

An Auto-Correction Teleoperation Method for a Mobile Manipulator Using Gaze Tracking and Hand Motion Detection ^{*}

Junshen Chen¹, Ze Ji^{1*}, Hanlin Niu¹, Rossitza Setchi¹, and Chenguang Yang²

¹ School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

* jiz1@cf.ac.uk

² Bristol Robotics Laboratory, University of the West of England,
Bristol BS16 1QY, UK

Abstract. Situational awareness in remote environments is crucial for human operators to teleoperate mobile manipulators confidently and reliably. Visual feedback is the most common way for environment perception, providing rich information to human operators. This paper proposes an intuitive teleoperation method by combining gaze tracking and hand motion detection to teleoperate a mobile manipulator. A camera is fixed on the end-effector of the mobile robot's arm to provide visual feedback, acting as the eye of the teleoperator. Rather than direct remote control of the robot, an adaptive auto-correction mechanism is introduced for helping human operators to achieve better hand-eye coordination of the teleoperation experiences. The mobile manipulator can adjust its behaviours, such as speed, while gaze and hand movements of the operator are in different states. The experiments carried out demonstrated the effectiveness of the proposed algorithm and the survey evaluation verified the practical application value of the system.

Keywords: Teleoperation · Hand-Eye Coordination · Mobile Manipulator

1 Introduction

Telerobotics has been widely applied in both research and industrial fields. Teleoperated robots, also known as telerobots, are often compared with fully automated control robots. Automated robots are often used to perform fixed tasks, while telerobots are designed to work in more unstructured environments and collaborate with humans on more diverse tasks.

A mobile manipulator is a system consisting of a mobile robot mounted with one or more robotic manipulators. There are a diverse range of applications of

^{*} This work is supported by the Advanced Sustainable Manufacturing Technologies (ASTUTE) 2020 and ASTUTE East projects. We would like to thank the visiting MSc students from INSA-Strasbourg, France, for their previous work on the Smart Eye system.

teleoperated mobile manipulators, ranging from large-scale outdoor applications, such as the Mars Rovers [9] for space exploration and military operations, to indoor domestic environments, such as home-care and health-care applications [8].

There are various modalities nowadays allowing human operators to interact with robots, including voice recognition, human body gestures, joystick, gaze tracking, hand gestures, and combinations of them. This work aims to investigate the performance of combining gaze tracking and hand gesture recognition to provide an intuitive and natural interactive modality for mobile manipulators.

Nowadays, gaze-tracking and hand-tracking systems are manifold, and the applications keep on increasing. Video games and virtual reality are two examples of general public usage. In the scientific community, new interfaces using such technologies are being researched and developed, in particular, for human-computer interactions, and notably for human-robot interactions. A gaze tracking teleoperation method was proposed in [11], and it also offered a 2-D mapping-based calibration method for a user's head movement. In [4], experiments were taken out for letting human operators teleoperate a robot arm to write and draw by using their gaze. An interface designed in [7] enables the user to teleoperate the robot using his hands' motion. The user can make an online adjustment for the autonomy of the robot. Another hand movements teleoperation system was introduced in [3]. In [5], it presented a robot teleoperation platform that used a user interface based on eye-gaze tracking that users can navigate a teleoperated mobile robot using their eyes. Interval Kalman filter and improved particle filter are applied to realise a markerless and contactless teleoperation interface.

However, for a hand-eye combined system of teleoperation, as developed in this work, hand-eye coordination is an important factor to gain better teleoperation experience. Eye-hand coordination, or hand-eye coordination, is the ability to simultaneous use of eyes and hands to fulfil tasks, much like an activity that uses the visual information to guide movements of hands. For example, drivers are required to constantly use hand-eye coordination that visual information needs to be translated into actions executed by hands on the steering wheels, in order to keep the car in the middle of the lane to avoid accidents.

In [10], the perspective of uncalibrated hand-eye coordination has been studied on a humanoid robot. Online image Jacobian matrix was identified by using Kalman filter. A microsurgery system used in minimally invasive surgery, which requires a high level of hand-eye coordination, was introduced in [6]. To eliminate the changing orientation of the object from visual feedback, a disorientation auto-recovery algorithm was applied to calibrate the rotation angle in real time. This helps to simplify manual dexterity among surgeons. However, no related researches have been carried out for enhancing hand-eye coordination while teleoperating mobile manipulators.

