ZoomTouch: Multi-User Remote Robot Control in Zoom by DNN-based Gesture Recognition

Ilya Zakharkin
Skolkovo Institute of Science and Technology (Skoltech)
Moscow, Russia
ilya.zakharkin@skoltech.ru

Arman Tsaturyan
Skolkovo Institute of Science and Technology (Skoltech)
Moscow, Russia
arman.tsaturyan@skoltech.ru

Miguel Altamirano Cabrera
Skolkovo Institute of Science and Technology (Skoltech)
Moscow, Russia
miguel.altamirano@skoltech.ru

Jonathan Tirado
Skolkovo Institute of Science and Technology (Skoltech)
Moscow, Russia
jonathan.tirado@skoltech.ru

Dzmitry Tsetserukou
Skolkovo Institute of Science and Technology (Skoltech)
Moscow, Russia
d.tsetserukou@skoltech.ru

ABSTRACT
We present ZoomTouch, a breakthrough technology for multi-user and real-time control of robot from Zoom by DNN-based gesture recognition. The users can have a video conferencing in a digital world and at the same time to perform dexterous manipulations with tangible objects by remote robot. As the scenario, we proposed the remote robotic COVID-19 test Laboratory to substitute medical assistant working in protective gear in close proximity with infected cells and to considerably reduce the time to receive the test results. The proposed technology suggests a new type of reality, where multi-users can jointly interact with remote object, e.g. make a new building design, joint cooking in robotic kitchen, etc, and discuss/modify the results at the same time.

CCS CONCEPTS
• Computing methodologies → Activity recognition and understanding; • Human-centered computing → Gestural input; Collaborative interaction; Graphics input devices; • Computer systems organization → Robotic control.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SA ’20 Emerging Technologies, December 04-13, 2020, Virtual Event, Republic of Korea © 2020 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8110-9/20/12.
https://doi.org/10.1145/3415255.3422892

KEYWORDS
Gesture Recognition, Hand Tracking, Robotics, Teleoperation

ACM Reference Format:
Ilya Zakharkin, Arman Tsaturyan, Miguel Altamirano Cabrera, Jonathan Tirado, and Dzmitry Tsetserukou. 2020. ZoomTouch: Multi-User Remote Robot Control in Zoom by DNN-based Gesture Recognition. In SIGGRAPH Asia 2020 Emerging Technologies (SA ’20 Emerging Technologies), December 04-13, 2020. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3415255.3422892

1 INTRODUCTION
Fully autonomous systems are already essential in industry, science, and medicine. However, an actively developing alternative is a robot-human interaction, where instead of replacing a person with a robot, a person controls the robot [Liu and Wang 2018]. Many methods have been proposed on how to arrange this interaction. Some works use gloves with sensors on hand to control the robotic arm or robot with a gripper [Fang et al. 2015]. To achieve smooth and safe guiding of a drone formation by a human operator, an impedance control and tactile feedback was proposed in [Tsykunov et al. 2019].

During the COVID-19 pandemic, Zoom became an essential tool for education, business meeting, family communication, etc. However, it is not possible to jointly develop a hardware technology in

Figure 1: a) Real-time gesture recognition in Zoom by DNN. b) Robot with a gripper that dexterously manipulates the pipettes. c) Available hand gestures for flexible remote control.
factory environment, or to make a medical tests and scientific experiments remotely. The paper focuses on ZoomTouch to allow several users of Zoom to control the robot located remotely. ZoomTouch consists of Zoom software platform, DNN gesture recognition engine, and remote robot with embedded high fidelity tactile sensor. We conducted an experiment on the gesture recognition and real-time robot control to perform the simulation of COVID-19 test. The demonstration revealed a very robust performance of gesture recognition algorithm and agility of robot manipulating soft object, i.e., pipette.

2 ZOOMTOUCH PIPELINE

The ZoomTouch system software consists of three parts: screen capture module, hand tracking module, and robot control module. Hardware part is made up of four components: user’s laptop with software developed by us Fig. 1(a), a 6 Degrees of Freedom (DoF) collaborative robot UR10 from Universal Robots (b), and a 2-finger gripper from Robotiq endowed with a tactile sensor array which comprises 10 by 10 pressure sensing points [Yem et al. 2019] (c).

