International Journal of Computer Vision and Signal Processing, 15(1), 130-138(2025) ORIGINAL ARTICLE

A Real-Time Vision-Based System for Human Gesture Recognition in Collaborative Work Cells

Natchanon Suppaadirek

Department of Life Science and Systems Engineering
Kyushu Institute of Technology, Japan

Shreyan Shukla

Department of Computer Science and Engineering
Dronacharya Group of Institutions, India

Piyush Mudgal

Department of Computer Science and Engineering (AI and ML)
Dronacharya College of Engineering, India

Tomohiro Shibata*

Department of Life Science and Systems Engineering
Kyushu Institute of Technology, Japan

Abstract

Human-robot interaction is a major field of investigation, focusing on the optimization of working processes as well as employee productivity. Despite an enormous amount of progress made in this direction, the necessity to develop systems targeting people with disabilities remains a pressing need. This report presents a novel paradigm for assistive robotics via the development of an intelligent work cell for aging individuals and physically disabled persons. The system merges depth camera technology, light machine learning, and MediaPipe-based human tracking to enable real-time human-robot interaction through accurate inference of user intent. Key innovations include a gimbal-mounted depth camera for motion tracking of the user, a modular 3D-printed gripper for easily customizing manipulation, and an efficient gesture classification pipeline. Experimental results demonstrate that the system achieves over 90% average gesture recognition accuracy, which is comparable to or higher than similar gesture-based systems, with real-time performance. The system bridges the gap between theoretical research and practical application in assistive robotics.

Contribution of the Paper: The contribution of this paper lies in the development and evaluation of an innovative system to interpret human gestures through vision-based technology, with a specific focus on improving human-robot collaboration.

Keywords: human-robot, computer vision, work cell

IJCVSP

ISSN: 2186-1390 (Online)
http://cennser.org/IJCVSP

Article History:
Received: 12/4/2025
Revised: 14/7/2025
Accepted: 1/11/2025
Published Online: 23/11/2025

1. INTRODUCTION

Growth in automation and robotics in workplaces has opened new possibilities for the inclusive employment of older people and persons with physical disabilities. Studies indicate that robotic workplaces have a high potential to allow the inclusion of workers with mobility disabilities by reducing physical stress and allowing the adaptation of tasks through the application of assistive technologies [1].

*Corresponding author
Email addresses: suppaadirek.natchanon733@mail.kyutech.jp (Natchanon Suppaadirek), shreyan.16166@gnindia.dronacharya.info (Shreyan Shukla), piyush.24248@ggnindia.dronacharya.info (Piyush Mudgal), tom@brain.kyutech.ac.jp (Tomohiro Shibata)

As the global workforce ages and the demand for diverse barrier-free workplaces increases, the demand for flexible human-oriented automation solutions is increasing [2].

One solution is the development of adaptive work cells, flexible workspaces where humans and robots collaborate seamlessly. These types of workstations are programmed to maximize productivity while serving the needs of the individual user [3]. Recent advances in human-robot collaboration have improved efficiency and safety of interaction in the workplace [4], [5].

However, while such technological advancements have been occurring, attempts at explicitly addressing assistance for individuals with disabilities in such collaborative working environments have been sparse [6],[7]. To address this problem, ergonomics-inspired solutions and advanced systems with real-time human gesture recognition are necessary [8].

This paper presents a vision-based human-robot collaboration system for inclusive manufacturing work cells. By leveraging RGB-D information, MediaPipe-based gesture tracking, and lightweight machine-learning models, our system monitors hand movements and orientations in real-time to infer user intent and guide robotic assistance. Our approach has three novel contributions over conventional systems:

An Intel RealSense D435 system mounted on a gimbal that dynamically tracks the user's position to provide continuous vision coverage during movement. A modular, 3D-printed gripper system that allows for on-demand end-effector and tool customization to support different assistive tasks and a real-time gesture recognition pipeline that fuses gesture filtering, 3D hand landmark extraction, and ensemble learning (e.g., XGBoost, Voting Classifier) optimized for fast training and deployment.

The system is designed to function with varying lighting and environmental conditions, making it more resilient than traditional marker-based motion capture systems[9], [10]. It decodes multiple vision streams simultaneously to construct a full spatial model of the workspace so that both the robotic arm and the user can naturally interact with shared tools and objects [11], [12].

To enable rapid prototyping of tools and task-specific customization, the system incorporates additive manufacturing directly into the work cell. This allows fast creation of specialized grippers and interfaces as the user's needs evolve. The robotic arm, which is equipped with interchangeable tools, is operated using gesture-based input, thereby enabling object handover, manipulation, or task assistance without requiring complex user input.

Compared to previous gesture-based human-robot collaboration systems, which typically base their camera configuration on fixed camera setups and restricted user locomotion, our solution brings three significant innovations [13]. Firstly, the use of a gimbal-mounted RealSense camera actively follows the motion of the user, providing uninterrupted vision coverage. Secondly, the modular 3D-printed gripper allows for quick adaptation to various assistive tasks. Third, the gesture recognition pipeline combines MediaPipe feature extraction with efficient ensemble learning models that can be run in real-time. Such features combined provide greater flexibility.