In this work, an intuitive teleoperation method for remote controlling a mobile manipulator is developed by combining gaze tracking and hand motion recognition. A camera is fixed on the end-effector of the mobile robot's arm to provide visual feedback, acting as the eye of the teleoperator. Rather than

direct remote control of the robot, an adaptive auto-correction method is introduced for helping human operators to gain better hand-eye coordination of the teleoperation experience. The mobile manipulator is able to adjust its behaviours, especially its speed, while gaze and hand movements of the operator are in different states. These methods will help to gain better experience while teleoperating a mobile manipulator using hand motion and gaze simultaneously. Also, it helps to minimise distractions while teleoperating, as well may help operators with poor hand-eye coordination.

2 System Description

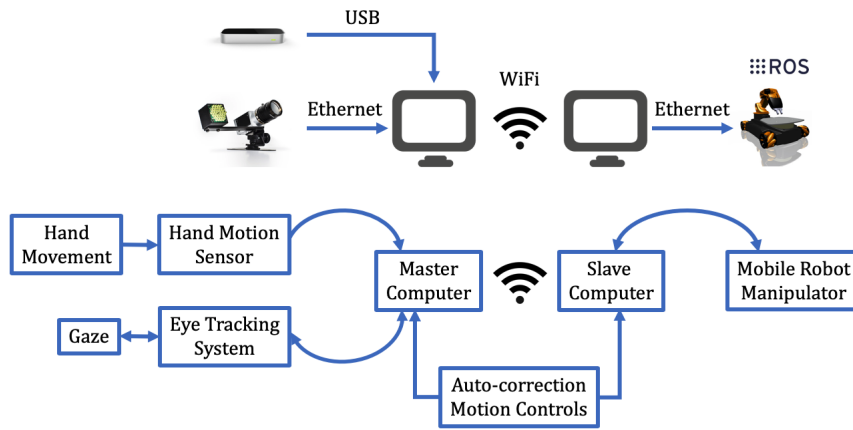


Fig. 1. The overall control system structure.

To enhance teleoperation experience, in this paper, a gaze tracking and hand motion detection combined method is used for teleoperating a KUKA Youbot. The Youbot is an omnidirectional mobile robot platform with a 5 Degree of Freedoms (DoFs) robot manipulator and a 2-finger gripper. A Smart Eye[®] system is used to track a human user's gaze, which is associated to the behaviours of the Youbot's manipulator in this work, mainly involving the rotational motions or orientations of the Youbot's manipulator. A camera is attached to the robot arm's end-effector, and allows visualisation to the users of the environment, such as obstacles, from the robot's perspective during teleoperation. User's hand motion is detected by a Leap Motion[®] sensor. The gestures of the right hand are translated to the Youbot's base behaviours for teleoperation, with the same direction as the operator's hand positioned over the sensor. The combined system provides the user an intuitive experience while operating the telebot.

In addition, the speed of Youbot has corrected adaptively according to the current states of the gaze tracked and hand motion. When the user is looking forward and pointing forward, the system will consider that the user has full attention on the robot so that the robot would run in full speed on its base. When the operator is looking and pointing to the same side, the speed of Youbot's base will be reduced. The movement of the robot's manipulator will also slow down until the direction of gaze and hand motion are both pointing forward again. The robot will gradually slow down to stationary when the user is looking and pointing in two different sides. The overall communications system is shown in Fig. 1.

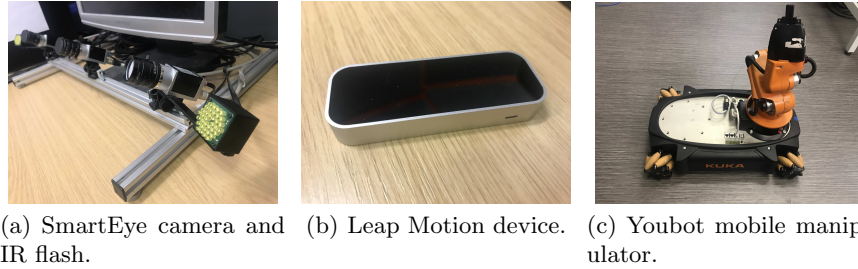


Fig. 2. Sensors and robot used for the teleoperation system.

