Remote Operated Human Robot Interactive System using Hand Gestures for Persons with Disabilities

Enamul Karim
The University of Texas at Arlington
Arlington, Texas, USA
enamul.karim@mavs.uta.edu

Harish Ram Nambiappan
The University of Texas at Arlington
Arlington, Texas, USA
harishram.nambiappan@mavs.uta.edu

Sneh Acharya
The University of Texas at Arlington
Arlington, Texas, USA
sxa6003@mavs.uta.edu

Fillia Makedon
The University of Texas at Arlington
Arlington, Texas, USA
makedon@uta.edu

ABSTRACT

This paper proposes a novel Human-Robot Interactive System where users can interact with the robotic system using hand gestures, even from a distance, through a smartphone-based IoT-Controller Framework. The system is primarily designed to help people with vocal and hearing impairments. A mobile application records user gestures shown in front of a smartphone and sends the data to a server. The server performs gesture recognition and forms a command which is sent to the robotic system. The robotic system performs the required task based on the given command. The robotic system is set to carry out an object pick and place operation based on the gesture commands in the system's initial experiments. The accuracy of both the gesture recognition model and the object grasping task were measured, along with the time taken for the system to complete the entire task. Experimental results showed that the system performed well, achieving 100% accuracy in object grabbing and gesture recognition, and the average time taken by the system to complete the robotic pick and place task was around 78 seconds.

CCS CONCEPTS

• **Human-centered computing** → **Smartphones**; • **Computer systems organization** → **Robotics**.

KEYWORDS

Internet-of-Things, Gesture Recognition, Human Robot Interaction

ACM Reference Format:
Enamul Karim, Harish Ram Nambiappan, Sneh Acharya, and Fillia Makedon. 2023. Remote Operated Human Robot Interactive System using Hand Gestures for Persons with Disabilities. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '23), July 05–07, 2023, Corfu, Greece.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3594806.3594832

PETRA '23, July 05–07, 2023, Corfu, Greece

ACM ISBN 979-8-4007-0069-9/23/07.
https://doi.org/10.1145/3594806.3594832

1 INTRODUCTION

People with vocal and hearing impairments usually have difficulties in communication. There have been research works in terms of aiding people with vocal and hearing impairments through developing systems that can be used for communication. Hand gesture recognition is one of the popular problems in computer vision that has its application in a variety of areas like healthcare [5, 6, 16], gaming [4] etc. Works such as [7, 14] have focused on developing wearable gloves using flex sensors to detect and recognize hand signs and gestures and produce output either through speech or display, which helps people with vocal and hearing impairments in communication. Additionally, there have been research works [2, 9] that focus on recognizing hand gestures and sign languages using computer vision and deep learning methods. The majority of research into developing systems to assist people with vocal and hearing impairments primarily focus on human-to-human communication scenarios, despite the need to develop systems that can assist these people in a human-robot interaction scenario. Although there are some works [1, 8] which utilize sign languages with robotic systems, they are limited to communication scenarios where the robotic system are used only for sign language interpretation. Robotic systems that can perform various tasks by understanding gestures or sign language need to be developed. Works like [3, 15] have used gesture recognition to control and teleoperate a robot, which requires the user to be present with the robot at all times. However, there are situations in which the user might be working on a task from a distance and needs to send commands to the robot to carry out the necessary robotic task, for which a proper system needs to be developed.

In this paper, we propose a novel and initial version of a Human-Robot Interactive system that allows users to instruct the robotic system to perform certain tasks using hand gestures through a smartphone application and a Smartphone-based IoT-Controller Framework [12], even when the user is at a distance. The IoT-Controller Framework recognizes and interprets the gestures sent by the smartphone application and delivers the appropriate commands to the robotic system to carry out the required task. This way, the IoT-Controller Framework connects the user's smartphone with the robotic system. Figure 1 demonstrates how the smartphone camera records hand gestures and transfers them to the server for gesture detection and command generation, which is then used by the robotic system to execute robotic movement. The MINA Robot

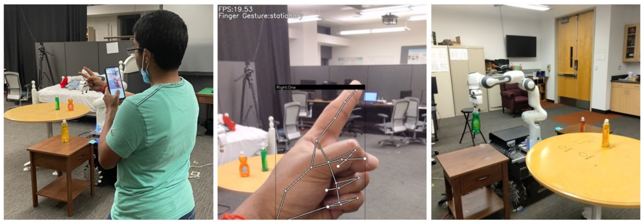

Figure 1: (Left) Smartphone Application recording the gestures shown by the user, (Middle) Gesture recognition performed at the server, (Right) Robotic Movement of pick and place task

