ASCE

Effects of Spatial Characteristics on the Human–Robot Communication Using Deictic Gesture in Construction

Sungboo Yoon, S.M.ASCE[1]; YeSeul Kim, S.M.ASCE[2]; Moonseo Park, M.ASCE[3]; and Changbum R. Ahn, A.M.ASCE[4]

Abstract: Construction robots are expected to frequently communicate in situ improvisations with human workers to adapt and change their workflow and methods. One way to achieve this is through deictic gestures that are one of the most effective forms of human–robot interaction (HRI) in delivering spatial information. Nevertheless, the limited coverage of deictic gestures in large-scale environments poses some challenges for both humans and robots in leveraging such techniques for HRI in construction. To identify the feasibility of deictic gestures in the construction domain and find applicable solutions for improving performance, this study aims to extend current knowledge on the performance in communicating positional information using deictic gestures by investigating the effects of spatial characteristics on spatial referencing, focusing on the target configuration, target distance, and relative position of human and robot. We observed that the recognition and estimation of deictic gestures were affected by the target plane, target position, and the target layout and that the robot performance was significantly reduced as the distance between the human and robot increased. The findings of this study demonstrate the challenges in spatial referencing within a large-scale environment and highlight the need for bidirectional communication in HRI. **DOI: 10.1061/JCEMD4.COENG-12997.**

Author keywords: Construction robotics; Human–robot interaction (HRI); Deictic gestures; Spatial communication.

Introduction

Robotic technologies are envisioned as a promising alternative for the construction industry, which constantly suffers from stagnant productivity and a shortage of skilled workers (Gharbia et al. 2020). Recent advances in artificial intelligence (AI) and AI-based perceptual and manipulative abilities in robotics have led to an unprecedented increase in robots' performance. As a result, robots have been introduced in many onsite construction tasks that benefit from their ability to detect minor deviations, handle heavy and hazardous building elements, and precisely repeat defined paths (Hentout et al. 2019; Wang et al. 2021). Examples include semi-automated bricklaying robots (FBR 2021), rebar-tying robots (Construction Robotics 2021), site layout robots (Civ Robotics 2021; Dusty Robotics 2021), and three-dimensional (3D) printing robots (COBOD 2021; MX3D 2021).

However, the employment of such robots in real construction job sites is still limited due to the unstructured and dynamic nature of construction environments (Carra et al. 2018; Feng et al. 2015; Wang et al. 2021). When deployed onsite, robots are often faced with site congestion and a multitude of interactions among workers, materials, and equipment, resulting in degraded performance or even robot failures (Park and Cho 2017). In these situations, robots must frequently adapt and change their workflows and methods (Follini et al. 2021). This underlines the importance of the in situ improvisations of humans since compared to robots, humans are more competent in making adaptive decisions based on perceptual understanding and previous work experiences (Wang et al. 2021). In this context, there is a growing need for direct and effective communication of in situ improvisations between human workers and robots. Nevertheless, state-of-the-art human–robot interaction (HRI) methods in construction (e.g., teleoperation using joysticks) are inefficient in terms of exchanging construction workers' improvisations (Kyjanek et al. 2019; Roldán et al. 2019; Wang et al. 2021).

During the last few decades, natural communication has been a central issue in human–robot interactive technology (Li 2020; Tölgyessy et al. 2017). Inspired by human-human interaction (HHI) methods, many previous studies implemented two main techniques for HRI: speech and gesture. Using speech based on natural language understanding is the most convenient, yet in construction environments, speech may be interfered with by the noise of construction equipment and activities. In such environments, the use of deictic gestures is an ideal way around this problem. Deictic gestures are the most important form of gesture in task-based contexts owing to their key role in expressing semantic information about the shared environment (Alibali 2005; Williams et al. 2019; Yongda et al. 2018). They are also highly operable, nonintrusive, and intuitive since they do not require additional devices such as control pads or wearable sensors (Li 2020).

Nevertheless, in construction job sites, there are still many challenges in using deictic gestures for HRI. First, construction operations often involve distal, adjacent, and noncoplanar objects. For example, PVC panels, widely used as a cladding material, are

[1]Ph.D. Student, Dept. of Architecture and Architectural Engineering, Seoul National Univ., Seoul 08826, Republic of Korea. ORCID: https://orcid.org/0000-0003-4997-5792. Email: yoonsb24@snu.ac.kr

[2]Ph.D. Student, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332. ORCID: https://orcid.org/0000-0001-7953-4176. Email: ykim858@gatech.edu

[3]Professor, Dept. of Architecture and Architectural Engineering, Seoul National Univ., Seoul 08826, Republic of Korea. Email: mspark@snu.ac.kr

[4]Associate Professor, Dept. of Architecture and Architectural Engineering, Institute of Construction and Environmental Engineering, Seoul National Univ., Seoul 08826, Republic of Korea (corresponding author). ORCID: https://orcid.org/0000-0002-6733-2216. Email: cbahn@snu.ac.kr

Note. This manuscript was submitted on August 5, 2022; approved on March 2, 2023; published online on April 28, 2023. Discussion period open until September 28, 2023; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364.

installed into grid openings in a ceiling frame or on walls or existing panels with a minimum gap between the adjoining panels (Liang et al. 2020). Such distal, adjacent, and noncoplanar targets may pose a significant effect on the recognition and estimation of deictic gestures because deictic gestures often suffer from limited coverage. Moreover, in shared workspaces on construction job sites, workers and robots form various human–robot relationships, resulting in constant changes in the relative positions between them. However, the reliability and readability of the exchanged spatial information are significantly affected by the relative position of the human (Gustavsson et al. 2018; Mayer et al. 2020). These challenges are critical because such spatial information, including geometric entities of a building component [domains, surfaces, edges, points on a surface, and bounding spaces (e.g., zone)] or geometric relationships between them (Kim and Cho 2015), is one of the most fundamental factors for in situ improvisations. For instance, when a worker intends to provisionally install a drywall panel for the ceiling, they should be aware of the information on possible ceiling joints to drive the fasteners in based on their spatial cognitive abilities. Meanwhile, these challenges may explain why, despite the advantages of deictic gestures, the application of the developed human–robot interface has been limited to specific settings that fundamentally deviate from construction environments, such as close-proximity objects located on tables (Jevtić et al. 2019; Jirak et al. 2021; Weerakoon et al. 2020; Weng et al. 2019; Whitney et al. 2017) or head-up displays (HUDs) (Brand et al. 2016).

To this end, this study aims to investigate to what extent and how spatial referencing using deictic gestures to communicate positional information about the construction objects is influenced by spatial characteristics, namely, target positions (with different target configurations) and human positions relative to the robot. In this study, the latest deictic gesture-based HRI method was adopted that utilizes the vision-based human pose estimation technique to estimate pointing directions and positions. We developed and performed spatial communication tasks for panel installation work, in the contexts of HRI and HHI. Finally, we evaluated the robot's performance in spatial communication and compared it to human performance. Based on the findings, we discuss the envisioned applications of deictic gestures to construction operations and implications for human–robot communication techniques in construction. This study contributes to the body of knowledge that addresses the need for spatial communication between the human worker and robot for onsite robot deployment.

Background

Human–Robot Interaction in Construction

HRI can be defined as the exchange of information and actions between a human and a robot to perform a given task by means of a user interface [ISO 8373:2021 (ISO 2021)]. The goal of HRI is to enable synergistic teams of humans and robots in which team members perform tasks according to their abilities (Burke et al. 2004). HRI is integrated into collaborative robots for various industrial applications, such as assembling (Feng et al. 2015; Grahn et al. 2018; Heydaryan et al. 2018; Makris et al. 2016), machine tending (Annem et al. 2019), packaging (Zwicker and Reinhart 2014), and palletizing (Ganglbauer et al. 2020; Lamon et al. 2020). According to Malik and Bilberg (2019), human–robot relationships in HRI can be classified into four categories: (1) coexistence—the human and robot work alongside each other but do not share a workspace, (2) synchronized—the human and robot share a workspace but only one of them is present at any one moment, (3) cooperation—the human and robot share a workspace and are present at the same time but work independently, and (4) collaboration—the human and robot share a workspace, are present at the same time, and work simultaneously.

HRI is an emerging research field in construction because it is one of the keys to the successful implementation of construction robotics (Adami et al. 2022). A large number of high-performance robots are in development or are being deployed on construction job sites. Construction workers have serious concerns about technological unemployment and worry that their jobs might be replaced by robots or that they would not be able to adjust to automation (Kim et al. 2022). Implementing an appropriate level of HRI could be one of the viable solutions to the concerns of construction workers (Kim et al. 2022). However, HRI methods in construction are still under development and are inefficient in terms of exchanging construction workers' improvisations. They usually suffer from accuracy reduction and time delays (Roldán et al. 2019; Wang et al. 2021) and require humans to continuously perform manual tasks during the whole work process (Wang et al. 2021).