2.1 Gaze Control with Smart Eye

The device used for this application is the Smart Eye Pro from the company Smart Eyes. It consists of two infra-red projectors and three high-quality cameras. The third camera is used to improve robustness and accuracy by creating redundant data. Fig. 2(a) shows one of the cameras and an infra-red projector. The system is combined with a desktop computer, whose screen is used to display video streams broadcasted from the embedded camera on the robot.

The device is provided with a specific piece of software called *Smart Eye*, which requires a two-phase calibration before use and also can be used to monitor gaze behaviours of operators. Those data can be stored in a log file or sent in real time to third-party software.

The Smart Eye system is a stereoscopic device, which requires a first calibration step to determine the relative position and orientation of the three cameras and the IR flashes. The calibration is done using a chessboard, which is monitored by all the cameras at the same time. The extrinsic parameters are stored and should not change for the duration of the experimentation.

The result of camera calibration can be evaluated by a calibration verification process. By presenting again the chessboard to the cameras, the SmartEye software computes and displays the average difference in pixels between its model

and the reality. The calibration is considered accurate, when the value displayed is sub-pixelic. Otherwise, a new calibration phase will be repeated.

The origin and the orientation of the 3D-world have also to be done by defining a world coordinate system. The origin of the head's roll, pitch and yaw and the intersection between the sight and the screen depends on how precisely the world coordinate system has been established.

The second calibration phase is to improve the performance of eye tracking by estimating and correcting the errors caused by the differences of each individual user. By asking the user to stare at four pre-defined targets on the screen, the Smart Eye system can compute the differences between the estimated coordinates of users' attention on the screen and the desired coordinates at the four pre-defined targets. The errors are compensated to make a correction for gaze tracking of each individual user. The direction of the sight is computed in real time. Fig. 3 shows the graphical interface of the gaze tracking system.

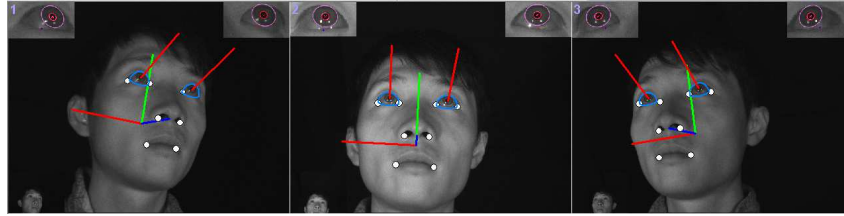


Fig. 3. User gaze tracking calibrated results.

2.2 Hand detection with Leap Motion

The tracking device used is the *Leap Motion*[®], which captures hands and fingers' motions as input, providing a low-cost structured light based motion capture system. Compared to a mouse, no contact or touching is required here. The device has been conceived to be embedded into a virtual reality headset, but also to be used as a standalone for various applications.

The device illustrated in Fig. 2(b) is a small module of size $80mm \times 30mm \times 11.25mm$. It is composed of two cameras and three infra-red lights. Those cameras track the interactions in an area shaped like an inverted pyramid in which field of view is $120^\circ \times 150^\circ$ for a depth of approximately $600mm$. According to the manufacturer, the accuracy of the Leap Motion is about $0.7mm$ inside this zone.

2.3 Youbot

The *Youbot* (Fig. 2(c)) is an omnidirectional mobile manipulator designed by the German robotic company *KUKA*. It was conceived for research and education

purposes [1]. This robot is composed of two parts: A rectangular-shaped mobile base moving in the plan using four Swedish wheels; A robotic arm made of five degrees of freedom. A gripper is located at the tip of the arm, as well as a bracket intended to fix a camera.

Some applications already use gaze to control mobile robots. In this work, the particularity of the Youbot, which is to combine the assets of mobile robots and robotic arm, will be fully exploited to develop the system. Indeed, as the camera is set at the tip of the arm, the whole kinematic chain of the robot is used to control precisely the pose of the camera as the user requires it.