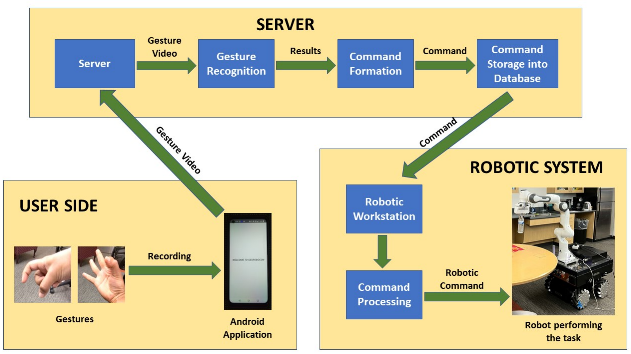

Figure 2: System Architecture

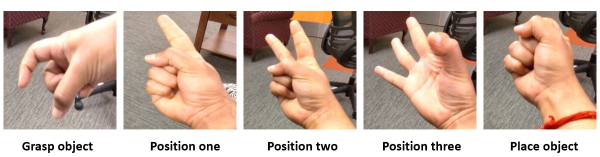

Figure 3: Gestures used for system interaction

[11, 13], which consists of a Summit-XL Steel Mobile Base with a Franka Emika Panda Arm, is used to implement the robotic tasks.

2 METHODOLOGY

The system architecture of the Human-Robot Interactive System is shown in Figure 2. In this paper, the robotic task focuses on picking up objects from a source location and placing it in a destination location. The Smartphone application uses the smartphone's camera sensor to record the gestures provided by the user. Figure 3 shows the gestures used for this application. The application sends the gesture video to the server workstation, where gesture recognition is performed. MediaPipe [10] is used for hand gesture recognition which employs machine-learning techniques to recognize the hand's landmarks from a single frame. The keypoints detected by the framework are initially recorded in a CSV file and labeled into one of the following five classes: *Pick, Place, One, Two, Three.* TF Lite is used to train the model for recognizing the various hand gestures. The model was fit using the Adam optimizer and was trained for

Figure 4: Positions at which gestures were recorded and sent to the server for testing the system

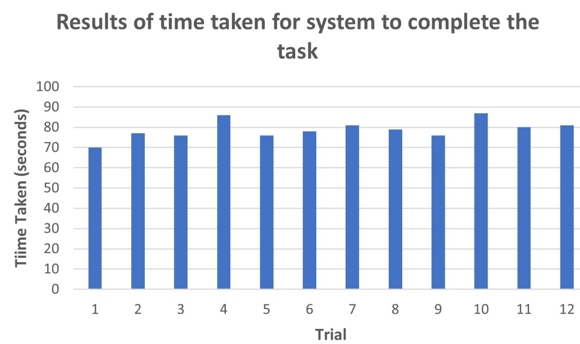

Figure 5: Results of time taken for system to complete task

1000 epochs. It attained a validation accuracy of about 98%. Based on the gesture recognition results, the resultant command is formed which contains the information regarding the object to be picked up by the robotic system and the destination location where it should be placed. The resultant command is stored in the server database which is being retrieved and sent to the robotic system. The robotic system processes the command to get the object and destination position and map those details with the joint and cartesian values that the robot will use to carry out the pick and place task.

3 EXPERIMENTATION

The initial version of the system was tested in an object pick and place scenario. Three objects, labeled as numbers 1, 2, and 3, were placed in front of the robot and two tables were placed on both sides of the robot which served as the destination locations. The user sends gesture commands to the system using a smartphone application in real-time and, based on the commands, the robotic system has to pick objects and place them on one of the destination tables. For safety purposes, the robot was set to move at 10 percent of its maximum velocity. The smartphone, server, and robotic system are all connected through a 2.4 GHz wireless network. The gestures for the commands were given in the following format: "Pick x Place y", where x = 1,2,3 refers to the object number to be picked up by the robot and y = 1,2 refers to the destination table number. Gesture commands were sent from four distinct locations around the experimental area to check whether the application delivers the commands to the robotic system through the network and see if our model recognizes the hand gestures properly regardless of the background scenarios. Figure 4 shows the experimental setup with the four different positions from which gestures were recorded and sent to the system. A total of 12 trials were conducted. Gesture recognition accuracy, object grasping accuracy and the task completion time were calculated. Gesture recognition accuracy refers to the accuracy of the gesture recognition model, object grasping accuracy refers to the number of times the robotic system grasped the correct object and placed it on the correct destination table based on the gesture command, and the completion time was calculated from the moment the gesture commands were sent from the smartphone application till the moment the robotic system completed the pick and place task.

Results showed that the initial version of the system performed well with 100 percent gesture recognition and object grasping accuracy. This indicates that, in every trial, the current gesture recognition model was able to recognize the given set of gestures properly to form the required command which is utilized by the robotic system to grasp the correct object and place it in the appropriate destination position. Figure 5 shows the time taken for the system to complete the task in every trial where the time taken ranges between 70 seconds to 86 seconds and the average time taken for the system to complete the task was around 78 seconds.