The goal of this project is to allow older workers and individuals with motor disabilities to remain productive contributors to industrial labor in a meaningful capacity. Rather than replacing human workers, though, our system augments them, enabling user control, reducing physical effort, and opening accessible paths to ongoing workforce participation.

2. RELATED WORKS

Vision-based gesture recognition has become a cornerstone in human-robot interaction (HRI), offering natural human movement to enable intuitive control. Earlier methods relied on hand-crafted features, but newer approaches use more and more machine learning and deep learning to give robust performance under different conditions [14]. Deep neural networks, such as 3D Convolutional
Neural Networks (CNNs) Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have worked well in constructing skeletal poses and gait images for action forecasting [15]. RGB-D sensors, namely cheap cameras such as the Microsoft Kinect and Intel RealSense, have also enabled real-time gesture detection using combined depth and color streams [16], [17]. Vision-based robotic systems for human-robot interaction have also evolved significantly, enabling more intuitive gesture-based control through real-time visual processing [18].

Gesture datasets from RGB-D inputs have made it possible to train and benchmark robust models in noisy or cluttered settings. The datasets have played a central role in the creation of real-time dynamic gesture recognition using skeletal tracking or temporal models such as Finite State Machines (FSMs) [19]. Applications span industrial automation to automotive systems, where depth-based gesture recognition enhances driver-vehicle interaction [20], [21]. Machine learning played a key role in the translation of raw gesture data into semantic robot commands. By acquiring predictive models from spatial and temporal characteristics of gestures, systems can accurately infer user intent with flexibility [22]. These models have several advantages, including data training without being specific to particular users, real-time inference, and generalizing well across users and environments. Lightweight classifiers, such as decision trees and ensemble algorithms (e.g., Voting Classifier), are especially promising for application in low-latency systems, where computational efficiency is of the utmost importance. Recent work also explores multimodal approaches that integrate vision and tactile sensing for robust real-time gesture and attention recognition in human-robot interaction [23]. Though many approaches focus on deep learning, recent work has shown that the combination of gesture filtering with efficient feature extraction (e.g., via MediaPipe or OpenPose) and classical

machine learning can result in highly accurate yet computationally light gesture recognition systems. These methods, however, assume fixed cameras and constrained user movement.

Robot assistive systems designed for users with motor disabilities or age-related limitations are increasingly using vision and gesture-based interfaces for simplicity of operation. Prior research has demonstrated the feasibility of using depth cameras and wearable sensors to enable gesture-based assistive tasks [17]. Most such systems, however, rely on fixed camera setups or limit users to remain within highly limited working areas.

While some assistive robot systems allow for basic customization, few support on-demand reconfiguration of tooling or real-time adjustment to user positioning. Moreover, there remains limited integration between gesture recognition, modular end-effectors, and camera motion capable of tracking mobile users in dynamic environments.
Our work addresses these gaps by introducing a gimbal-mounted depth camera system, 3Dprinted modular tooling and a training gesture recognition framework for adaptable, inclusive smart work cells. These technologies overall enable more intuitive and effective human-robot collaboration, particularly for older adults and individuals with disabilities, by supporting adaptive and responsive robot assistance in task execution [24].

Relative to past work that relied on fixed tooling setups and static cameras, our system stands out in its combination of gimbal-based user tracking, modular tooling through 3D printing, and an ultra-light machine learning pipeline for real-time gesture detection. These features in combination enable greater flexibility and usability, particularly for dynamic and unstructured user workspaces for individuals with physical disabilities. Recent studies in adaptive HRI [25], multimodal user gesture recognition, and real-time assistive robotics increasingly call for flexible, lightweight systems, which this research explicitly targets [26], [27].

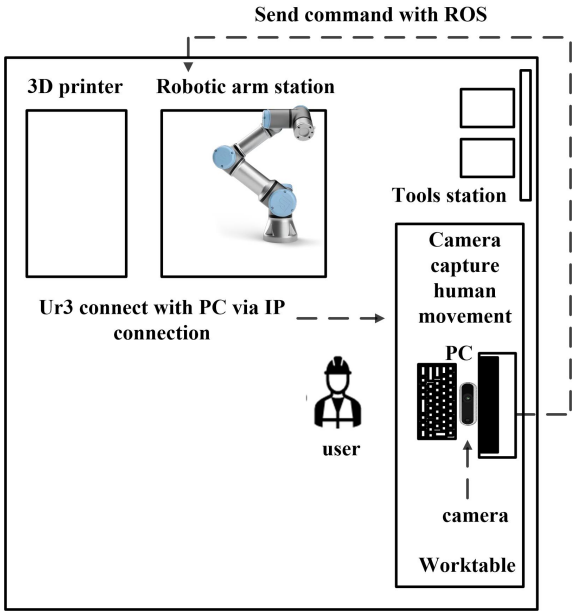

Figure 1: Workcell layout

Figure 2: Robotic arm station

3. SYSTEM OVERVIEW

This chapter provides an overview of the system at the heart of our research to produce an efficient workspace. It outlines how the various systems and subsystems that make up our research collaborate in a way that enhances usability and efficiency.

