Using Deep Learning to Predict User Intention in Virtual Environments

Advisor:

Vincent Guigue email: vincent.guigue@lip6.fr web : http://www-connex.lip6.fr/~guigue/
Gilles Bailly email: gilles.bailly@isir.upmc.fr web: www.gillesbailly.fr

Keywords

Machine Learning, Deep Learning, Human-Computer Interaction, VR, User Intention

Summary

The goal of this internship is to build, implement and evaluate an algorithm to predict user intention in a virtual environment. The result will be used to control a robot to provide rich haptic feedback in Virtual Environment.

Description

VR context. While Virtual Reality (VR) successfully simulates the visual and auditory senses, it is not the case for the sense of touch (haptic). Indeed, haptic technology is still poor compared to the rich ways humans can interact with their environment.

We are currently building a robotic system. This system displaces physical objects to dynamically overlay the virtual ones. Users do not need to use controller and can directly touch the target object. However, we need to know in advance which the user is likely to interact with.

Challenge. We need to elaborate a **user intention model** to know where to move the robot before the interaction occurs.

Approach. The approach consists of elaborating a Deep Learning model. Given the user's movements, the model will predict the next object the user will interact and where she will touch it.

Historically, these approaches have mainly relied on simulation for human behavior and gestures. The predictive aspect is tackled by geometric modelling of gaps between actors and movement speeds. The idea of this internship is to address this issue from the data point of view. Indeed, most inputs in VR can be seen as multivariate time-series, tracking bodies, hands or even fingers over time. As far as the output are concerned, the different actions, choices, ways of picking up an object are all classification problems.
Thus, we are going to exploit recent advances in deep learning to learn a relevant representation of the multi-variate time series [1, 2]. This first step will enable us to build a strong baseline on the target tasks but this internship should enable us to tackle two more ambitious objectives:
(1) understanding the different categories of motions and the related contextual factors: *which individual prefers which type of gesture in which situation?*
(2) This in-depth understanding of the signals should allow us to consider a transfer of models between different motion capture databases in the same way as is done with language models.
The first objective requires disentangling and generative architectures on which we are currently working [3]. Regarding the second target, the idea is to rely on an unsupervised framework that would be able to capture relevant aspect of signals in large datasets. Then, those features & classical temporal patterns will be an asset to be more efficient on small data use-cases [4].

We will work with the student to (1) identify the relevant features, (2) design the model and (3) evaluate the model. We anticipate that this work will lead to a publication in a conference in Human-Computer Interaction and/or Machine Learning.

The internship may last from 4 to 6 months and could serve as the foundation for a Phd thesis.

Required skills

- Basic knowledge about Machine Learning
- Programming skills (e.g. python)

Context

The Multi-Scale Interaction and MLIA teams at Sorbonne Université have a strong track record in HCI and Machine Learning conferences.

[1] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828.
[2] Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, *33*(4), 917-963.
[3] Cribier-Delande, P., Puget, R., Guigue, V., & Denoyer, L. (2020, April). Time Series Prediction using Disentangled Latent Factors. In *ESANN 2020-28th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*.
[4] Franceschi, J. Y., Dieuleveut, A., & Jaggi, M. (2019). Unsupervised scalable representation learning for multivariate time series. In *Advances in Neural Information Processing Systems* (pp. 4650-4661).