3 Method

This section describes the strategies of how user gestures are linked to the behaviours of the robot. The joint angles \underline{q} of the robot can be described as

$$\underline{q} = f^{-1}(\underline{X}) \quad (1)$$

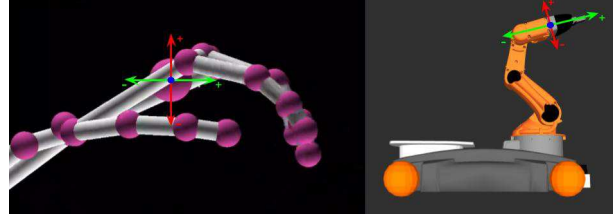
where $\underline{q} \in R^5$ is the joint angles of the manipulator, \underline{X} is the end-effector coordinate, represented as $\underline{X} = [X_E, Y_E, Z_E, \phi, \Omega, \psi]'$ and X_E , Y_E and Z_E are the corresponding positions on the X , Y and Z axes of the end effector, whereas ϕ , Ω and ψ are the roll, pitch and yaw of the end effector, respectively.

Controlling the robot using hand gestures is straightforward. Fig. 4(a) shows the relationship between the left hand gesture and the robot end-effector, and Fig. 4(b) illustrates the relationship between the right hand gesture and robot displacement. The right hand controls the base movement in all directions, by translating the relative pose between the hand and the sensor to speed control commands for the robot, where the distance from the hand to the sensor is proportional to the robot speed. The left hand controls the end-effector; however, it only deals with its coordinate, (X_E, Y_E, Z_E) , rather than orientations (β, θ_1) , which are controlled by the gaze gestures.

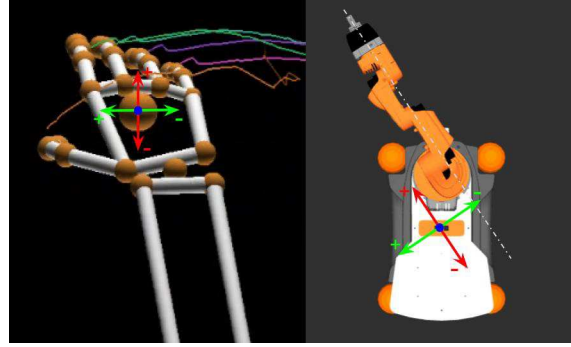
3.1 Impact of gaze commands

The arm receives three values from the user intended to orientate the end-effector: x_{screen} , y_{screen} and $head_{roll}$, where x_{screen} represents the horizontal position of the sight on the screen. 0 means that the user is looking at the left edge of the screen, 0.5 at the middle and 1 at the right edge, y_{screen} represents the vertical position of the sight on the screen. 0 means that the user is looking at the bottom edge of the screen, 0.5 at the middle and 1 at the top edge and $head_{roll}$ represents the user's head roll in radian. A positive angle means that the user is bending its head to the left and a negative means that the head bends to the right.

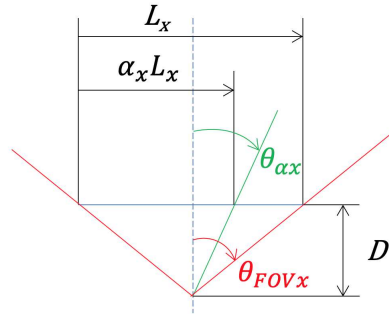
Considering that the screen displays in real-time the images captured by the camera, Fig. 5 illustrates the relationship between the field of view of the camera (horizontal here) and the computer screen, where:



(a) Left hand motion control and end-effector displacement



(b) Right hand motion control and robot base displacement

Fig. 4. Hand motion control and robot displacement**Fig. 5.** Orientation of the camera according to the sight.

- L_x represents the width of the computer screen.
- θ_{FOVx} represents the horizontal camera field of view
- α_x is equal to x_{screen} .
- θ_{ax} is the output angle corresponding at the angular displacement of the camera.
- D is a variable which represents the virtual focal of the system represented.

According to the Fig. 5, the following equations can be written:

$$\begin{cases} \tan \theta_{FOVx} = \frac{L_x}{2D} \\ \tan \theta_{\alpha x} = \frac{L_x \alpha_x - L_x/2}{D} \end{cases} \quad (2)$$

which leads easily to the relation:

$$\theta_{\alpha x} = \text{atan}[(2\alpha_x - 1)\tan\theta_{FOVx}] \quad (3)$$

Likewise, the following equation can be obtained:

$$\theta_{\alpha y} = \text{atan}[(2\alpha_y - 1)\tan\theta_{FOVy}] \quad (4)$$

where θ_{FOVy} represents the vertical camera field of view, α_y is equal to y_{screen} , $\theta_{\alpha y}$ is the output angle corresponding at the angle displacement of the camera.