3.1. Work cell concept

The assistive work cell is designed to mimic a typical office setting, with a desk and a computer as the central focus. To increase functionality, a different support system for tool placement has been implemented and with it, the robotic arm can be used productively. The design of the work area is discussed in Fig. 1. A gimbal-like depth camera is constantly monitoring the movements of the user in a way that ensures the user's actions remain within the camera's field of view. The system makes recording human movement more precise so that the system can interpret the gestures of the user better and transmit this to the robotic arm for execution. The hardware system consists of an Intel Core i7 CPU, NVIDIA GeForce GTX Titan X GPU, and Ubuntu 20.04.6 operating system. The configuration is consistent with human-centered robotic workspace practices, with the aim of intuitive and ergonomic collaboration [28].

3.2. Robotic arm station

Fig. 2 illustrates a tool station with sections containing various tools that are mounted on grappling hooks and a robotic arm station. This is an effective and efficient way of utilizing space and obtaining tools in time. It particularly

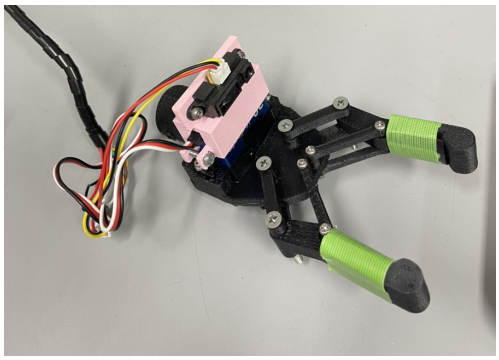

Figure 3: Gripper model

proves to be effective in work cell spaces where the robotic arm should have easy access to other tools.

The robotic arm station features the Universal Robots UR3e arm due to its precision, reliability, and fine manipulation capability. The UR3e has a total of six joints, a payload of 3 kg, a reach of 500 mm, and a weight of 11.2 kg. The robot arm is placed on the left side of the user to facilitate easy access to the tool and efficient task accomplishment.

For user choice and work processes, two functioning modes are in place: Manual mode using a control pad and an automatic Gesture Recognition mode.

Manual Control: The Robotic arm is controlled with a pad with direct, easy change adjustments in real-time. Implication is the offering of a natural and easy-to-read interface for facilitating the fine control of instructions.

Automatic Gesture Reading: This mode utilizes a vision-based approach to the interpretation of human gesture [23]. Integrated with depth camera feedback and robot motion, machine learning algorithms read and predict what the user wants to do to allow human-robot interaction.

3.3. Gripper units

One of the most notable aspects of this research is the gripper that has been specially made, which was created using 3D printing for modularity and flexibility to be able to do a variety of tasks [29]. It is easy to remove and install, allowing effortless switching between tasks. 3D printing makes it of immense advantage with the ability to quickly produce and adjust the design incrementally. A 3D print station, in the back of the robot arm, allows rapid replacement of failed parts or new design generation and testing. Flexibility through adaptation allows increased versatility in using the system for many different applications, to meet different users' needs, and various activities. As shown in Fig. 3, the standard gripper has a short extension on the tip for disengaging toolboxes from grappling hooks.

4. HUMAN GESTURE RECOGNITION

This research aims to develop vision-based human gesture identification based on the analysis and tracking of human body movement. This is acquired by real-time head orientation, facial pose, and hand gesture monitoring to ascertain user activity. Converting camera image data into usable coordinate systems for use in robotic systems represents a key application of this process, and it involves top-down mathematical computation. Furthermore, the sole reliance on visual data in machine learning algorithms might lead to diminished accuracy, especially in ambiguous or complex situations.

For such challenges to be addressed, in the current research work, Intel® RealSense™ Depth Camera D435 is utilized for video capture, supported by the MediaPipe Holistic model. Through this combination, a holistic analysis of human facial behavior, body pose, and hand gestures can be made. MediaPipe Holistic extracts significant key points from images, enabling accurate interpretation of human pose and gesture. Such key points provide significant information on body location and movement, enabling the system to be more efficient in analyzing human motion. Fig. 4 illustrates the overall system of the assistive robotic arm. The first step is to acquire human gesture data by using the depth camera. The data, after being acquired, is processed and analyzed in order to create machine-learning models that are capable of decoding human gestures. These models, in return, assist in guiding the robotic arm in determining human gestures.

The MediaPipe Holistic model provides a real-time output of 543 landmarks: 33 for body pose, 468 for facial expressions, and 42 for hand gestures (21 for each hand). The landmarks are connected by lines that draw the body structure, face, and hands in real-time.

