FaceDrive: Facial Expression Driven Operation to Control Virtual Supernumerary Robotic Arms

Masaaki Fukuoka
Keio University

Adrien Verhulst
The University of Tokyo

Fumihiko Nakamura
Keio University

Ryo Takizawa
Keio University

Katsutoshi Masai
Keio University

Maki Sugimoto
Keio University

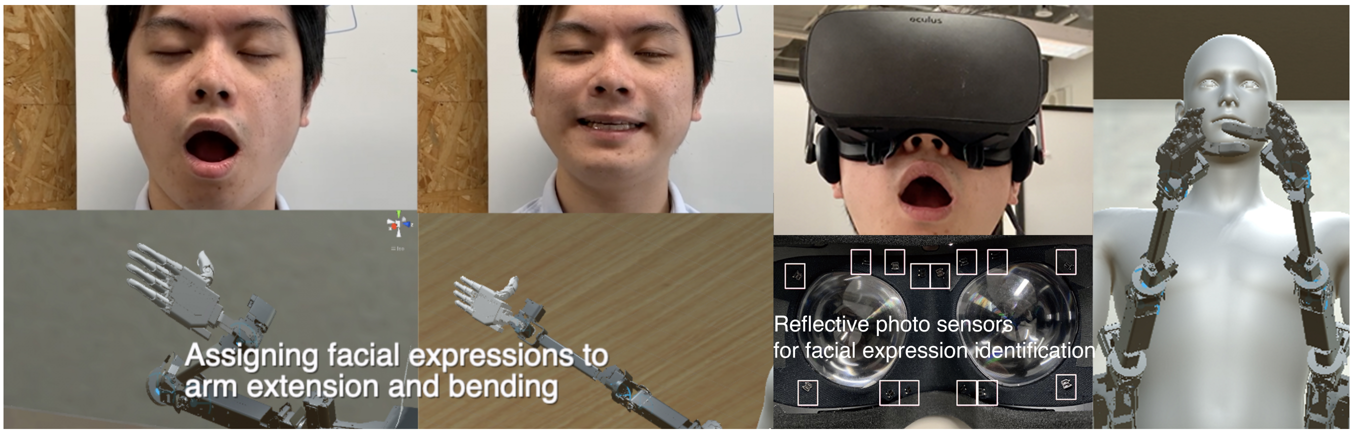

Figure 1: Snapshots of FaceDrive and virtual robot arm using facial expression of the HMD user.

ABSTRACT

Supernumerary Robotic Arms (SRAs) can make physical activities easier, but require cooperation with the operator. To improve co-operation, we predict the operator's intentions by using his/her Facial Expressions (FEs). We propose to map FEs to SRAs commands (e.g. grab, release). To measure FEs, we used a optical sensor-based approach (here inside a HMD), the sensors data are fed to a SVM classifying them in FEs. The SRAs can then carry out commands by predicting the operator's FEs (and arguably, the operator's intention). We made a Virtual reality Environment (VE) with SRAs and synchronizable avatar to investigate the most suitable mapping between FEs and SRAs. In SIGGRAPH Asia 2019, the user can manipulate virtual SRAs using his/her FEs.

CCS CONCEPTS

• **Computer systems organization** → *External interfaces for robotics*; *Real-time operating systems*; • **Software and its engineering** → Virtual worlds training simulations.

SA '19 XR, November 17–20, 2019, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-6947-3/19/11.
https://doi.org/10.1145/3355355.3361888

KEYWORDS

datasets, support vector machine, augmented human

ACM Reference Format:
Masaaki Fukuoka, Adrien Verhulst, Fumihiko Nakamura, Ryo Takizawa, Katsutoshi Masai, and Maki Sugimoto. 2019. FaceDrive: Facial Expression Driven Operation to Control Virtual Supernumerary Robotic Arms. In *SIGGRAPH Asia 2019 XR (SA '19 XR), November 17–20, 2019, Brisbane, QLD, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3355355.3361888

1 INTRODUCTION

Supernumerary Robotic Arms (SRAs) can support people in physical activities and are especially useful when the user wants to: (i) reduce physical workload; and (ii) use more than 2 arms (c.f. Fig. 2). Although SRAs can also be operated remotely, SRAs are usually operated by the user "wearing" them (the operator), with: (i) a direct mapping of the operator's own arms to the SRAs (e.g. robot arms strapped to the operator); and / or (ii) joysticks / buttons. However, cooperation with SRAs is challenging, as it requires the operator to balance his/her arms movements and the SRAs movements. Indeed the SRAs are "moving with" the operator's arms, making it difficult to keep the SRAs in place while the operator moves his/her own arms. Moreover, the SRAs cannot guess the intends behind the operator movements. In this work, we investigated the use of Facial Expressions (FEs) based commands to control virtual SRAs (here *only* 2 arms). Indeed we argue that FEs are: (i) an usable control method to give continuous or discrete commands (e.g. to move the robotic arms along an axis, to record / play a command, etc.);

