Neural Sign Language Translation for Virtual Reality

Xavier Giró-i-Nieto and Francesc Moreno-Noguer

RobIRI: Robot Perception and Manipulation at IRI
Institut de Robòtica i Informàtica Industrial UPC-CSIC, Carrer de Llorens i Artigas, 4, 08028 Barcelona

Project Description

Technical advances in videoconferencing have been recently boosted thanks to the major adoption of teleworking after the global COVID-19 health crisis. However, these systems are far away from providing a realistic experience, so augmented and virtual reality solutions are being improved to provide commercial solutions in the short term. The popularization of wearable headsets to interact in virtual environments (eg. Facebook Oculus in Metaverse) also provides an opportunity to remove the communication barriers between the Deaf and Hearing communities. Some Deaf have acquired skills to understand lip-reading, but this is an additional effort prone to communication errors. On the other hand, the large majority of the Hearing community does not possess sign language skills.

This research project aims at exploiting the recent advances in neural machine translation, together with the latest technologies, for virtual reality human-to-human communication. The main goal is exploring the capabilities of the Facebook Oculus 2 headset for sign language translation, that is, to recognize a continuous sequence of signs and translate them to spoken language. While Oculus 2 possesses a high precision hands tracker[1], it also covers the face of the signer, where a small portion of the sign language expression occurs. The project will quantify the loss of accuracy due to the loss of these facial features on a novel dataset of 4 hours to be recorded during this project. The recording will contain the same contents of the test partition of How2Sign[2], a dataset published by the project advisors in the International Conference in Computer Vision and Pattern Recognition (CVPR) 2021, ranked #4 in the most impactful publications across science according to the Google Scholar metrics 2021.

The student will work in the *RobIRI: Robot Perception and Manipulation at IRI* and will join a multidisciplinar team of experts in computer vision and natural language processing. The candidate should have a strong background in training deep neural networks, preferably, with the PyTorch framework. Upon reaching the project goals, a scientific publication is expected to be submitted to a major conference or journal.

References

1. Han, S. *et al.* Megatrack: monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graph. (TOG)* **39**, 87–1 (2020).

2. Duarte, A. *et al.* How2sign: a large-scale multimodal dataset for continuous american sign language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2735–2744 (2021).

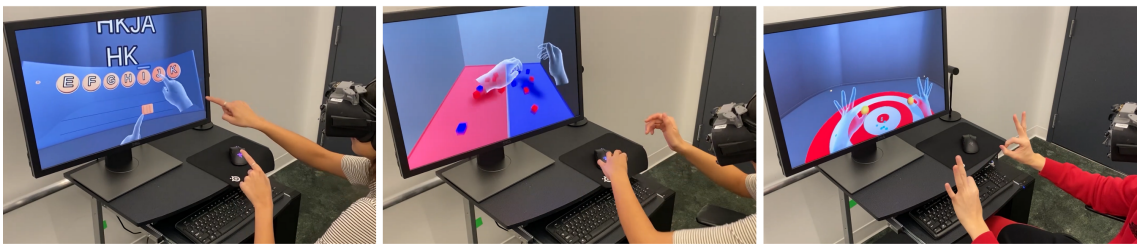

Figure 1. Hand tracking with Oculus[1].