In [13], a human gesture prediction system using a full-body dataset and several training models are presented. Despite the accuracy of the system being good, it must employ a big dataset, leading to long data acquisition, long training duration, and huge computational overhead. The system also requires the camera to remain stationary at

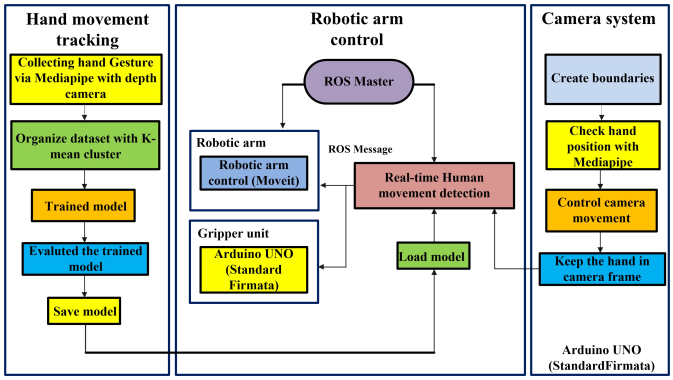

Figure 4: Workflow of the system

the same location where data collection was done for it to be accurate. To get around these constraints, our work focuses on eliminating redundant information and retaining the most informative features. Since most productive behaviors are achieved through the hands, our system focuses on hand gesture data, loading filters to recognize informative data.

The system initially automatically picks an available RealSense camera and sets it up to stream high-definition depth and color data. Real-time preprocessing techniques, such as horizontal flip and Gaussian blur, enhance data quality before landmark detection. MediaPipe is used to obtain hand landmarks from each frame. Due to variability in hand size and camera distance in users, raw landmark coordinates may not be reliable. For this, the wrist is set as the origin point, and angle-based features are calculated to account for hand size variation.

Hand orientation—quantified in terms of roll, pitch, and yaw—is computed by constructing a rotation matrix from vectors between selected hand and arm landmarks, and then transforming into Euler angles. Gestures are tagged, and corresponding information is stored in a CSV file. Hand motion is tracked in 3D by the system, with the X-axis representing horizontal movement, the Y-axis vertical location, and the Z-axis depth in camera space. 12 gestures were recorded over 76,296 frames.

In the approach of [13], various machine learning classifiers were tried out, including logistic regression, random forest, gradient boosting, support vector classifier, and ridge classifier. Voting Classifier with hyperparameter tuning had slightly higher accuracy than single models, albeit at the cost of increased training time. To reduce computational load and maintain or improve performance, our strategy incorporates K-Means clustering to divide the dataset into 100 groups and eliminate duplicate samples. Correlation analysis also helps in identifying important features and reducing dimension. Machine learning pipelines integrate StandardScaler feature scaling and Ridge Classifier, XGBoost, and Voting Classifier classifiers. Training optimization techniques such as K-Means clustering and feature correlation analysis reduce data redundancy and training time.

System performance in gesture recognition is graphed in the format of a confusion matrix in Fig. 5. After successful model validation and training had been conducted, the system was integrated with a robotic arm platform. With such an integration, the robot can classify gestures and respond accordingly based on them, achieving real-time detection and user intent response.

To facilitate dynamic situations, an Intel® RealSense™ Depth Camera D435 was mounted on a gimbal stabilizing system to allow the camera to continuously track the user even while roaming in the workspace. This enables a stable field of view and reduces the likelihood of losing gesture tracking due to user movement. The gimbal pan-tilt feature enhances overall system stability and accuracy during dynamic situations. The camera-gimbal assembly is shown in Fig. 6.

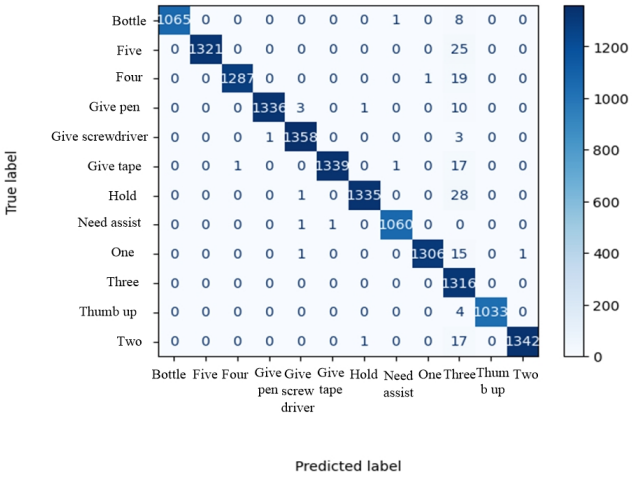

Figure 5: Confusion Matrix

Figure 6: Depth camera unit

RealSense camera captures RGB and depth at 640×480 pixels at 30 fps. Depth stream is utilized to enable depth-based perception, which enables hand tracking, occlusion, and gesture. Raw depth values are converted to real-world metric values using the depth scale factor of the camera.