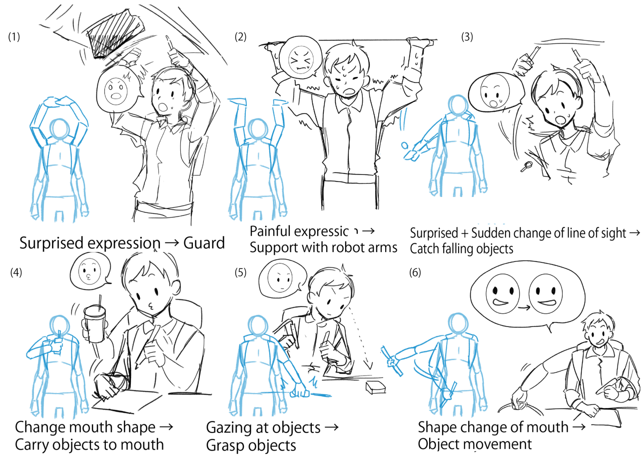

Figure 2: The operator has 2 hands occupied and uses a FE command to move the SRAs.

and (ii) able to convey "emotion-based" commands (e.g. to have a kill-switch if the operator looks frighten, to move slowly if the operator looks tired, etc.).

We used a VE with virtual SRAs instead of using real SRAs. Indeed VR is a cost-effective, safer and faster approach than developing with real SRAs. Since the HMD is covering the face (and a RGB camera cannot be used to get FEs), here facial movements are measured with 16 reflective optical sensors built-in the HMD, and classified in FEs thanks to a SVM (c.f. Sec. 2). The novelty of this work is attempting to connect robots actions to human motions and emotions.

2 SYSTEM OVERVIEW

We used the "AffectiveWear" approach [Suzuki et al. 2017] to capture FEs when wearing a HMD. Here, a set of 16 reflective photo-sensors is fixed inside an Oculus CV1 (c.f. Fig. 1). The distance value of every sensors is calculated and fed to a set of SVM classifiers (each of them trained to correspond to a FE, such as raising eyes, moving left cheek, etc.). The highest confident classifier predicts the corresponding FE.

The task we developed uses 4 SRAs actions: (i) extending / contracting a SRA to reach an object (done by changing the elbow angle); and (ii) opening / closing the SRA hand to grab an object. To associate these actions to FEs, we used the "Tumori control" method [M.Niwa et al. 2012]: We playback SRA actions in the VE (here "reaching" actions and "grab" actions); When the SRA does an action, the operator does a FE to associate the action to a FE (e.g. when the SRA extends to reach an object, the operator smiles, thus associating the "reaching action" to his/her smile).

The "Tumori control" method works especially well between 1-1 associations (1 operation linked to 1 FE). While this is an issue for advanced controls, here we are targeting simple controls (moreover FEs on their own must be kept simple to be done effectively). In our system, if there is a change of FE when a motion is executed, then the motion is automatically associated to the FE at the time of the expression change Fig. 3.

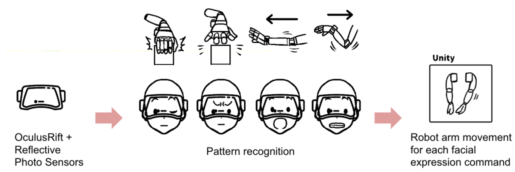

Figure 3: System execution. A FE commands a SRA.

3 DEMONSTRATION

Our method can estimate the facial expression of the HMD user as probabilistic values. There are 2 SRAs localized on each side of the avatar's waist (c.f. Fig. 1). Since the operator can only control 1 SRA at a time, the right / left SRA is controllable when the operator is looking at respectively his/her right / left side (when grabbing an object, the right / left switching is automatically turned off to easily move the object). The "controlled" SRA follows the head direction thanks to Inverse Kinematics (IK). The VE receives the operator's current FE every 15ms from our FEs recognition system (sent by UDP). In addition, the FE commands (e.g. opening / closing) and the elapsed time since the beginning of the task are displayed against a wall in front of the operator. The user experience is as follows: Every user receives an explanation of the system and the task, and then put on the HMD. Once in the VE, after a quick calibration, participants look at a SRA actions (i.e. extending / contracting arm and opening / closing hand) and ask for each actions to associate the action to a FE of their choice. We collect 2,000 sensor data samples of every FEs (to find out the FEs used and also to prepare a "global" classifier). After associating every SRA actions to a FE, the participants perform a manipulation task (i.e. reach for a sphere positioned by extending a SRA arm, then grab the sphere by closing the SRA hand, move the sphere inside a cube and release it by opening the SRA hand.

4 CONCLUSION

In this demonstration, we present our mapping method between FEs and SRAs commands by using a facial expression recognition technology by embedded photo-reflective sensors and machine learning. By using this technology, we are able to reflect facial expression of HMD users in various robot arm manipulation in Virtual Reality systems.

ACKNOWLEDGMENTS

This work was supported by JST ERATO(JPMJER1701).

REFERENCES

M.Niwa, H.Iizuka, H.Ando, and T.Maeda. 2012. Tumori Control : Manipulate Robot with detection and transmission of an Archetype of Behavioral Intention. *Transactions of the Virtual Reality Society of Japan* 17, 1 (2012), 3–10. https://doi.org/10.18974/tvrsj.17.1_3

K. Suzuki, F. Nakamura, J. Otsuka, K. Masai, Y. Itoh, Y. Sugiura, and M. Sugimoto. 2017. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *2017 IEEE Virtual Reality (VR)*. 177–185. https://doi.org/10.1109/VR.2017.7892245