Preprogramming has the lowest level of robot autonomy yet is the most common HRI method in construction that involves programming a robot with a predefined sequence of activities (Liang et al. 2021b). Preprogramming accounts for most of the HRI methods in the manufacturing industry, environments in which the robots perform the same tasks repetitively with minimum human intervention (Inkulu et al. 2022). In the construction industry, preprogramming is widely employed in operations such as welding (Tavares et al. 2019) or assembly (Ding et al. 2020; Feng et al. 2015), often supported by building information modeling (BIM). 3D BIM models are manually created by human labor or reconstructed through 3D vision techniques and tasks are allocated separately for human operators and robots based on BIM models. Despite the high accuracy of this method, it faces some challenges in expressing real-time onsite information due to the dynamic and unstructured nature of a construction site (Ding et al. 2020), making this method time-consuming and burdensome since the user must manually update BIM models during operation to handle in situ variations.

Teleoperation has been suggested as an alternative to BIM-based preprogramming due to its capability in handling unexpected situations (Liang et al. 2021b) and therefore has been deployed in dynamic and unsafe construction operations such as excavation (Okishiba et al. 2019) and maintenance (David et al. 2014; Koh et al. 2021). Teleoperation involves real-time control of a robot's motion by a human operator from a remote site through a communication channel [ISO 8373:2021 (ISO 2021)] (Zhou et al. 2020). Yet, the difficulties for novice users in the operation of the robots utilizing the traditional communication technologies of teleoperation (i.e., joysticks and control pads) led to an increasing number of investigations on the intuitive and user-friendly human–robot communication technologies (Chen et al. 2022; David et al. 2014; Roldán et al. 2019).

Deictic Gesture-Based Human–Robot Interaction

Deictic gestures, often referred to as pointing gestures, are a form of gesture commonly performed by extending the arm and index finger (Mayer et al. 2020). Deictic gestures are one of the most fundamental techniques of HHI (Oosterwijk et al. 2017). Humans learn to use deictic gestures from infancy and continue to rely on them as a core means of communication (Butterworth 2003; Williams et al. 2019). In particular, deictic gestures are an effective nonverbal

communication technique for establishing joint attention based on mutual understanding of an environment in which people find it difficult or impossible to communicate through verbal descriptions, e.g., noisy factory environments (Williams et al. 2019). Currently, deictic gestures are considered one of the most acceptable techniques to support the exchange of spatial information in complex environments and are used for such things as describing the shape of an object, giving a location in space, or giving directions (Alibali 2005).

Deictic gestures have been widely accepted for HRI or human-computer interaction (HCI), beginning with the preliminary work of Bolt (1980), in which deictic gestures were used as an input for the HCI to move virtual objects on a screen (also known as *Put-That-There*). Recent applications of deictic gestures in HRI include indicating target points (Dhingra et al. 2020; Lai et al. 2016), labeling information on areas (Zamani et al. 2018), selecting objects for object-fetching or attention-directing tasks (Canal et al. 2016; Sauppé and Mutlu 2014; Whitney et al. 2017), and setting a destination for robot navigation (Gromov et al. 2020; Medeiros et al. 2021; Tölgyessy et al. 2017).

Nevertheless, a major challenge of using deictic gestures for HRI is their limited performance. Jevtić et al. (2019) compared two different interaction modalities for robot-assisted dressing: speech and gesture command. The experimental results confirmed that the pointing gesture command has significantly worse performance compared with speech input in terms of the user workload (i.e., physical demand, performance, and effort) and robot performance (i.e., number of corrections). Also, Jevtić et al. (2015) compared three different interaction modalities for mobile robot guidance in an indoor environment: direct physical interaction, person following, and pointing control. The experimental results confirmed that pointing control has significantly worst performance among the three interaction modalities in terms of user workload, task completion time, and accuracy, whereas direct physical interaction showed the highest performance. Mayer et al. (2015) showed that even with the motion capture system that estimates the absolute position of markers attached to users, the interface had limited accuracy in selecting target points on a wall display. These investigations show the general performance of the deictic gesture as an input modality for HRI systems, yet performance in a large-scale construction environment has not been studied in depth. Moreover, to expand the opportunities for utilizing deictic gestures for human–robot spatial communication in construction, it must be verified that the robot performance would reach human performance in identical conditions.

Deictic Gesture Recognition

Deictic gesture-based HRI relies on accurate and robust deictic gesture recognition. Deictic gesture recognition techniques can be divided into two main categories: sensor-based approach and vision-based approach. The sensor-based approach measures electrical muscle stimulation (EMS) with an electromyography (EMG) (Ameri et al. 2018; Navas Medrano et al. 2020), the specific force, angular rate, and orientation of the arm with inertial measurement units (IMUs) (Gromov et al. 2018, 2019, 2020; Walkowski et al. 2011), or motion data with data gloves (Kumar et al. 2012) and reconstructs 3D rays in the user's own local reference frame using the motion signals (Gromov et al. 2018). The sensor-based approach is usually an accurate, sensitive, and reliable method owing to its direct acquisition of a human deictic posture. However, the sensor-based approach may not be considered a realistic method for onsite applications as it relies on the existence of a wearable data acquisition device, resulting in a lack of convenience and user-friendliness (Li et al. 2019; Medeiros et al. 2021).

The vision-based approach collects images with one or more cameras, e.g., monocular cameras (Nickel and Stiefelhagen 2003), stereo cameras (Keskin et al. 2003), or RGB-D cameras (Dhingra et al. 2020), and reconstructs 3D rays from the estimated 3D human pose in the surrounding space using the user actions or state information. The vision-based approach does not usually require the wearing of external devices and has the benefit of being natural, nonintrusive, intuitive, and highly operable. Nevertheless, the vision-based approach has unavoidable problems, such as image noise mainly caused by illumination and occlusion by other objects or users (Li 2020).

The deictic gesture recognition technique used in this study belongs to the category of vision-based approach using an RGB-D camera for human pose estimation. RGB-D cameras are widely used for data collection in HRI studies due to their affordability and sufficient information retrieved from 3D sensing of a robot's surrounding environment (Tashtoush et al. 2021). Using RGB-D cameras for deictic gesture recognition provides numerous advantages, including higher robustness to variations in lighting conditions and accuracy in human contour extraction compared with RGB cameras (Li 2020).

Methodology

To evaluate the impact of spatial characteristics on spatial referencing using deictic gestures during HRI in construction, we developed a spatial communication task that involved one or more addressers (human pointers), addressees (human and robot observers), and referents (panels) (see Fig. 1). HRI in construction can involve the communication of many types of information in both human-to-robot and robot-to-human directions (Weng et al. 2019). This study is focused on the human-to-robot communication of positional information on a target panel for highlighting where in the workspace (wall and ceiling) the robot should execute its panel installation work. This study aims to evaluate both robot performance in spatial communication in HRI contexts and human performance in HHI contexts and compare the results. Therefore, we had a human observer for every robot's observation position while maintaining the other experimental conditions.

The experimental design is focused on how the recognition and estimation of deictic gestures were affected by the target configuration, target distance, and relative position of humans and robots. Information about these factors is often involved in reasoning about target locations within a large-scale environment. Referencing an object, for example, requires reasoning about its distance and direction, which can be determined relative to a human or robot moving through the surrounding environment, as well as its spatial features (Vasilyeva and Lourenco 2012). These spatial characteristics are specific to construction, compared to the scope of prior studies in related fields such as robotics and computer science. To account for distal, adjacent, and noncoplanar objects frequently encountered in construction, the present study explored target positions with different target configurations. Furthermore, given the dynamic nature of construction environments, the positions of humans relative to robots were manipulated.

First, we manipulated the target configuration by altering the panel layouts and sizes to examine to what level of pointing difficulty the interface showed acceptable recognition accuracy. Second, we manipulated the target distance by changing the distance between the pointer and the panels. Last, we manipulated the relative position of the pointer and observer by changing the relative

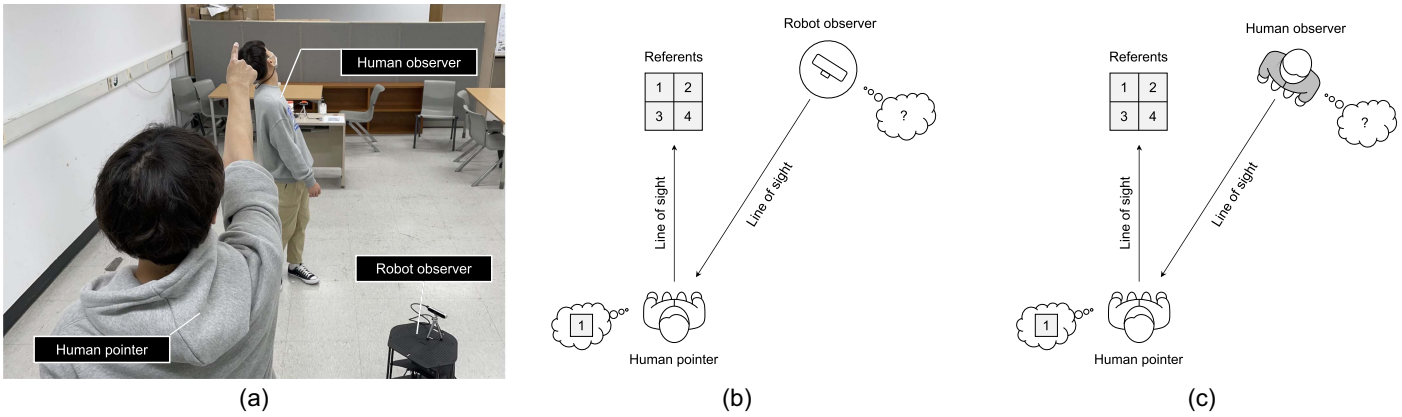

Fig. 1. Spatial communication task: (a) photograph of the task setup; schematic illustration of the task setup in (b) HRI context; and (c) HHI context.

distance and angle of the observer in relation to the pointer. Meanwhile, we manipulated the number of pointers and human observers to disambiguate differences in individuals. The experimental conditions for each experiment are as follows.