4 CONCLUSION AND FUTURE WORK

A novel hand gesture based Human-Robot Interactive System is proposed for helping people with vocal and hearing impairments where the users use hand gestures to instruct the robotic system to perform a robotic task through a smartphone-based IoT-Controller Framework with initial results showing that the system took an average time of 78 seconds to complete the required robotic task. Future work includes developing an eyeglass-based prototype with embedded cameras to record hand gestures which would be user-friendly and convenient for people with vocal, hearing, and vision impairments, along with incorporating gesture and speech recognition to create a multimodal human-robot interaction system, as well as integrating a mobile robot base and robotic arm functionality to carry out various robotic tasks. Also in terms of hand gesture recognition, the model will be improvised to make it better such that it can verify and interpret gestures made by various kinds of people, including those with Parkinson's disease.

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation under award number 1719031.

REFERENCES

[1] Dina A Alabbad, Nouha O Alsaleh, Naimah A Alaqeel, Yara A Alshehri, Nashwa A Alzahrani, and Maha K Alhobaishi. 2022. A Robot-based Arabic Sign Language Translating System. In *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*. IEEE, 151–156.

[2] Kshitij Bantupalli and Ying Xie. 2018. American sign language recognition using deep learning and computer vision. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 4896–4899.

[3] Alex Chico, Patricio J Cruz, Juan Pablo Vásconez, Marco E Benalcázar, Robin Álvarez, Lorena Barona, and Ángel Leonardo Valdivieso. 2021. Hand gesture recognition and tracking control for a virtual UR5 robot manipulator. In *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM)*. IEEE, 1–6.

[4] Farnaz Farahanipad, Harish Ram Nambiappan, Ashish Jaiswal, Maria Kyrarini, and Fillia Makedon. 2020. HAND-REHA: dynamic hand gesture recognition for game-based wrist rehabilitation. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 1–9.

[5] Preetham Ganesh, Reza Etemadi Idgahi, Chinmaya Basavanahally Venkatesh, Ashwin Ramesh Babu, and Maria Kyrarini. 2020. Personalized system for human gym activity recognition using an RGB camera. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 1–7.

[6] Srujana Gattupalli, Ashwin Ramesh Babu, James Robert Brady, Fillia Makedon, and Vassilis Athitsos. 2018. Towards deep learning based hand keypoints detection for rapid sequential movements from rgb images. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. 31–37.

[7] Kuruvilla John, Aparna Sreekumar, and Lija Rachel Mathew. 2021. Wearable Gesture Detection Glove for Mute People. In *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*. IEEE, 97–101.

[8] Nazerke Kalidolda and Anara Sandygulova. 2018. Towards Interpreting Robotic System for Fingerspelling Recognition in Real Time. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 141–142.

[9] MM Kamruzzaman. 2020. Arabic sign language recognition and generating Arabic speech using convolutional neural network. *Wireless Communications and Mobile Computing* 2020 (2020).

[10] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[11] Harish Ram Nambiappan, Stephanie Arevalo Arboleda, Cody Lee Lundberg, Maria Kyrarini, Fillia Makedon, and Nicholas Gans. 2022. MINA: A Robotic Assistant for Hospital Fetching Tasks. *Technologies* 10, 2 (2022). https://doi.org/10.3390/technologies10020041

[12] Harish Ram Nambiappan, Enamul Karim, Md Jillur Rahman Saurav, Anushka Srivastav, Nicholas Gans, and Fillia Makedon. 2022. Smartphone Based IoT-Controller Framework for Assisting the Blind in Human Robot Interaction. In *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '22)*. Association for Computing Machinery, New York, NY, USA, 514–516. https://doi.org/10.1145/3529190.3534782

[13] Harish Ram Nambiappan, Krishna Chaitanya Kodur, Maria Kyrarini, Fillia Makedon, and Nicholas Gans. 2021. MINA: A Multitasking Intelligent Nurse Aid Robot. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*. 266–267.

[14] Atik Shahariar Patwary, Zunaid Zaohar, Asmany Akter Sornaly, and Riasat Khan. 2022. Speaking System for Deaf and Mute People with Flex Sensors. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 168–173.

[15] Wen Qi, Salih Ertug Ovur, Zhijun Li, Aldo Marzullo, and Rong Song. 2021. Multi-sensor guided hand gesture recognition for a teleoperated robot using a recurrent neural network. *IEEE Robotics and Automation Letters* 6, 3 (2021), 6039–6045.

[16] Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Ashish Jaiswal, Alexis Lueckenhoff, Maria Kyrarini, and Fillia Makedon. 2020. A multi-modal system to assess cognition in children from their physical movements. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 6–14.