Demo: gesture based interaction with the Hololens 2

Marco Emporio
marco.emporio@univr.it
University of Verona
Italy

Ariel Caputo
ariel.caputo@univr.it
University of Verona
Italy

Deborah Pintani
deborah.pintani@univr.it
University of Verona
Italy

Fedrico Cunico
federico.cunico@univr.it
University of Verona
Italy

Federico Girella
federico.girella@univr.it
University of Verona
Italy

Andrea Avogaro
andrea.avogaro@univr.it
University of Verona
Italy

Marco Cristani
marco.cristani@univr.it
University of Verona
Italy

Andrea Giachetti
andrea.giachetti@univr.it
University of Verona
Italy

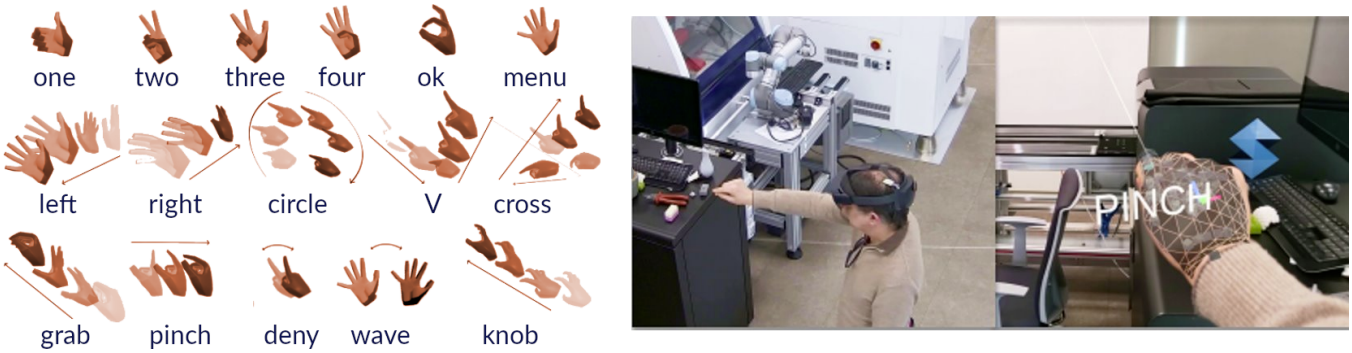

Figure 1: Left: dictionary of the gesture recognized by our Hololens 2 app. Right: the gesture recognition can be used to interact effectively with machines in an augmented workplace demo.

ABSTRACT

Gesture recognition is one of the default interaction modalities in many XR applications, although the gesture types recognized by many applications is typically limited to few static poses. In this demo we show that a recent network-based solution for online, sliding window, gesture classification from hand pose streams (On-Off deep Multi-View Multi-Task) can be used for the simultaneous detection and recognition of heterogeneous gestures, including dynamic coarse and fine grained ones, enabling interaction designers to create novel ways to perform interactive tasks that can be applied to different domains.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Human-centered computing** → **Gestural input**.

CHItaly 2023, September 20–22, 2023, Torino, Italy

ACM ISBN 979-8-4007-0806-0/23/09.
https://doi.org/10.1145/3605390.3610827

KEYWORDS

Mid-air Interaction, Gestures, Neural Networks

ACM Reference Format:
Marco Emporio, Ariel Caputo, Deborah Pintani, Fedrico Cunico, Federico Girella, Andrea Avogaro, Marco Cristani, and Andrea Giachetti. 2023. Demo: gesture based interaction with the Hololens 2. In *15th Biannual Conference of the Italian SIGCHI Chapter (CHItaly 2023), September 20–22, 2023, Torino, Italy.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3605390.3610827

1 INTRODUCTION

Gesture recognition is already popular in the domain of extended reality (XR). Headets like Hololens 2 or Meta Quest feature reliable finger tracking and API for the development of immersive application, can exploit simple gesture recognition methods. However, to fully exploit the potential advantages of gestural interface, the use of more advanced gesture recognition tools able to recognize not only static poses, but also dynamic gestures of different types is required. Several methods for the detection of heterogeneous hand gestures including hand motion and articulation have been proposed in the scientific literature, but not applied in the practice. In this demo we show an Hololens 2 app that can recognize in

real time a variety of heterogeneous gestures defined in a recent benchmark ([2]) and that can be used to design complex interactive task in immersive/augmented environment.

2 GESTURE DICTIONARY

The recognizer is trained to detect with a short delay the gestures included in a training dictionary proposed for the SHREC 2022 contest on online detection of heterogeneous gestures [2]. This dictionary includes some classical static poses, often used in interactive tasks, and dynamic gestures characterized by global hand trajectories and/or finger articulation. The task is not easy as, for a generic interaction, the gestures should be detected within sequences where they are interleaved with non meaningful gesticulation. The complete set of the gestures of the dictionary is shown in Figure 1, left.

In our demo, the user will be able to wear a Hololens 2 and observe that the app can automatically detect the execution of a gesture included in the dictionary with a rather short delay. Additional demos where the gestures can trigger specific actions will also be shown. Currently, we use the system in our Industrial Computer Engineering lab to control selected actions on a test production line, as shown in Figure 1, right.

3 GESTURE RECOGNIZER

The application recognizes gestures using a simple 1D convolutional network structure trained with the On-Off deep Multi-View Multi-Task paradigm, described in [1]. The algorithm exploits multiple time-local views (e.g., feature vectors) related to hand pose and movement to generate rich gesture descriptions, and is trained on multiple tasks (gesture start/end, gesture class) As not all the tasks-related labels are available for the training data, the method allows specific tasks to be switched on/off depending on whether they can apply to the input.

The method has proven to provide state-of-the-art performances on the dictionary used, in terms of gesture classification accuracy, segmentation accuracy, false positives, and decision latency while maintaining real-time operation.

4 THE DEMO SYSTEM

The demo presented is based on a client-server architecture sending the hand pose data stream to a server. As the network used is lightweight, there are no strong requirements for the computer used as server (e.g. an average laptop is sufficient to ensure real-time performances). We plan to test a simplified version that can work directly on the client side. Wearing a Hololens 2 headset, users will be able to check the accuracy of the online detection recognition of the gestures of the dictionary and to use gesture-based shortcuts to perform simple interaction tasks in a realistic context.

ACKNOWLEDGMENTS

This work was partially partially carried out within the PNRR research activities of the consortium iNEST (Interconnected North-Est Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.5 – D.D. 1058 23/06/2022, ECS00000043).

REFERENCES

[1] Federico Cunico, Marco Emporio, Federico Girella, Andrea Giachetti, Andrea Avogaro, and Marco Cristani. 2023. OO-dMVMT: A Deep Multi-view Multi-task Classification Framework for Real-time 3D Hand Gesture Classification and Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2744–2753.

[2] Marco Emporio, Ariel Caputo, Andrea Giachetti, Marco Cristani, Guido Borghi, Andrea D'Eusanio, Minh-Quan Le, Hai-Dang Nguyen, Minh-Triet Tran, Felix Ambellan, et al. 2022. SHREC 2022 track on online detection of heterogeneous gestures. *Computers & Graphics* 107 (2022), 241–251.