

Deep Learning for Hand Motion Representation

About Gleechi:

Gleechi is a Stockholm-based startup that have developed the first software to make it possible to animate hands that can move and interact freely and realistically in games and Virtual Reality. The technology is based on 8 years of robotics research, and the company now has customers including one of the top 10 largest VR developers in the world as well as a world-leading automation company. Gleechi has received several awards, including Super Startup of 2015 by Veckans Affärer and ALMI Invest and Winner of the European competition EIT Digital Idea Challenge 2015.

Video demo: <https://www.youtube.com/watch?v=xkCt17JHEzY>

Introduction:

With the recent growth of virtual reality (VR) applications, there is a demand to create highly immersive environments in which the avatar that the user embodies reflects any kind of actions in the virtual world as precise as possible. The major action humans use for interacting with the world is grasping of objects with their hands. Until now, the visual representation of grasping in VR has been resolved by very simple means only, such as attaching a rigid hand to the object that does not adapt to the shape, or manually animating a sparse set of grasps for pre-defined objects, or just not showing hands at all. Initial experiments have shown that hands that are too human-like and do not match the the players' expectations in appearance or behavior, often leads to a loss of the feeling of presence (i.e. making the players feel they are not really in the game). The effect is closely related to the "Uncanny Valley" effect, which refers to when features look and move almost, but not exactly, like natural beings or fits to users' intention, it causes a response of revulsion among the observers.

Description:

Gleechi provides a software solution called VirtualGrasp which makes it possible to animate natural looking grasping interactions in real-time based on the constraints of the virtual world (such as shape of objects, kinematics of the hand, etc). This solution is not a hand tracking algorithm, but a tool that animates a given hand model. In VR applications, an important measure of success for such a system is to create hand and finger motions that both satisfy the physical constraints placed by the object, and are natural and realistic to the human eyes. The first is easy to measure, the second however is difficult to achieve. We believe a data-driven approach exploiting machine learning techniques is a good solution to quantify the "realism" and "naturalism" of the grasps. Such an approach also provides a foundation to synthesize grasps that satisfy user's intention when interacting in the virtual world.

Recently machine learning technique that exploits the deep structure of neural networks has achieved significant progress towards many practical industrial problems. In the context of modeling 3D human motion, deep neural network (DNN) has been successfully applied to represent the spatial-temporal structure of the skeletal pose and motion, and can be used for both action classification and motion prediction and generation [1][2]. The goal of this thesis is to exploit DNN for the purpose of representing human hand grasping and interaction motions.

Tasks:

- Summarize state-of-the-art of deep learning study aimed at modeling human 3D motion, and evaluate which network structure(s) are most suitable for hand motion representation.
- Collect training database from human subject when grasping/interacting with different objects.
- Implement modeling and training of DNNs, using *Caffe deep learning framework* [3], in C++.
- Test, optimize and evaluate the implemented process using the database.
- Summarize and discuss the findings in a report / thesis.

Supervisor at Gleechi: Dr. Dan Song

References:

- [1] H. Liu and T. Taniguchi. Feature extraction and pattern recognition for human motion by a deep sparse autoencoder. In IEEE International Conference on Computer and Information Technology, 2014
- [2] Judith B. Michael J. Black et al. Deep representation learning for human motion prediction and classification. CVPR 2017
- [3] Y. Jia, E. Shelhamer, et al. Caffe: Convolutional architecture for fast feature embedding. 2014.

Application info:

Last apply date:	2017-07-31
Project work period:	Estimated to be 2017 Sep - 2018 Feb
Assignment type:	Degree project
Credits:	30 hp
How to apply:	Please email us your CV, transcript and an one-page personal letter.