Experiment 1 (E1): Multiple pointers performed pointing tasks at a fixed position, and the robot faced the pointer at a fixed position. Large-sized panels were installed on both wall and ceiling in a horizontal grid layout without space between them.

Experiment 2 (E2): A single pointer performed pointing tasks in three different positions. The robot and human observer faced the pointer at a fixed position. Small-sized panels were installed on the ceiling in both horizontal and vertical grid layouts without space between them.

Experiment 3 (E3): A single pointer performed pointing tasks at a fixed position. The robot and multiple human observers faced the pointer in nine different positions. Small-sized panels were installed on the ceiling in both horizontal and vertical grid layouts without space between them.

Hardware Setup

The robot used for the experiments is shown in Fig. 2. The robotic platform is Turtlebot v2 that comprises a mobile base called Kobuki (Yujin Robot, Seoul, Republic of Korea), a front-mounted RGB-D camera, and computing hardware. The Intel RealSense Depth Camera D435 (Intel Corporation, Santa Clara, California) was selected as an RGB-D camera, which can capture RGB and depth images simultaneously. It was connected to Intel Next Unit of Computing (NUC), a small form factor computer, with Intel Core i3-8109U CPU @ 3.60GHz and 32GB RAM. The system was implemented in a robot operating system (ROS).

Experimental Setup

The participants were asked to perform a spatial communication task for each experiment. At the start of each trial, the pointer was visually informed about the panel number. The pointer was asked to reference the target panel with a pointing gesture and hold until the next panel number was shown. No further instructions were given to the pointer (i.e., to refer to the center of the panel). Meanwhile, the observer was asked to infer the correct panel number without informing the pointer to eliminate the influence on further pointing motions of the pointer. When the next panel number was presented to the pointer, the pointer returned to the initial posture and repeated the previous steps. We set the duration of each trial as 5 s.

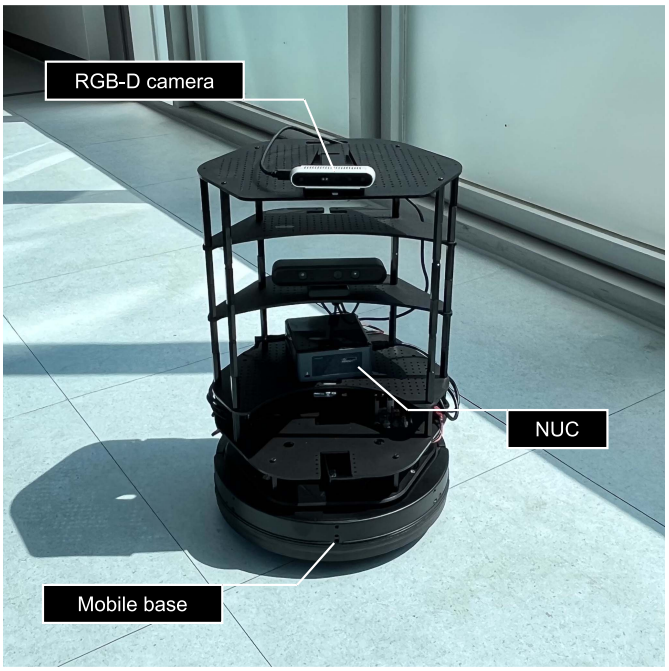

Fig. 2. Robotic platform used in this study.

Three experiments were conducted in a laboratory setup. The experimental setup of Experiment 1 is depicted in Fig. 3(a). Five panels with an equal size of 0.7×0.7 m were installed side by side on both ceiling and wall. The robot was located at Position O1, and the front-mounted RGB-D camera faced the pointer at a height of 0.7 m and roll angle of 90° (the positive z-axis points forward). The pointer stood at Position P1, 3.0 m away from the RGB-D camera. All the intrinsic and extrinsic parameters were identified and used for RGB-D camera calibration. Four right-handed participants (two male and two female, ages 20–29) were asked to perform the spatial communication tasks for Experiment 1. All the participants were assigned the role of pointer. No observers were involved in Experiment 1. Each participant performed 130 trials, grouped into 13 blocks. A single block consisted of 10 trials: 5 ceiling panels (C1–C5) and 5 wall panels (W1–W5).

The experimental setup of Experiment 2 is illustrated in Fig. 3(b). Nine panels with an equal size of 0.6×0.6 m were installed on the ceiling in a 3×3 grid layout. The robot was located at

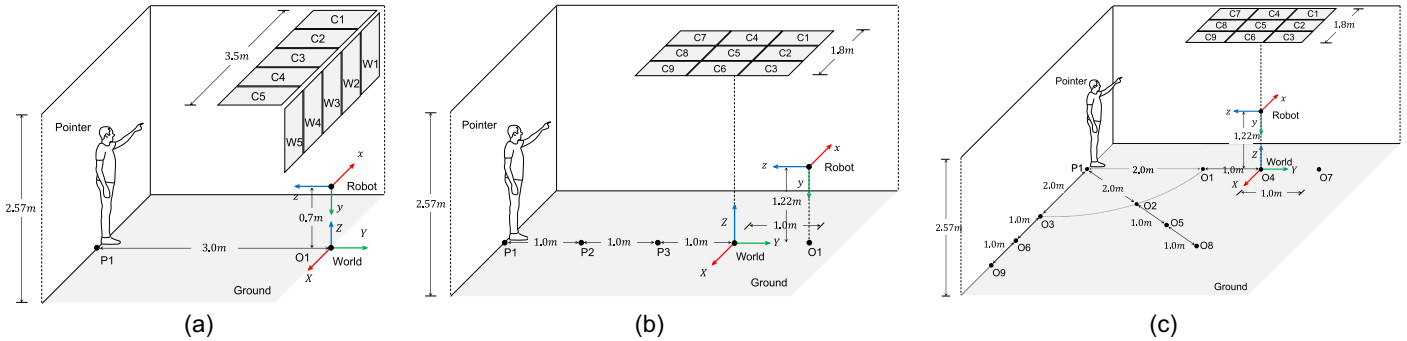

Fig. 3. Experimental environments of (a) Experiment 1; (b) Experiment 2; and (c) Experiment 3.

Position O1, 1.0 m away from the world origin, and the RGB-D camera faced the pointer at a height of 1.22 m and roll angle of 83°. The observer also stood at Position O1. The pointer stood at three different positions (P1–P3), each position with a different distance to the world origin (2–4 m, 1 m steps), which was located at the projected position of the center point of the C5 panel. Two right-handed participants (two males, ages 25–27) were recruited to participate in Experiment 2. One of the participants was assigned the role of pointer, while the other was assigned the role of observer. For each pointing position, the pointer performed 90 trials, grouped into 10 blocks. A single block consisted of nine trials: nine ceiling panels (C1–C9).

The experimental setup of Experiment 3 is similar to the experimental setup of Experiment 2 [see Fig. 3(c)]. However, in Experiment 3, the pointer stood at a fixed Position P1, while the positions of the observers and robot varied from O1 to O9, each position with a different distance (2–4 m, 1 m steps) and angle (0°–90°, 45° steps) to the pointer. Eleven right-handed participants (seven male and four female, ages 24–32) were recruited to participate in Experiment 3. One of the participants was assigned the role of pointer, and the others were assigned the role of observer. For each observation position, the pointer performed 45 trials, grouped into 10 blocks. A single block consisted of nine trials: nine ceiling panels (C1–C9).