Gesture recognition is stabilized using a Kalman filter over the 3D landmark coordinates. A gimbal-mounted RealSense camera enhances robustness by maintaining focus on the user's hand movements even during motion, reducing tracking failures. The gimbal mechanism is constructed with two high-torque 20 kg servo motors (270-degree rotation) for pan and tilt motion. They are powered through an Arduino UNO with the StandardFirmata protocol, which provides real-time serial control by a host system. Servo motion ranges are restricted to avoid overextension and offer safety. A depth-based occlusion detection system is

also implemented. Objects within 0.5 meters of the camera are labeled as potential occlusions. Upon detection, a visual alert is presented, requesting the user to reposition, thereby guaranteeing gesture visibility.

The robotic arm is controlled through the Robot Operating System (ROS), employing "rospy" to facilitate intercommunication between the human gesture recognition module and robotic arm control system. The driving mechanism to manipulate the robotic arm is done utilizing the "moveit" library to manage individual joint movement to bring it to a specific destination. Arduino UNO controls the gripper section by using the Stand Firmata library, which makes it possible for Arduino to be coded in Python. The motor used in the gripper unit is a 20 kg servo motor with 270-degree control. The example of the system is discussed in Fig. 7. The program detects user gestures in real-time. In this case, it recognizes the gesture as "Bottle", which is a command instructed to the robotic arm to remove an empty bottle off the worktable, which, in this case, the robotic arm will carry a bottle and move it into the trash zone.

5. EXPERIMENT

The experimental setup for the validation of the performance of the system in detecting human gesture and controlling the robotic arm is illustrated in Fig. 8. Systematic experiments were conducted to evaluate the system's ability to accurately interpret human gestures and translate them into correct robotic movements.

To ensure robustness and avoid spurious activations, a gesture confidence threshold of 80 was set. This is a filtering technique that discards low-confidence predictions, normally produced by the system when attempting to interpret non-gesture or indeterminate movement. To further enhance reliability, the system conducts five consecutive gesture tests. The robot arm receives a command only when all five tests produce the same result and are above the confidence level.

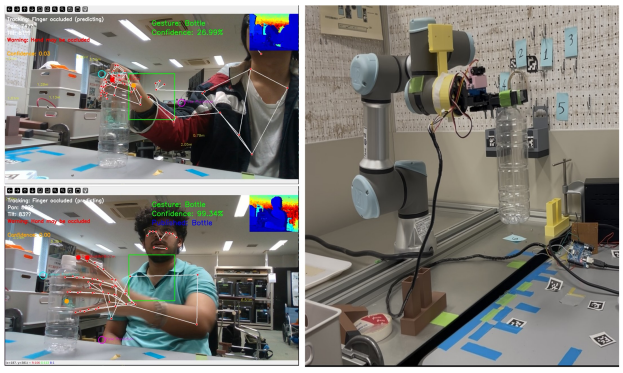

Figure 7: Gesture recognition and robotic response in the 'Bottle' scenario

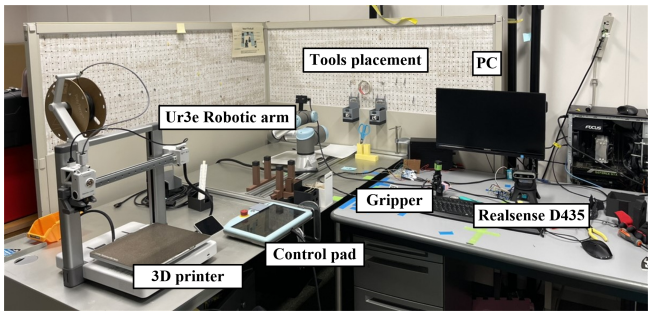

Figure 8: Test bench used for evaluating the system

Fig. 9 depicts pre-defined gesture patterns used in experiments.