When user launches the video conference on the laptop, the same conference is opened on the server, that runs the algorithm to perform screen capture, allowing hand tracking and gesture recognition right from the video conference stream. Eventually, the information is sent to the robot control module to move the robot to the target position. Thus, ZoomTouch allows a person to control a robot remotely in real-time using only laptop webcam and video conference software, such as Zoom, Google Meet, Microsoft Teams.

The hand tracking module is implemented on the base of Mediapipe framework. It tracks the landmarks of a hand (or both hands), using them we recognize different hand gestures in real-time on CPU. In this work we classify 5 gestures: move, angle, grab, finger distance, and no gesture (Fig. 1(c)). The distance of finger distance gesture is controlled by grab gesture on the other hand. We trained Gradient Boosting classifier on manually labeled gesture dataset of 1000 images: 200 per move, angle and grab classes, and 400 for no gesture class. We used normalized landmarks, angles between joints, and pairwise landmark distances as features to predict the gesture class. It resulted in multi-class accuracy of 91% on a test set. The landmarks and gesture on each video frame are passed to the robot control module via the high-performance asynchronous messaging library ZeroMQ.

The robot control module uses current hand location and gesture information to change the position of the 6 DoF UR10 robot Tool Center Point (TCP) position, angle, and the state of the 2-finger gripper (e.g., close, open, or an open distance).

The robotic gripper is embedded with tactile sensor arrays. Each array is capable of sensing a maximum frame area of 5.8 cm2 with a resolution of 100 points per frame. The sensing frequency equals 120 Hz (frames per second). Robot can precisely detect the pressure applied to instruments or objects grasped by the fingers. The pressure data can be used for dexterous and safe manipulations with fragile objects and tactile feedback in teleoperation systems.

3 APPLICATIONS

During the demo, a person will seat in front of a laptop in a videoconference. The user will see the videos from several cameras including that installed on the end-effector of the robotic arm and will be able to control manipulator by hand gestures. The participants of the demonstration will perform the remote medical test. The user will guide the robot with hand gestures to take the pipette from the table, put it in a test tube with simulated blood, take a blood sample, and then bring it to the antibody test, and finally drop blood into the blood compartment.

It is possible to achieve effective remote collaboration thanks to the Zoom’s ‘spotlight’ function, which enables the host to pin the user’s video for all participants of the call. The host-user manually ‘spotlights’ the video from current robot operator, so that the screen capture module processes the picture of the correct person.

Our system can be used in many future teleoperation scenarios, e.g., for making experiments on International Space Station or providing multi-user access to robots of Mars 2020 space mission to get the insight of Martian soil composition. ZoomTouch can potentially be used to control other types of robots, e.g. mobile robots, swarm of flying robots equipped with cameras, etc.

ACKNOWLEDGMENTS

The authors would like to thank Professor Hiroyuki Kajimoto for providing tactile sensor arrays.

REFERENCES

Bin Fang, Di Guo, Fuchun Sun, Huaping Liu, and Yupei Wu. 2015. A robotic hand-arm teleoperation system using human arm/hand with a novel data glove. In 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2483–2488.

Hongyi Liu and Lihui Wang. 2018. Gesture recognition for human-robot collaboration: A review. International Journal of Industrial Ergonomics 68 (2018), 355–367.

E. Tsykunov, R. Agishev, R. Ibrahimov, L. Labazanova, A. Tleugazy, and D. Tsetserukou. 2019. SwarmTouch: Guiding a Swarm of Micro-Quadrotors With Impedance Control Using a Wearable Tactile Interface. IEEE Transactions on Haptics 12, 3 (2019), 363–374.

Vibol Yem, Hiroyuki Kajimoto, Katsunari Sato, and Hidekazu Yoshihara. 2019. A System of Tactile Transmission on the Fingertips with Electrical-Thermal and Vibration Stimulation. In Human Interface and the Management of Information. Information in Intelligent Systems, Sakae Yamamoto and Hirohiko Mori (Eds.). Springer International Publishing, Cham, 101–113.