Gesture Recognition and Pointing Target Estimation

Fig. 4 shows the deictic gesture-based HRI method adopted in this study, which utilizes human pose estimation with deep learning. In this method, the 3D human skeletal data extracted from the RGB and depth images are used to detect deictic gestures. This study employs OpenPose (Cao et al. 2021), a real-time multiperson two-dimensional (2D) pose estimation library, to estimate the 2D skeletal data. The BODY-25 model from the OpenPose library detects 25 human body joints from each RGB image frame. The pixel coordinates (x, y) from the 2D image are transformed into corresponding world coordinates (X, Y, Z) in 3D space by inverse perspective projection (Kim et al. 2015) [Eq. (1)]:

$$[X, Y, Z, 1]^T = \boldsymbol{P}^{-1}[x, y, 1]^T \tag{1}$$

where $\boldsymbol{P} = 3 \times 4$ camera projection matrix. The 3D human pose obtained from the projection transformation is then utilized for estimating the pointing direction. In the estimation of the pointing direction, three 3D coordinates of the body joints are used: (1) shoulder P_s, (2) elbow P_e, and (3) wrist P_w. We used wrist position instead of fingers to enhance the computation efficiency, considering the onsite applications (Sprute et al. 2018). Given the position of the three body joints, the elbow joint angle θ is defined as Eq. (2)

$$\cos\theta = \frac{\mathbf{v}_{se}\mathbf{v}_{sw}}{|\mathbf{v}_{se}||\mathbf{v}_{sw}|} \tag{2}$$

where $\mathbf{v}_{se}$ = vector from the shoulder to the elbow joint; and $\mathbf{v}_{sw}$ = vector from the shoulder to the wrist joint. If θ is below a predefined angle, the system assumes that the person is stretching their arm for pointing. The pointing direction is defined by a straight line starting from the shoulder to the wrist joint [Eq. (3)]:

$$\mathbf{p} = P_s + d(P_w - P_s) \tag{3}$$

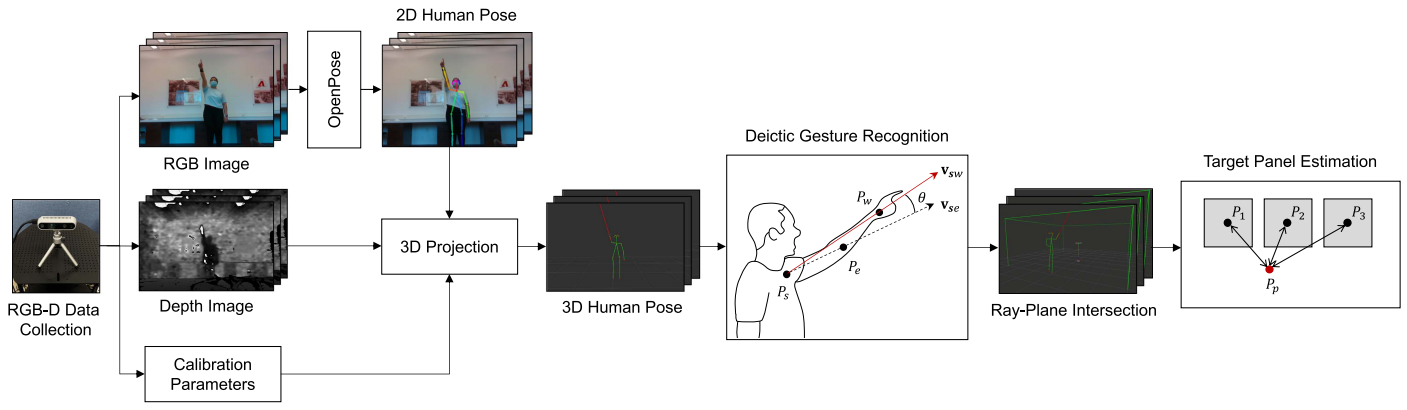

Fig. 4. Target panel estimation process of the current HRI method using deictic gestures.

where $d \in \mathbb{R}$. The pointing position P_p on a ceiling or wall is calculated by the ray-plane intersection using Eq. (4):

$$d = \frac{(P_0 - P_s) \cdot \mathbf{n}}{(P_w - P_s) \cdot \mathbf{n}} \tag{4}$$

where P_0 = arbitrary point on the plane; and $\mathbf{n}$ = normal vector of the plane. The target panel is then predicted using the pointing position. Let P_i be the center point of the target panel index $i \in \mathbb{N}$. A panel with the closest Euclidean distance from the center point is selected as a target panel i_t, which is defined as Eq. (5)

$$i_t = \underset{i}{\operatorname{argmin}}(|P_p - P_i|) \tag{5}$$

This study adopts two techniques proposed from the state-of-the-art HRI, commonly used for enhancing performance. First, only stabilized pointing motions were used for evaluation (Dhingra et al. 2020). Those motions were extracted by excluding the first and fourth quartiles of the candidate frames. Second, the pointing calibration was performed before the main experiments. The estimated pointing positions may differ significantly from what the pointer intended because the pointing movements and their trajectories may vary with individuals (Jevtić et al. 2019; Navas Medrano et al. 2020). Therefore, the pointing calibration locations were calibrated to compensate for the estimation error. In the calibration phase, the participants were asked to point at every corner panel once (C1, C4, C7, and C9). For each corner panel, the mean x- and y-coordinates of the pointing position are calculated using all estimated positions for a single pointing task. The parameters of the linear fitting function are then computed using the calculated mean 2D coordinates and the ground truth 2D coordinates. The linear regression model was applied for the collected data points for evaluation.

Evaluation

We used two common metrics for HRI to evaluate the performance of deictic gestures for spatial communication (Steinfeld et al. 2006). To measure recognition accuracy, we use the F1-score, which is calculated by Eqs. (6)–(8)

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

For each pointing task, the true positive (TP) is the case in which the estimated position is classified as a correct target panel; the false negative (FN) is the case in which the estimated position is classified as other target panels; the false positive (FP) is the case in which the subject is pointing at other target panels; and the true negative (TN) is the case in which the subject is pointing at other target panels but classified correctly.

To measure the point estimation accuracy, we used the deviation from the target, which is defined as the Euclidean distance between the estimated pointing position P_p and the center point of the target panel P_t [Eq. (9)]. We assumed that when asked to refer to a target panel, the participants would tend to point at its center

$$\epsilon = |P_p - P_t| \tag{9}$$

Results

A total of 1,195 pointing trials (520 trials in Experiment 1, 270 trials in Experiment 2, and 405 trials in Experiment 3) were evaluated in an offline setting. In the analysis for each experiment, trials were averaged for each experimental condition and the mean values were put into a statistical analysis.

Experiment 1

There were several differences in the performance depending on the locations of the panels. First, the mean deviation from the targets of the ceiling panels [C1–C5; $M = 1.125$, $SD = 0.263$; Fig. 5(a)] was significantly higher than the wall panels [W1–W5; $M = 0.410$, $SD = 0.174$; Fig. 5(b)], $t(80.502) = 1.702$, $p < 0.001$. Meanwhile, the mean F1-score did not significantly differ (ceiling: $M = 0.815$, $SD = 0.256$ versus wall: $M = 0.895$, $SD = 0.187$), $t(4{,}384.3) = 60.276$, $p = 0.093$. Second, the mean deviation from the targets of the side panels (C1/C5 and W1/W5; $M = 0.836$, $SD = 0.241$; darker box plot in Fig. 5) was significantly higher than the panels near the center (C2–C4 and W2–W4; $M = 0.721$, $SD = 0.210$; lighter box plot in Fig. 5), $t(4{,}292.9) = 4.9359$, $p < 0.001$. Again, the mean F1-score did not show a significant difference (side: $M = 0.881$, $SD = 0.199$ versus center: $M = 0.838$, $SD = 0.243$), $t(3) = 2.5046$, $p = 0.087$. Last, the experimental results were not specific to any pointer; the results of the one-way ANOVA did not present a significant effect of the individual differences on both the deviation from the target, $F(1, 38) = 1.302$, $p = 0.261$, $\eta_p^2 = 0.033$ and the F1-score, $F(1, 88) = 0.186$, $p = 0.668$, $\eta_p^2 = 0.002$.

Experiment 2

We observed three main findings from the evaluation results of Experiment 2. First, the robot performance tended to be higher when the pointer was closer to the target; the results of the one-way ANOVA showed a significant effect of the target distance on the mean F1-score (unfilled dots in Fig. 6), $(1, 25) = 42.27$, $p < 0.001$, $\eta_p^2 = 0.628$ and on the deviation from the target, $F(1{,}261) = 106.2$, $p < 0.001$, $\eta_p^2 = 0.289$. Fig. 7 shows the center point of each target panel (*Ground truth*), average estimated positions (*Predicted*), and their deviations from the target for three target distances. The results showed that the mean deviation from the target was lower than 0.3 m, half the width or height of the panel, when the target distance was 1 or 2 m. Thus, the estimated positions mostly lay within the target panels under such conditions. Second, the target distance also had a significant effect on the F1-score of the human observer, $F(1, 25) = 18.75$, $p < 0.001$, $\eta_p^2 = 0.429$, as shown by the filled dots in Fig. 6. Moreover, the results of the paired t-test showed no significant difference between the HRI and HHI contexts on the F1-score, $t(2) = 1.859$, $p = 0.204$; precision, $t(2) = 1.814$, $p = 0.211$; and recall, $t(2) = 2.078$, $p = 0.173$. Last, the human observer was more resistant to the change in target distance compared to the robot; while the distance increased from 1 to 3 m, the mean F1-score of the human observer decreased by 24%, whereas the mean F1-score of the robot decreased by 55% (see Fig. 6).