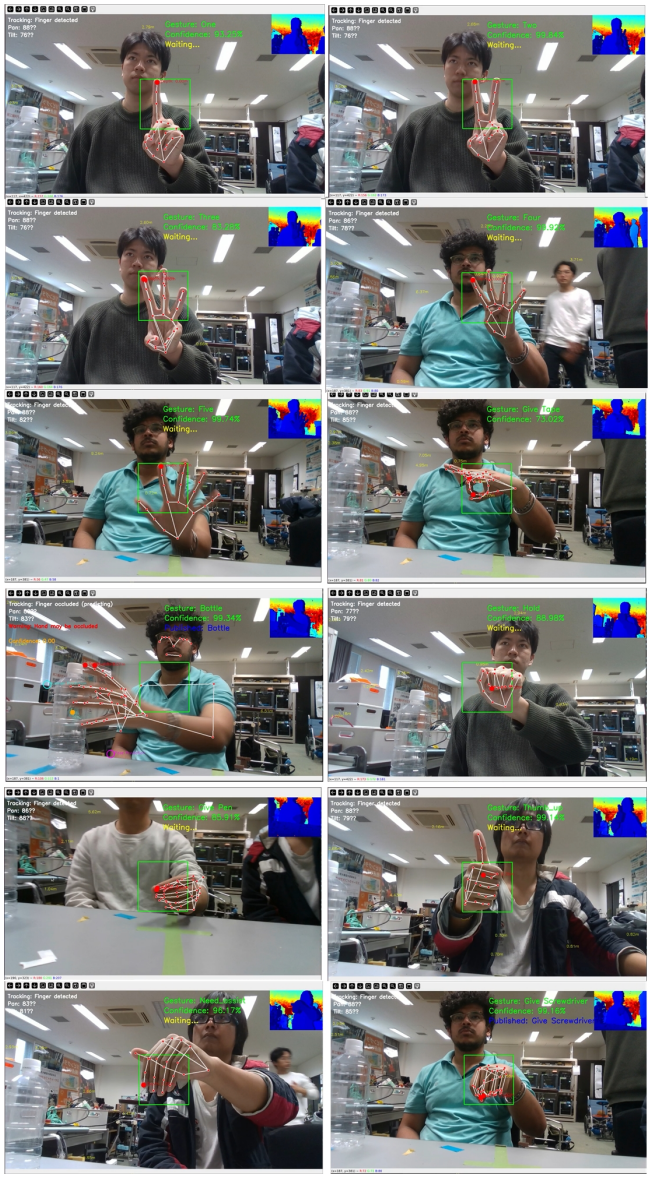

Figure 9: Gestures used in the experiment

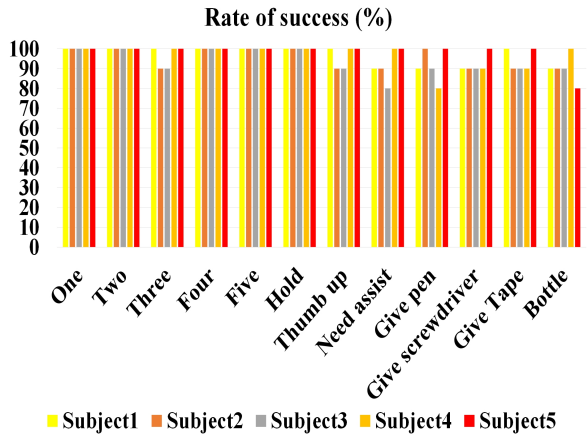

Figure 10: User Accuracy for Human Gesture Reading

The gestures were chosen deliberately for testing the system over a large range of real-world usage scenarios with different body postures, motion profiles, and interaction distances. The subjects were also asked to move about freely within the workspace, intentionally testing the gimbal mechanism to evaluate its performance for accurate tracking under dynamic situations.

The tests involved five healthy male subjects between 22 and 32 years of age. The current experiment was conducted with healthy male participants to validate the system's basic operation and safety. The system targets older adult individuals and disabled individuals, but this proof-of-concept test is just a beginning. Trials with older adults and motor-impaired persons will be part of subsequent experiments to confirm the effectiveness and usability of the system in our target population. The subjects were oriented with the experiment's purpose, safety procedures, and data confidentiality practices before the tests were carried out.

Most significantly, three out of five subjects (Subjects 2, 3, and 5) did not participate in data collection for training. This made it possible to evaluate the capability of the system to generalize across new users. User Accuracy in the experiment can be seen from Table I and Fig. 10. Each subject performs each task 10 times. Recognition performance was satisfactory across all the subjects, regardless of their orientation and location in front of the camera. This success is credited largely to the training method, which employed a broad range of hand positions, angles, and distances to build a robust model.

Performance results show good recognition. Average accuracy exceeded 90% across all gestures and subjects. Misclassifications primarily occurred between visually similar gestures (e.g., 'Three' vs. 'Bottle' or 'Give pen' vs. 'Give screwdriver') but were reduced using gesture verification.

Average gesture recognition response times are quite difficult as they vary from each person, where they position themselves or how fast they move their body, which made user accuracy a more relevant metric than response time.

Notably, Subject 3, who had a smaller body size than others, showed no detectable decline in recognition accuracy. This further suggests that the system appropriately generalizes between users with varying physical characteristics. Moreover, performance was significantly improved utilizing the gimbal-based camera tracking system by maintaining focus on the user's hand movements for guaranteed gesture detection even during motion.

As a whole, experimental results validate the system's ability in gesture recognition and robotic arm actuation for different user conditions, affirming its viability in real-world applications of human-robot collaboration.

Future work will extend testing to target users, including older adults and individuals with disabilities, to validate usability across a broader population.

6. CONCLUSIONS

This research was intended to develop a system that can analyze the gestures of humans to enable a robotic arm to provide services for users, especially those who are physically disabled or older adults, to perform work-related tasks. With a focus on real-world work situations, the system is developed to improve the efficiency of tasks and provide more autonomy for physically disabled users.

One of the most distinctive features of the system is the modular, 3D-printed gripper unit mounted on the robotic arm, showcasing both the versatility and the practicality of the platform. The addition of a gimbal mechanism also makes the system more robust by minimizing tracking errors due to user motion or position shifts.