Therefore, $\theta_{\alpha x}$ and $\theta_{\alpha y}$ are the angular variation of the end-effector, with respect to x_{screen} and y_{screen} . Eq.(3) and Eq.(4) describe the calculations of $\theta_{\alpha x}$ and $\theta_{\alpha y}$ that will be used to control the movement of the camera to the desired position of the sight, hence producing a natural way to move the robot arm.

Moreover, the *arctangent* form of Eq.(3) and Eq.(4) produces a smoothing of commands which will reduce the high amplitude sight motion and will bond the maximal motion amplitude to the field of view of the camera. The user is then assured of controlling the camera motion easily.

3.2 Adaptive speed control

The speed control algorithm for the Youbot's base unit is explained in Algorithm 1, and is further explained and visualised in Fig. 7(a). It should be noted that only the right hand is considered in the context of adaptive base speed control. When the hand is pointing forward, and eyes are looking at the forward direction at the same time, the robot will also go forward in full speed. In Fig. 7(b), when the operator is looking and pointing to the same side, the speed of Youbot's base is reduced. The movement of the robot's manipulator also slows down until the direction of gaze and hand motion are pointing forward both again. The robot gradually slows down to stationary when the user is looking and pointing in two different sides, as shown in Fig. 7(c).

4 Experiments

From a user point of view, it is intuitive for an operator to manipulate the robot using the proposed teleoperation system. Once the calibrations phases are completed with the Smart Eye software, users, either skilled or unskilled, are expected to be able to control the robot within negligible learning time.

The gaze tracking device, however, introduces some noticeable noises with its measurements. To understand the noise characteristics of gaze tracking, three tests were carried out with one user performing three tasks, namely eye moving from centre to the right horizontally, eye moving from centre to the left,

Algorithm 1 Speed control algorithm of a mobile robotic arm.

Input: D_e : direction of user's gaze; D_h : direction of user's hand; k_a and k_b : control gain of robot arm speed and base speed;
Output: optimal movement speed of robot's arm V_a ; optimal movement speed of robot's base V_b ;
1: initial $k_a = k_b = 1$;
2: **repeat**
3: compute difference of directions $D' = D_e - D_h$;
4: **if** $D' \leq \frac{\pi}{2}$ **then**
5: let base follow arm $k_a = 0.5$, $k_b = 0.7$;
6: **end if**
7: **if** $D' > \frac{\pi}{2}$ **then**
8: let base wait arm $k_a = 0.7$, $k_b = 0$;
9: **end if**
10: **until** $D_e = D_h = Forward$
11: set back $k_a = k_b = 1$.

and staring at one fixed point. As shown in Fig. 6, the left column shows the trajectories of eye movements in three tests. The right column shows the statistical distributions respectively. The first two show the distributions of robot orientation motions in the vertical axis (θ_{ay} in radian), calculated from the raw measurement in the image pixel space using Eq. (4), and the third one shows the distribution of robot motions in both directions (θ_{ax} in Eq.(3), and θ_{ay} in Eq.(4)). The figures shown are the results after applying a simple moving average (MA) filter with 20 samples for each window. Thanks to the averaging effects from both the MA filter and also the latency introduced by the robot controllers, the robot motions are not as jerky as the raw measurement and look satisfactorily smooth.

The TCP client/server between the subsystems works correctly, and no problems were noticed during the user commands transmissions. Images are transmitted with a latency of approximately 200ms, which was small enough to control the arm correctly. During the tests, a user could control the robot with his/her hand gestures and gaze while monitoring the reactions of the arm in real time. The auto-correction system would step in automatically to help the user gain better hand-eye cooperation during teleoperation.

The method was implemented and tested with nine untrained operators. The task was to navigate the robot around the laboratory. All participants were given a brief introduction of how the gaze and hand motion control interface works. Each participant was asked to complete a survey for evaluating the system performance. The survey used is a System Usability Scale (SUS) survey designed by [2]. Users were asked to evaluate the system based on 10 statements, to rate them from 1 to 5 (will be scaled to 0 to 4 for calculating the final score).