Experiment 3

We observed three main findings from the evaluation results of Experiment 3. First, the results of the paired t-test showed that the mean F1-scores of all nine observation positions did not significantly differ between the HRI and HHI contexts (human: $M = 0.567$, $SD = 0.097$ versus robot: $M = 0.569$, $SD = 0.251$),

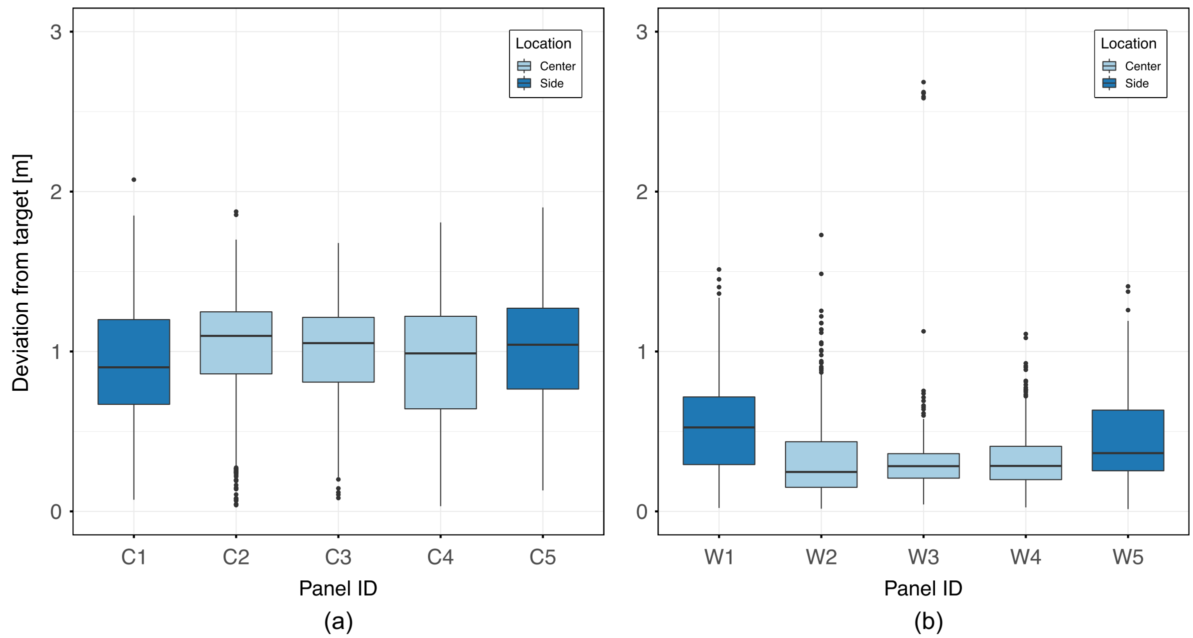

Fig. 5. Deviation from the target of the (a) ceiling panels (C1–C5); and (b) wall panels (W1–W5).

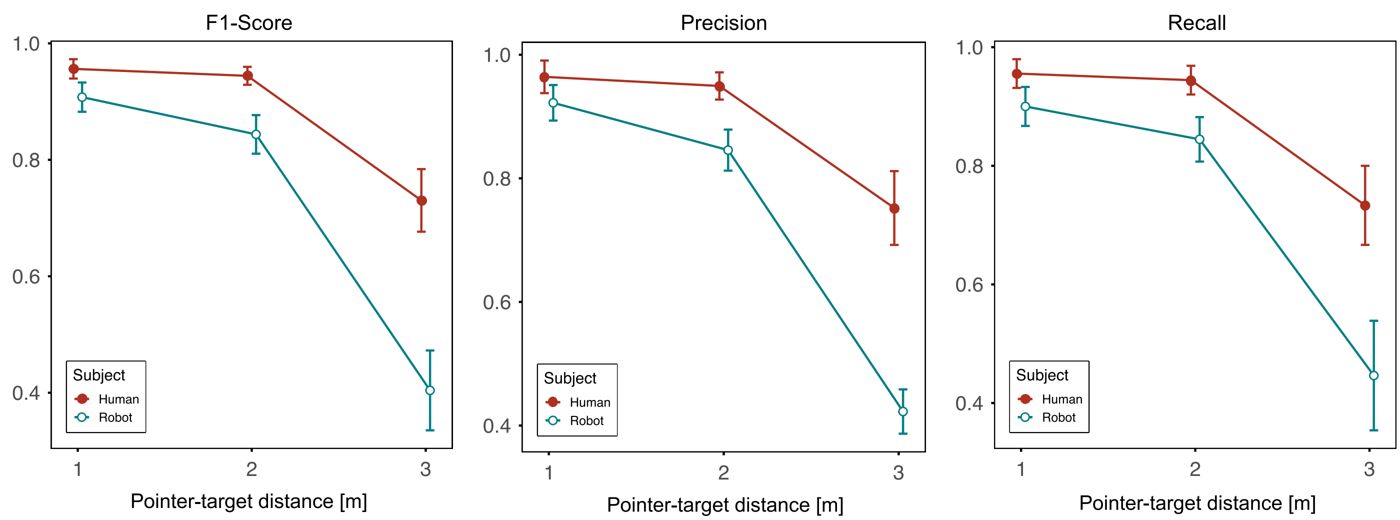

Fig. 6. Recognition accuracy of the human observer and robot (Experiment 2).

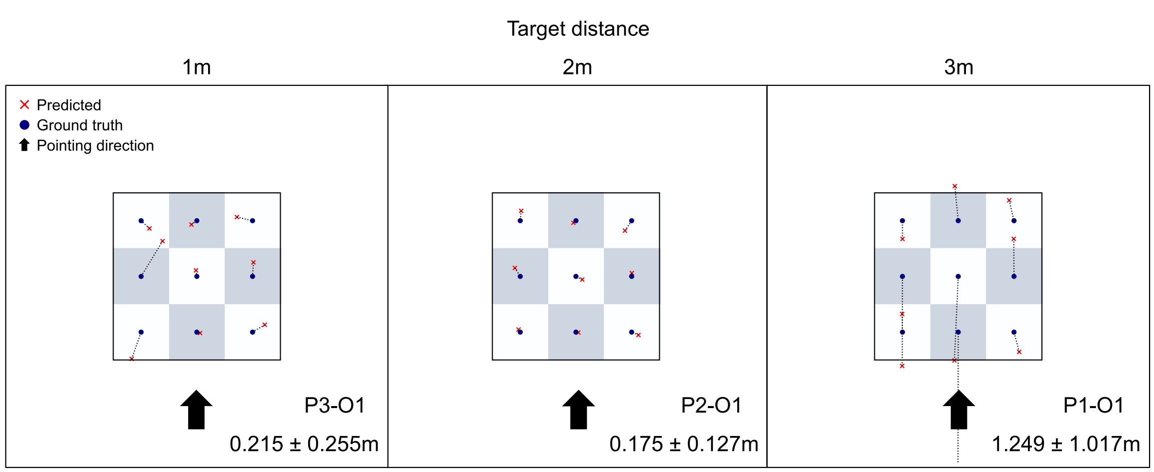

Fig. 7. Estimated positions and their deviations from the target (Experiment 2).