The experimental results indicate that some user training is required for optimal system performance. This includes acclimatization to the gesture vocabulary, learning optimal positioning, and understanding the range of operations that the robotic arm can perform. While preliminary tests with healthy participants demonstrate promising performance, the system has yet to be validated with older adults and individuals with disabilities, the intended user population. Future work will involve trials with diverse participants to assess real-world usability and inform further refinements.

References

[1] N. Mandischer, M. Gürtler, C. Weidemann, E. Hüsing, S.-O. Bezrucav, D. Gossen, V. Brünjes, M. Hüsing, B. Corves, Toward adaptive human–robot collaboration for the inclusion of people with disabilities in manual labor tasks, Electronics 12 (5) (2023) 1118. doi:10.3390/electronics12051118.
URL http://dx.doi.org/10.3390/electronics12051118

[2] A. Bonello, E. Francalanza, P. Refalo, Smart and sustainable human-centred workstations for operators with disability in the age of industry 5.0: A systematic review, Sustainability 16 (1) (2023) 281. doi:10.3390/su16010281.
URL http://dx.doi.org/10.3390/su16010281

[3] E. Hüsing, C. Weidemann, M. Lorenz, B. Corves, M. Hüsing, Determining robotic assistance for inclusive workplaces for people with disabilities, Robotics 10 (1) (2021) 44. doi:10.3390/robotics10010044.
URL http://dx.doi.org/10.3390/robotics10010044

[4] F. Mohammadi Amin, M. Rezayati, H. W. van de Venn, H. Karimpour, A mixed-perception approach for safe human–robot collaboration in industrial automation, Sensors 20 (21) (2020) 6347. doi:10.3390/s20216347.
URL http://dx.doi.org/10.3390/s20216347

[5] J. Arents, V. Abolins, J. Judvaitis, O. Vismanis, A. Oraby, K. Ozols, Human–robot collaboration trends and safety aspects: A systematic review, Journal of Sensor and Actuator Networks 10 (3) (2021) 48. doi:10.3390/jsan10030048.
URL http://dx.doi.org/10.3390/jsan10030048

[6] K. Tylutki, Ż. Konopacka, J. Woźniak, D. Sołatycka, Adaptation of the workplace for disabled people—sustainable participation in the labor market, Sustainability 16 (17) (2024) 7473. doi:10.3390/su16177473.
URL http://dx.doi.org/10.3390/su16177473

[7] S. Drolshagen, M. Pfingsthorn, A. Hein, Context-aware robotic assistive system: Robotic pointing gesture-based assistance for people with disabilities in sheltered workshops, Robotics 12 (5) (2023) 132. doi:10.3390/robotics12050132.
URL http://dx.doi.org/10.3390/robotics12050132

[8] O. P. Narenthiran, J. Torero, M. Woodrow, Inclusive design of workspaces: Mixed methods approach to understanding users, Sustainability 14 (6) (2022) 3337. doi:10.3390/su14063337.
URL http://dx.doi.org/10.3390/su14063337

[9] J. Zhao, Y. Wang, Y. Cao, M. Guo, X. Huang, R. Zhang, X. Dou, X. Niu, Y. Cui, J. Wang, The fusion strategy of 2d and 3d information based on deep learning: A review, Remote Sensing 13 (20) (2021) 4029. doi:10.3390/rs13204029.
URL http://dx.doi.org/10.3390/rs13204029

[10] N. Chen, C. Zuo, E. Lam, B. Lee, 3d imaging based on depth measurement technologies, Sensors 18 (11) (2018) 3711. doi:10.3390/s18113711.
URL http://dx.doi.org/10.3390/s18113711

[11] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: CVPR 2011, IEEE, 2011, p. 1297–1304. doi:10.1109/cvpr.2011.5995316.
URL http://dx.doi.org/10.1109/CVPR.2011.5995316

[12] N. Dimitropoulos, G. Michalos, Z. Arkouli, G. Kokotinis, S. Makris, Industrial collaborative environments integrating ai, big data and robotics for smart manufacturing, Procedia CIRP 128 (2024) 858–863. doi:10.1016/j.procir.2024.04.027.
URL http://dx.doi.org/10.1016/j.procir.2024.04.027

[13] N. Suppaadirek, M. Sonnic, R. A. D. Jimenez, T. Shibata, Design and development of a work cell with a one-handed soldering tool for enhanced human-robot collaboration, in: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2024, p. 7123–7130. doi:10.1109/iros58592.2024.10801987.
URL http://dx.doi.org/10.1109/IROS58592.2024.10801987

[14] Y. Kong, Y. Fu, Human action recognition and prediction: A survey (2018). doi:10.48550/ARXIV.1806.11230.
URL https://arxiv.org/abs/1806.11230

[15] D. Gupta, A. K. Singh, N. Gupta, D. K. Vishwakarma, Sdlnet: A combined cnn & amp; rnn human activity recognition model, in: 2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT), IEEE, 2023, p. 1–5. doi:10.1109/apsit58554.2023.10201657.
URL http://dx.doi.org/10.1109/APSIT58554.2023.10201657