Fig. 8 presents the average scores of the survey results. 10 questions were grouped into 5 sets to demonstrate different characteristics of the system [5]. And the final SUS score was calculated as an average of 69.8, out of 100. Taking

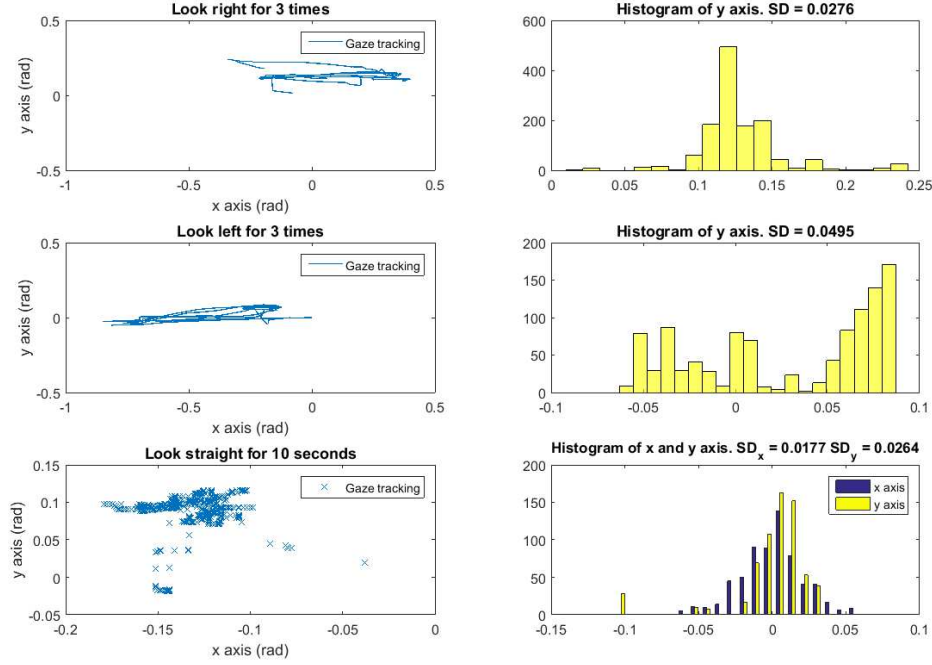
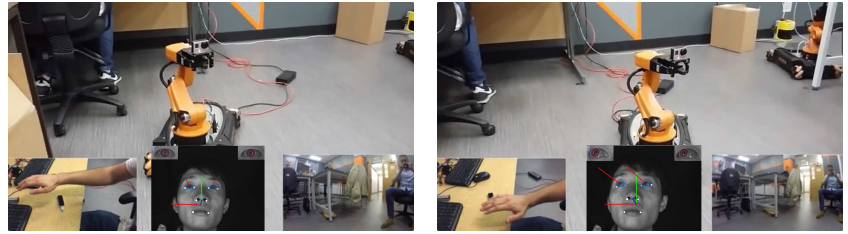


Fig. 6. Gaze tracking input data with their corresponding standard deviations (SD) in three different scenarios.

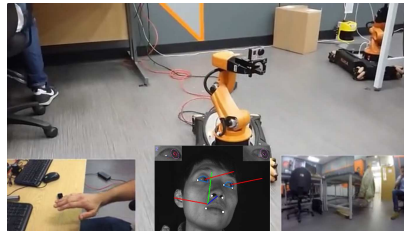
both the average scores from Fig. 8 and the final SUS survey score into account, it can be concluded that the proposed teleoperation method has been proved to be usable and acceptable to general users. However, more user trials will be needed with more complex scenarios.

5 Conclusion

This paper presents a combination method of using gaze tracking and hand motion detection to teleoperate a mobile manipulator. A camera was attached to the robot manipulator's end-effector to provide the user with visual feedback. To allow the user to control the robot in a more natural way, gaze tracking is used to orientate the camera, providing an intuitive method of interaction with the robot. The right hand motion tracking is used for moving the robot base on the ground and the left hand motion controls the translation of the manipulator end-effector. This combination method offers the operator to view the robot's workspace from a first-person perspective to operate the robot intuitively. Also, the operator can manipulate the orientations of the robot's end-effector through intuitive head motions. Hand-eye coordination was enhanced during multi-sensor teleoperation with the applied algorithm. The proposed system was validated with tests on a KUKA Youbot and demonstrated the improved performance of



(a) Teleoperating while user pointing forward and looking forward. (b) Teleoperating while user pointing and looking in same directions.



(c) Teleoperating while user pointing and looking in different directions.

Fig. 7. The illustration of auto-correction teleoperation method. Photos captured from the video of tests.

