layer, a fully connected layer and an output layer for a total of 11 layers and 4.5×10^6 weights.

We gathered pictures from four databases to create one containing 13100 images. Half of the images were of opened hands, and the other half of faces. We then split our database in three sets: training set, containing 60% of the database, as well as validation and test sets containing 20% of the entire data set each. Preliminary results showed an accuracy of 99.6% on the validation set, and 99.7% on the test set.

These last results are promising, although they are of opened hands with a uniform background. Therefore, we would like to push forward the recognition of hands by creating a classifier which can recognize hands even with complex backgrounds. Then we aim to do the same for the robot's hand during motor babbling. We hope to have a robust 'high five' behavior once the human hand can be robustly localized, and the correct arm joint pose generated.

[1]National Institute of Applied Sciences of Lyon, France
[2]Centre for Robotics and Neural Systems, Plymouth University, UK

References

[1] A. Cangelosi, M. Schlesinger, and L. B. Smith, *Developmental robotics: From babies to robots*. MIT Press, 2015.

[2] R. J. Gerber, T. Wilks, and C. Erdie-Lalena, "Developmental milestones: motor development," *Pediatr Rev*, vol. 31, no. 7, pp. 267–276, 2010.

[3] D. Caligiore, T. Ferrauto, D. Parisi, N. Accornero, M. Capozza, and G. Baldassarre, "Using motor babbling and hebb rules for modeling the development of reaching with obstacles and grasping," in *International Conference on Cognitive Systems*, pp. E1–8, 2008.

[4] M. F. Stoelen, A. Cangelosi, and F. Bonsignorio, "Co-exploring actuator antagonism and bio-inspired control in a printable robot arm," *The 14th International Conference on the Simulation of Adaptive Behavior*, 2016.

[5] H. Cecotti and A. Graser, "Convolutional neural networks for p300 detection with application to brain-computer interfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 433–445, 2011.

[6] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, *et al.*, "Comparison of learning algorithms for handwritten digit recognition," in *International conference on artificial neural networks*, vol. 60, pp. 53–60, 1995.