[16] R. C. Hsu, P.-C. Su, J.-L. Hsu, C.-Y. Wang, Real-time interaction system of human-robot with hand gestures, in: 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), IEEE, 2020, p. 396–398. doi:10.1109/ecice50847.2020.9301957.
URL http://dx.doi.org/10.1109/ECICE50847.2020.9301957

[17] Y. Cheng, P. Yi, R. Liu, J. Dong, D. Zhou, Q. Zhang, Human-robot interaction method combining human pose estimation and motion intention recognition, in: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2021, p. 958–963. doi:10.1109/cscwd49262.2021.9437772.
URL http://dx.doi.org/10.1109/CSCWD49262.2021.9437772

[18] N. Robinson, B. Tidd, D. Campbell, D. Kulić, P. Corke, Robotic vision for human-robot interaction and collaboration: A survey and systematic review, ACM Transactions on Human-Robot Interactiondoi:10.48550/ARXIV.2307.15363.
URL https://arxiv.org/abs/2307.15363

[19] S. Jeeru, A. K. Sivapuram, D. G. León, J. Gröli, S. R. Yeduri, L. R. Cenkeramaddi, Depth camera based dataset of hand gestures, Data in Brief 45 (2022) 108659. doi:10.1016/j.dib.2022.108659.
URL http://dx.doi.org/10.1016/j.dib.2022.108659

[20] A. Ramey, V. González-Pacheco, M. A. Salichs, Integration of a low-cost rgb-d sensor in a social robot for gesture recognition, in: Proceedings of the 6th international conference on Human-robot interaction, HRI'11, ACM, 2011, p. 229–230. doi:10.1145/1957656.1957745.
URL http://dx.doi.org/10.1145/1957656.1957745

[21] N. Zengeler, T. Kopinski, U. Handmann, Hand gesture recognition in automotive human–machine interaction using depth cameras, Sensors 19 (1) (2018) 59. doi:10.3390/s19010059.
URL http://dx.doi.org/10.3390/s19010059

[22] S. Gore, S. Hamsa, S. Roychowdhury, G. Patil, S. Gore, S. Karmode, Augmented intelligence in machine learning for cybersecurity: Enhancing threat detection and human-machine collaboration, in: 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), IEEE, 2023, p. 638–644. doi:10.1109/icaiss58487.2023.10250514.
URL http://dx.doi.org/10.1109/ICAISS58487.2023.10250514

[23] C. Y. Wong, L. Vergez, W. Suleiman, Vision- and tactile-based continuous multimodal intention and attention recognition for safer physical human-robot interaction (2022). doi:10.48550/ARXIV.2206.11350.
URL https://arxiv.org/abs/2206.11350

[24] P. Prajod, M. L. Nicora, M. Malosio, E. André, Gaze-based attention recognition for human-robot collaboration (2023). doi:10.48550/ARXIV.2303.17619.
URL https://arxiv.org/abs/2303.17619

[25] B. Karbouj, K. A. Rashwany, O. Alshamaa, J. Krüger, Adaptive behavior of collaborative robots: Review and investigation of human predictive ability., Procedia CIRP 130 (2024) 952–958. doi:10.1016/j.procir.2024.10.190.
URL http://dx.doi.org/10.1016/j.procir.2024.10.190

[26] S. Trick, D. Koert, J. Peters, C. Rothkopf, Multimodal uncertainty reduction for intention recognition in human-robot interaction (2019). doi:10.48550/ARXIV.1907.02426.
URL https://arxiv.org/abs/1907.02426

[27] P. Franceschi, D. Cassinelli, N. Pedrocchi, M. Beschi, P. Rocco, Design of an assistive controller for physical human–robot interaction based on cooperative game theory and human intention estimation, IEEE Transactions on Automation Science and Engineering 22 (2025) 5741–5756. doi:10.1109/tase.2024.3429643.
URL http://dx.doi.org/10.1109/TASE.2024.3429643

[28] B. Venkatesh, J. Enright, K. Rackley, H. Asada, Towards human-robot collaborative workcells for industrial assembly applications, in: IEEE/RSJ IROS, 2019, pp. 7258–7265.

[29] M. Tenorth, D. Jain, M. Beetz, Robotic picking and placing of objects in cluttered scenes, in: ISER, Springer, 2013, pp. 311–320.

Table 1: User Accuracy for Human Gesture Reading

Class	**Test Subject**				
	Subject1	**Subject2**	**Subject3**	**Subject4**	**Subject5**
One	100	100	100	100	100
Two	100	100	100	100	100
Three	100	90	90	100	100
Four	100	100	100	100	100
Five	100	100	100	100	100
Hold	100	100	100	100	100
Thumb up	100	90	90	100	100
Need assist	90	90	80	100	100
Give pen	90	100	90	80	100
Give screwdriver	90	90	90	90	100
Give Tape	100	90	90	90	100
Bottle	90	90	90	100	80