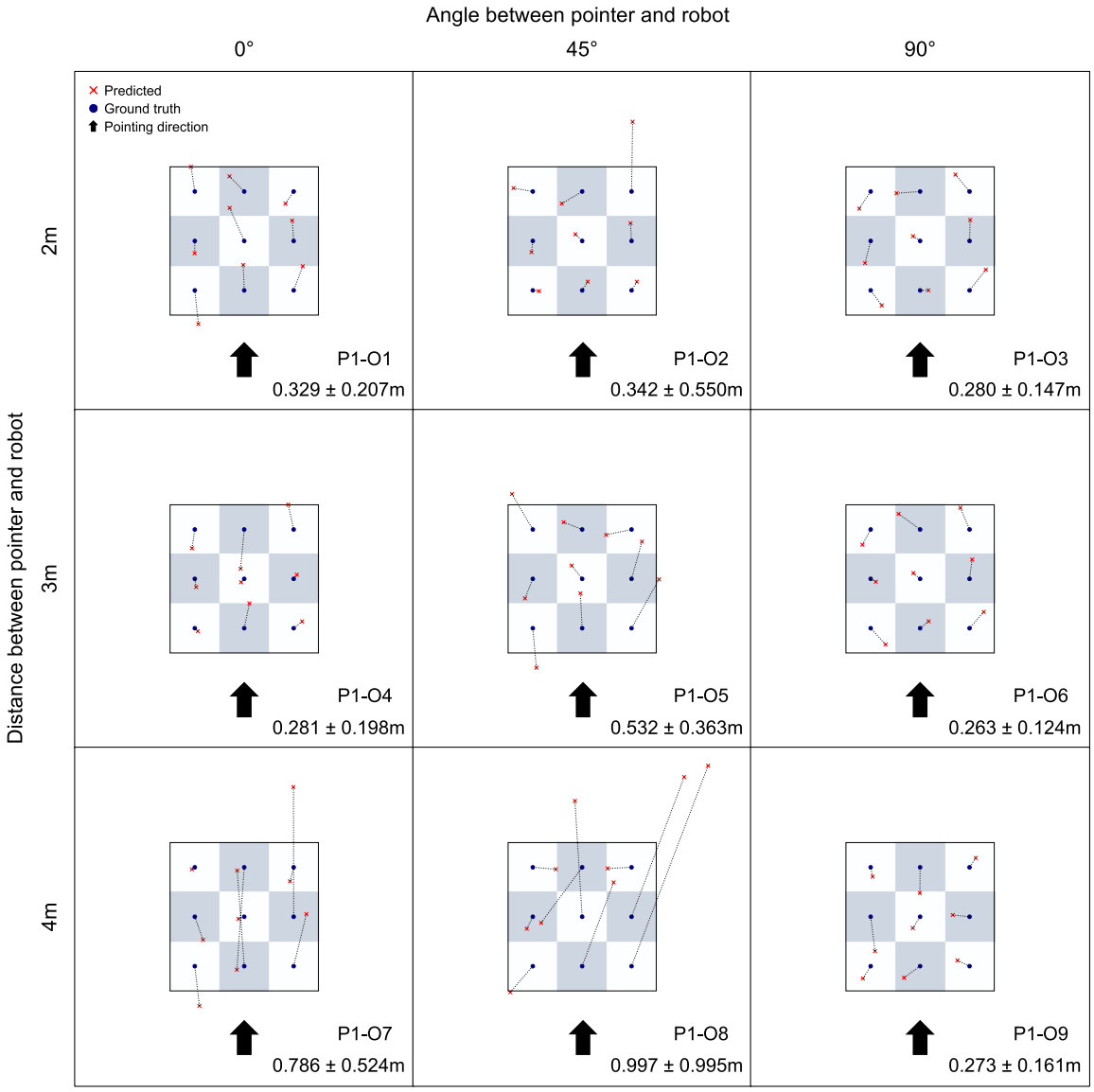

Fig. 8. Estimated positions and their deviations from the target (Experiment 3).

$t(8) = 0.046$, $p = 0.964$. Likewise, the mean precision (human: $M = 0.599$, $SD = 0.034$ versus robot: $M = 0.590$, $SD = 0.191$), $t(8) = 0.135$, $p = 0.896$ and recall (human: $M = 0.562$, $SD = 0.039$ versus robot: $M = 0.591$, $SD = 0.175$), $t(8) = 0.499$, $p = 0.631$ did not show significant differences. Nevertheless, the difference in the standard deviation of the results suggested that humans were more consistent in their performance compared to the robot. Second, the effect of the distance between the pointer and the observer (hereafter referred to as *distance*) and the angle between the pointer and the observer (hereafter referred to as *angle*) on the performance was only significant for the robot. We conducted a two-way ANOVA to verify whether the robot performance was significantly influenced by *distance* × *angle*. The results confirmed a significant effect of *distance* × *angle* on the deviation from the target [*distance*, $F(1,381) = 37.23$, $p < 0.001$, $\eta_p^2 = 0.089$; *angle*, $F(1,381) = 10.53$, $p < 0.01$, $\eta_p^2 = 0.027$; *distance* × *angle* interaction, $F(1,381) = 10.4$, $p < 0.01$, $\eta_p^2 = 0.027$]. Fig. 8 shows the center point of each target panel (*Ground truth*), average estimated positions (*Predicted*), and their deviations from the target for three *distances* and three *angles*. The estimated positions generally lay within the target panels when the *distance* was 2 m, whereas in the observation Position O8, the estimated positions clearly deviate from the target panels, with increased mean deviation from the target up to 0.997 m. The robot's recognition accuracy according to *distance* is shown by the unfilled dots in Fig. 9. It was confirmed that only the *distance* had a significant effect on the F1-score [*distance*, $F(1,77) = 42.444$, $p < 0.001$, $\eta_p^2 = 0.355$; *angle*, $F(1,77) = 2.896$, $p = 0.093$, $\eta_p^2 = 0.036$; *distance* × *angle* interaction, $F(1,77) = 2.387$, $p = 0.126$, $\eta_p^2 = 0.03$]. Meanwhile, the filled dots in Fig. 9 show the recognition accuracy of the human observers. We conducted a two-way repeated-measures ANOVA (RM-ANOVA) to verify whether the human performance was significantly influenced by *distance* × *angle*. The results confirmed no significant effect of *distance* × *angle* on the F1-score [*distance*, $F(2,18) = 0.204$, $p = 0.817$, $\eta_p^2 = 0.003$; *angle*, $F(1.23,11.07) = 1.489$, $p = 0.256$, $\eta_p^2 = 0.025$; *distance* × *angle* interaction, $F(4,36) = 0.536$, $p = 0.71$, $\eta_p^2 = 0.02$]. Last, the mean F1-scores of each human observer did not differ significantly, $F(1,79) = 2.099$, $p = 0.151$, $\eta_p^2 = 0.026$.

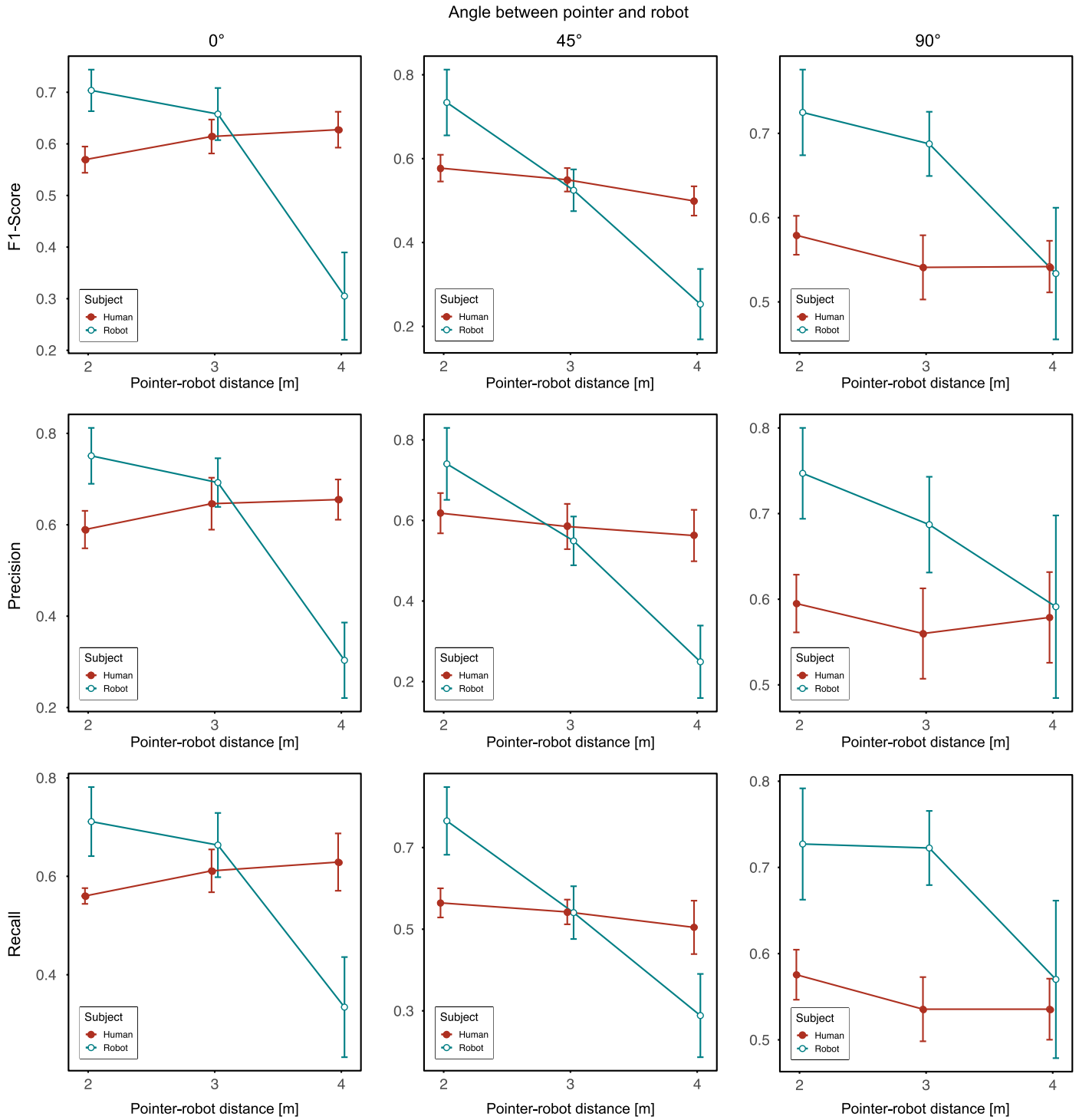

Fig. 9. Recognition accuracy of the human observers and the robot (Experiment 3).

