A Neural Network Approach For Supporting Gesture Keyboards in Augmented Reality

Junxiao Shen[a,b,c]

[a] *Trinity College*
[b] *Department of Engineering, University of Cambridge, UK*
[c] *js2283@cam.ac.uk*

Keyword: Human-Computer Interaction, Deep Learning, Machine Learning, Text Input, Augmented Reality, Gesture-based Keyboard

Abstract: Augmented Reality (AR) can superimpose 3D virtual objects in a real environment to create a mixed reality experience that allows the user to simultaneously interact with digital and physical objects. The core value of AR is not a simple display of data, but the integration of immersive sensations and the natural interactions between the user and the virtual objects.

Text input is a key interaction method that allows users to enter text such as for sending messages or writing documents. Automatic Speech Recognition (ASR) is one of the methods. However, ASR is infeasible in noisy environments such as in factories or in crowded and public places. We propose the AR Gesture Keyboard (ARGK), which is an analogy to the word-gesture keyboard on smart phones [1].

The ARGK consists of a recogniser and a virtual keyboard. The recogniser maps the gesture traces to text, and we decided to use a machine learning model to build this recogniser. A simulated recogniser that simulates this machine learning-based recogniser was firstly designed to full-fill three goals. First, it can allow us to carry out user experiments to study user behaviour through the computation of micro metrics. Different metrics that measure the characteristics of the gesture data were proposed and analysed so that the user's behaviour can be explained in a more systematic way. The micro metrics describe the accuracy of the users while they gesture on the keyboard, how the users perceive where the keyboard lies and how consistent they are at gesturing the phrases on the plane relative to the virtual keyboard plane. Second, the text entry rate revealed from user experiments can show the potential of this new gesture input method to beat the state-of-the-art AR text entry techniques. Third, a large amount of data is required to build the machine learning-based recogniser. We can use the simulated recogniser to collect gesture trace data. Therefore, The outcomes of the simulated recogniser, not only the micro metrics but also the collected gesture trace data, of this project were anticipated to allow future work to generate artificial traces.

We carried out 20 user experiments to collect gesture data and performed a detailed analysis. The data collection experiment requires not only the HoloLens but also a tracking system that tracks the user's fingertip and the HoloLens so that the relative position and orientation between them can be computed. OptiTrack is used as the tracking

system since it offers a system with low latency and high accuracy. This simulated recogniser combined with the user experiments can not only enable us to study the user's natural behaviours but also give us insights on the characteristics of natural gesture traces so that we can implement these characteristics into the data synthesiser in the future. This is because a deep learning model may also require some handcrafted features implemented to improve the overall performance. Therefore, the project's goal is to capture and analyse the characteristics of the gesture trace, user behaviour and to assess the potential of the ARGK.

Since this project has completed the primary objectives, we began to investigate the generation of artificial traces. Therefore, the future work is to use deep learning to generate artificial traces and to map the natural and artificial traces to text. We proposed a neural network to generate artificial traces. The neural network consists of two LSTM layers and one attention mechanism. Different optimisers and training techniques were used to optimise the training process and performance. The performance of this synthesiser can be measured not only by classifiers (if the synthesiser has a representative output, then the classifier can not distinguish artificial and natural gesture traces) but also by the micro metrics proposed in the early stage.

In general, the whole project contains three main stages. First, develop the ARGK with a simulated recogniser; Second, run user experiments to collect gesture trace data and obtain the micro metrics quantitatively from the collected data; Third, develop a full ARGK with a machine learning-based recogniser by using a combination of collected data and synthetic data (which is synthesised from the collected data using another neural network). The fourth-year project contains the first and second stage. The second stage is critical for the third stage. The trace data is limited and the second stage can help us to generate synthetic data. Moreover, the second stage can quantitatively show the great potential of the ARGK system with the machine learning-based recogniser. Nevertheless, preliminary results from the simulated recogniser show the potential of the ARGK to support text entry in AR.

References

[1] P. O. Kristensson, Discrete and continuous shape writing for text entry and control, Ph.D. thesis, Linköping UniversityLinköping University, MDA - Human Computer Interfaces, The Institute of Technology (2007).