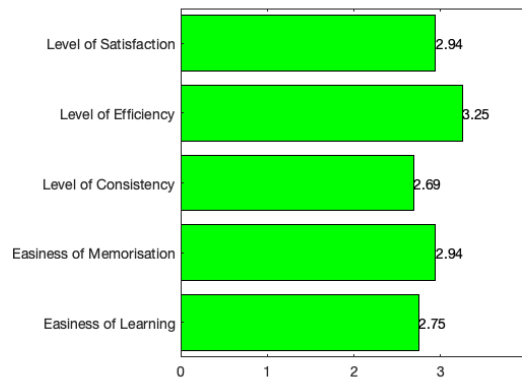


Fig. 8. Average scores of SUS survey results. 10 questions was grouped into 5 different characteristics.

the designed auto-correction teleoperation method. User experience survey was carried out to prove the feasibility of the system. Future works can be developed based on the survey to further improve the robot teleoperation experience for unskilled operators.

References

1. Bischoff, R., Huggenberger, U., Prassler, E.: Kuka youbot-a mobile manipulator for research and education. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on. pp. 1–4. IEEE (2011) [6](#)
2. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996) [9](#)
3. Du, G., Zhang, P., Liu, X.: Markerless human-manipulator interface using leap motion with interval kalman filter and improved particle filter. IEEE Transactions on Industrial Informatics **12**(2), 694–704 (April 2016). <https://doi.org/10.1109/TII.2016.2526674> [2](#)
4. Dziemian, S., Abbott, W.W., Faisal, A.A.: Gaze-based teleprosthetic enables intuitive continuous control of complex robot arm use: Writing and drawing. In: 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob). pp. 1277–1282 (June 2016). <https://doi.org/10.1109/BIOROB.2016.7523807> [2](#)
5. Ggo, D., Carreto, C., Figueiredo, L.: Teleoperation of a mobile robot based on eye-gaze tracking. In: 2017 12th Iberian Conference on Information Systems and Technologies (CISTI). pp. 1–6 (June 2017). <https://doi.org/10.23919/CISTI.2017.7975673> [2](#), [9](#)
6. Li, X., Choi, T., Chun, H., Gim, S., Lee, S., Kang, S., Kim, K.: Active cannula robot with misorientation auto-recovery camera: A method to improve hand-eye coordination in minimally invasive surgery. In: 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013). pp. 276–280 (Oct 2013). <https://doi.org/10.1109/ICCAS.2013.6703908> [2](#)
7. Peppoloni, L., Brizzi, F., Avizzano, C.A., Ruffaldi, E.: Immersive ros-integrated framework for robot teleoperation. In: 2015 IEEE Symposium on 3D User Interfaces (3DUI). pp. 177–178 (March 2015). <https://doi.org/10.1109/3DUI.2015.7131758> [2](#)
8. Qiu, R., Ji, Z., Noyvirt, A., Soroka, A., Setchi, R., Pham, D.T., Xu, S., Shivarov, N., Pigni, L., Arbeiter, G., Weisshardt, F., Graf, B., Mast, M., Blasi, L., Facal, D., Rooker, M., Lopez, R., Li, D., Liu, B., Kronreif, G., Smrz, P.: Towards robust personal assistant robots: Experience gained in the srs project. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 1651–1657 (Oct 2012). <https://doi.org/10.1109/IROS.2012.6385727> [2](#)
9. Tunstel, E., Maimone, M., Trebi-Ollennu, A., Yen, J., Petras, R., Willson, R.: Mars exploration rover mobility and robotic arm operational performance. In: Systems, Man and Cybernetics, 2005 IEEE International Conference on. vol. 2, pp. 1807–1814. IEEE (2005) [2](#)
10. Xiang, Z., Su, J., Ma, Z.: Uncalibrated hand-eye coordination based hri on humanoid robot. In: Proceedings of the 33rd Chinese Control Conference. pp. 8438–8443 (July 2014). <https://doi.org/10.1109/ChiCC.2014.6896416> [2](#)
11. Yu, M., Wang, X., Lin, Y., Bai, X.: Gaze tracking system for teleoperation. In: The 26th Chinese Control and Decision Conference (2014 CCDC). pp. 4617–4622 (May 2014). <https://doi.org/10.1109/CCDC.2014.6852997> [2](#)