Discussion

Deictic gestures are one of the most frequently used HHI techniques due to their intuitive nature and quality of communication. Nevertheless, the limited accuracy of humans when pointing targets in remote locations is widely accepted and it poses some challenges for both humans and robots in leveraging deictic gestures for HRI (Mayer et al. 2020). Therefore, it is important to identify the feasibility of deictic gestures and find applicable solutions for improving performance. This study aims to extend current knowledge on the performance in communicating positional information using deictic gestures in the construction domain. We evaluated the estimation and recognition accuracy of spatial communication with experimental tasks developed in this study, which involved human pointers, human observers, and a robot. There are two main findings regarding the challenges in spatial communication using deictic gestures in large-scale environments. First, the effects of the target plane, target angle from the pointer, distance between the pointer and the target, and target layout contribute to the challenges in both pointing targets and interpreting their locations within

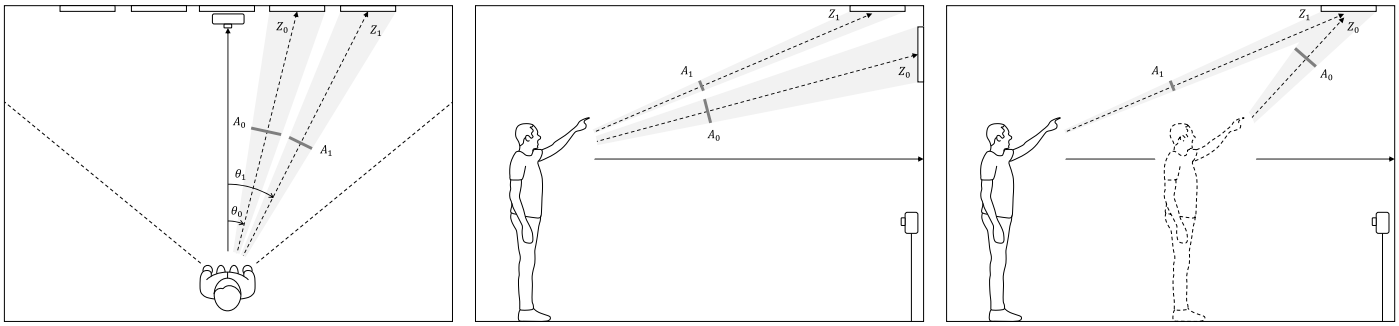

Fig. 10. The top and side views of the experimental environment: (a) center versus side panels; (b) wall versus ceiling panels; and (c) 3 m versus 1 m target distance.

large-scale environments like construction sites. Second, for the ceiling targets, HRI performance was more prone to change in the positional relationship compared with HHI performance, especially to change in the distance between the pointer and the observer.

The mean deviation from the target was decreased by 31.07% through the alteration in the target plane (from ceiling to wall) by 13.76% through the difference in the target angle (from side to center) and by 82.79% through the difference in the target distance (from 3 to 1 m). One possible explanation of these tendencies, which were common in all participants, may lie in the human vision via pinhole projection, in which the 3D visible world (on the world coordinate system) is projected onto a 2D projection plane (2D retina in the case of the human); this plane is focal length d away from the projective center along with the Z_h axis (on the human coordinate system), the gaze direction (Sharma et al. 2020). Fig. 10 shows schematic illustrations of the experimental environment of Experiments 1 and 2. A_0 and A_1 refer to the projected area of the target panels, perpendicular to the gaze directions Z_0 and Z_1, respectively (this study assumed that a person gazed at the center of the target panels when pointing). The differences in target planes and angles elicit the differences in the projected area: wall or center panels have larger projected areas A_0 compared with the projected areas of the ceiling or side panels A_1. We concluded that a smaller projected area hindered humans from pointing precisely while maintaining consistency. In addition, the panel layout had an influence on the robot's recognition accuracy. Under the same condition of the spatial communication excluding the layout of the panels, in other words, comparing the ceiling pointing tasks in Experiment 1 and observation Position O4 in Experiment 3, the mean F1-score dropped by 19.26% when the panels were installed both in a horizontal and vertical layout (E1: $M = 0.815$, $SD = 0.256$; E3: $M = 0.658$, $SD = 0.151$). These results are in line with previous evaluations on humans' limited pointing performance on distant targets in collaborative virtual environments (CVEs) (Mayer et al. 2018, 2020). We add to that literature by showing that similar tendencies appear in the real world and that humans' limited pointing accuracy is also influenced by the target plane, the horizontal angle between the pointer and the target, and the target layout.

The mean F1-score of each experimental condition did not show a significant difference between HHI and HRI in both Experiments 2 and 3. This indicates that regardless of the changes in pointing positions or observation positions, the robot's recognition accuracy in estimating the pointing position in the ceiling is mostly comparable to that of humans. Moreover, among nine observation positions in Experiment 3, no significant differences in human performance were observed. In other words, humans tend to have a consistent recognition accuracy for the ceiling targets regardless of the observation positions. Meanwhile, in the case of the robot, we observed that the deviation from the target was significantly influenced by both the distance and angle between the pointer and the robot. Most importantly, the distance between the pointer and the robot also had a significant effect on the robot's recognition accuracy. In particular, the accuracy dropped significantly at a distance of 4 m. This is not surprising because 0.3–3 m is specified as the ideal range for the RGB-D camera used in this study (Intel 2022). Therefore, we show that securing a close distance from a pointer is most effective and should be a priority for reliable performance in communicating positional information with a robot.

The robot performance remained high at a pointer-robot angle of 90°, compared with an angle of 0° and 45°, in which the performance dropped as the robot moves away from the pointer. We interpret this as a result of the deictic gesture recognition based on the pose estimation technique; the more right arm joints (shoulder, elbow, and wrist) utilized for the deictic gesture recognition that came into the camera view without depth occlusion, the more likely the system was to detect the exact position of the joints. The results confirm the intuition of the low possibility of depth occlusion while stretching the right arm for pointing at an angle of 90°, and thus, we suggest positioning a robot at 90° from a pointer as another way of enhancing the performance of human–robot spatial communication.

The pointing calibration showed a considerable improvement in the mean deviation from the targets. Without pointing calibration, it was observed that the mean deviation in Experiment 3 increased to 1.694 m from 0.449 m and the mean F1-score decreased to 0.064 from 0.569. Pointing calibration is generally considered one of the components of the current HRI methods using deictic gestures. However, in situations in which pointing calibration is unable to be performed before the use of the interface, such low performance could pose a major challenge in practical applications.

Envisioned Applications and Future Works

Deictic gestures are scalable for a wide range of HRI in construction in terms of their capability to represent various types of spatial information. These include target points and areas, as well as target objects. First, spatial communication on target points can be one of the envisioned applications in construction. Deictic gestures can be applied to construction tasks that require the positioning of end-of-arm tools (EOATs) at start points of seams, e.g., robotic welding (Lei et al. 2020) and caulking (Lundeen et al. 2017). They can also be used for tasks that need EOATs to be positioned at discrete

locations in a workspace, e.g., robotic drilling and anchoring (Gawel et al. 2019). Furthermore, information about target areas can be delivered by selecting the potential edges of a workspace using deictic gestures. The target areas can be floors, ceilings, walls, or facades, where construction tasks such as robotic painting (Asadi et al. 2018) or plastering (Mitterberger et al. 2022) are performed. In addition to target points and areas, deictic gestures can be employed to pick-and-place tasks in construction, e.g., robot-to-human construction material handover (Liu et al. 2021) and waste collection on sites (Chen et al. 2022).

The panel installation work, the scope of this study, would be another practical application of spatial communication on target objects. The workflow of human–robot collaborative panel installation is suggested by Liang et al. (2021a). This workflow begins with the robot setup process in which the robot navigates to the desired working station and acquires the geometric information of the as-built structure using the scene understanding methods (Liang et al. 2021a; Lundeen et al. 2017, 2019). Next, among the reachable panel grids, the human worker indicates the target grid location (Liang et al. 2021a). Then, the installation worker or robot manipulates and places a panel at the specified location (Liang et al. 2020). Last, the human worker performs quality checks and manually reworks if necessary (Liang et al. 2021a). Considering the second step of the workflow, spatial communication between the human and robot is needed to deliver the specified target location. This study is focused on the deictic gesture-based HRI as a spatial communication method.

Despite the high potential of the deictic gestures present as the HRI technique, applying such a method for spatial communication in the field is still challenging due to its low performance in large-scale construction environments. It is recommended that further research should be undertaken in consideration of the two-way communication of the HRI assisted by real-time robot feedback. As stated by Gromov et al. (2020), the lack of real-time robot feedback is one of the main reasons for low performance because the user is informed about the pointing position perceived by the robot only when a robot ends its motion, making it difficult for the user to correct the pointing input. In this regard, Medeiros et al. (2021) integrated visual feedback from the point of view of an unmanned aerial vehicle (UAV) and showed that it could enhance the pointing accuracy of users by allowing the users to review the location of the referent inferred by the system. Whitney et al. (2017) showed that social feedback from an item-fetching robot, such as asking questions to clarify the target object, could improve the accuracy of the system by 11.1%. Furthermore, Sprute et al. (2019) used a laser pointer as an interaction device to provide users with direct visual feedback. The proposed method achieved high accuracy (84.6% of the Jaccard index) and was proven to be applicable to novice users. All the aforementioned studies confirmed the effectiveness of two-way communication in improving performance.

Conclusion

This study investigated the effects of spatial characteristics on spatial referencing using deictic gestures, focusing on the exchange of positional information about panel objects in HRI and HHI contexts. The results of three experiments revealed that the recognition and estimation of deictic gestures were influenced by the target plane, target position, and the target layout and that the robot performance was significantly reduced as the human–robot distance increased. The robot achieved an F1-score of 0.895 in the wall panels and 0.815 in the ceiling panels. In the ceiling pointing tasks with a more complex target configuration, the F1-score of the robot dropped to the minimum of 0.404, while humans ranged from 0.730 to 0.956. These results demonstrate the challenges in spatial referencing within a large-scale environment and highlight the need for exchanging feedback from a robot to employ deictic gestures and robotics in construction work.

The findings of this study extend the body of knowledge in the construction domain by providing significant implications on gesture-based human–robot interfaces in construction job sites that have recently garnered much interest. This study offers generalizable and objective criteria for determining the level of precision required for effective communication between humans and robots using those gesture-based interfaces and suggests applicable solutions for improvement. Furthermore, the physical settings of the experiments are generalizable and representative and can be applied in the field to investigate the effects of spatial characteristics on human–robot communication.

However, we are aware that this study may have several limitations. First, this study assumes that the participants would tend to point at the center of the target panel. Nonetheless, this tendency was observed in the experiments and was later confirmed verbally by all participants, even though they were only instructed to refer to the panel. Second, this study assumes that the robot has a full understanding of its surrounding environment and workspace geometry that is completely accurate with no measurement errors. Future works are needed to investigate robot performance when the robot must learn its surrounding environment. Next, the experiments are conducted in one of the most ideal environments where single-task construction robots for interior finishing tasks can perform given tasks with minimum failure, i.e., an indoor environment with no other materials, equipment, and workers hindering the robot's field of view and workspace. Therefore, factors of construction sites such as dust, moving objects, and occlusion were not included in this study. In addition, while deictic gestures are one of the universal gestures in construction sites, construction professionals of the participants were not considered. However, such an environmental setup that does not involve the abovementioned factors would be less common in real construction sites. Last, robots that actually perform construction operations were not deployed in this study. Research into communicating in situ improvisations with colocated construction robots is currently in progress by the authors.

Data Availability Statement

All data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This study was supported by the New Faculty Startup Fund and the Institute of Construction and Environmental Engineering (ICEE) at Seoul National University (SNU). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of ICEE or SNU.

References

Adami, P., P. B. Rodrigues, P. J. Woods, B. Becerik-Gerber, L. Soibelman, Y. Copur-Gencturk, and G. Lucas. 2022. "Impact of VR-based training on human–robot interaction for remote operating construction robots." *J. Comput. Civ. Eng.* 36 (3): 04022006. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001016.

Alibali, M. W. 2005. "Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information." *Spatial Cognit. Comput.* 5 (4): 307–331. https://doi.org/10.1207/s15427633scc0504_2.

Ameri, S. K., M. Kim, I. A. Kuang, W. K. Perera, M. Alshiekh, H. Jeong, U. Topcu, D. Akinwande, and N. Lu. 2018. "Imperceptible electrooculography graphene sensor system for human–robot interface." *NPJ 2D Mater. Appl.* 2 (1): 19. https://doi.org/10.1038/s41699-018-0064-4.

Annem, V., P. Rajendran, S. Thakar, and S. K. Gupta. 2019. "Towards remote teleoperation of a semi-autonomous mobile manipulator system in machine tending tasks." In *Proc., ASME 2019 14th Int. Manufacturing Science and Engineering Conf., MSEC 2019*. New York: ASME.

Asadi, E., B. Li, and I. M. Chen. 2018. "Pictobot: A cooperative painting robot for interior finishing of industrial developments." *IEEE Rob. Autom. Mag.* 25 (2): 82–94. https://doi.org/10.1109/MRA.2018.2816972.

Bolt, R. A. 1980. "'Put-that-there': Voice and gesture at the graphics interface." In *Proc., 7th Annual Conf. on Computer Graphics and Interactive Techniques*, 262–270. New York: Association for Computing Machinery.

Brand, D., A. Meschtscherjakov, and K. Büchele. 2016. "Pointing at the HUD: Gesture interaction using a leap motion." In *Proc., AutomotiveUI 2016—8th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications, Adjunct Proceedings*, 167–172. New York: Association for Computing Machinery.

Burke, J. L., R. R. Murphy, E. Rogers, V. J. Lumelsky, and J. Scholtz. 2004. "Final report for the DARPA/NSF interdisciplinary study on human-robot interaction." *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 34 (2): 103–112. https://doi.org/10.1109/TSMCC.2004.826287.

Butterworth, G. 2003. "Pointing is the royal road to language for babies." In *Pointing*, 17–42. East Sussex, UK: Psychology Press.

Canal, G., S. Escalera, and C. Angulo. 2016. "A real-time human-robot interaction system based on gestures for assistive scenarios." *Comput. Vision Image Understanding* 149 (Aug): 65–77. https://doi.org/10.1016/j.cviu.2016.03.004.

Cao, Z., G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh. 2021. "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields." *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (1): 172–186. https://doi.org/10.1109/TPAMI.2019.2929257.

Carra, G., A. Argiolas, A. Bellissima, M. Niccolini, and M. Ragaglia. 2018. "Robotics in the construction industry: State of the art and future opportunities." In *Proc., 35th Int. Symp. on Automation and Robotics in Construction (ISARC)*, 866–873. Cambridge, UK: International Association for Automation and Robotics in Construction.

Chen, X., H. Huang, Y. Liu, J. Li, and M. Liu. 2022. "Robot for automatic waste sorting on construction sites." *Autom. Constr.* 141 (Sep): 104387. https://doi.org/10.1016/j.autcon.2022.104387.

Civ Robotics. 2021. "CivDot." Accessed December 16, 2021. https://www.civrobotics.com/product.

COBOD. 2021. "The BOD2." Accessed December 16, 2021. https://cobod.com/solution/bod2/.

Construction Robotics. 2021. "SAM." Accessed December 16, 2021. https://www.construction-robotics.com/sam-2/.

David, O., F. X. Russotto, M. da Silva Simoes, and Y. Measson. 2014. "Collision avoidance, virtual guides and advanced supervisory control teleoperation techniques for high-tech construction: Framework design." *Autom. Constr.* 44 (Aug): 63–72. https://doi.org/10.1016/j.autcon.2014.03.020.

Dhingra, N., E. Valli, and A. Kunz. 2020. "Recognition and localisation of pointing gestures using a RGB-D camera." In *Communications in Computer and Information Science*, 205–212. Cham, Switzerland: Springer.

Ding, L., W. Jiang, Y. Zhou, C. Zhou, and S. Liu. 2020. "BIM-based task-level planning for robotic brick assembly through image-based 3D modeling." *Adv. Eng. Inf.* 43 (Jan): 100993. https://doi.org/10.1016/j.aei.2019.100993.

Dusty Robotics. 2021. "FieldPrinter." Accessed December 16, 2021. https://www.dustyrobotics.com/product.

FBR (Fastbrick Robotics). 2021. "Hadrian X." Accessed December 16, 2021. https://www.fbr.com.au/view/hadrian-x.

Feng, C., Y. Xiao, A. Willette, W. McGee, and V. R. Kamat. 2015. "Vision guided autonomous robotic assembly and as-built scanning on unstructured construction sites." *Autom. Constr.* 59 (Nov): 128–138. https://doi.org/10.1016/j.autcon.2015.06.002.

Follini, C., V. Magnago, K. Freitag, M. Terzer, C. Marcher, M. Riedl, A. Giusti, and D. T. Matt. 2021. "BIM-integrated collaborative robotics for application in building construction and maintenance." *Robotics* 10 (1): 1–19. https://doi.org/10.3390/robotics10010002.

Ganglbauer, M., M. Ikeda, M. Plasch, and A. Pichler. 2020. "Human in the loop online estimation of robotic speed limits for safe human robot collaboration." *Procedia Manuf.* 51 (Jan): 88–94. https://doi.org/10.1016/j.promfg.2020.10.014.

Gawel, A., et al. 2019. "A fully-integrated sensing and control system for high-accuracy mobile robotic building construction." In *Proc., IEEE Int. Conf. on Intelligent Robots and Systems*, 2300–2307. New York: IEEE. https://doi.org/10.1109/IROS40897.2019.8967733.

Gharbia, M., A. Chang-Richards, Y. Lu, R. Y. Zhong, and H. Li. 2020. "Robotic technologies for on-site building construction: A systematic review." *J. Build. Eng.* 32 (Nov): 101554. https://doi.org/10.1016/j.jobe.2020.101584.

Grahn, S., V. Gopinath, X. V. Wang, and K. Johansen. 2018. "Exploring a model for production system design to utilize large robots in human-robot collaborative assembly cells." *Procedia Manuf.* 25 (Jan): 612–619. https://doi.org/10.1016/j.promfg.2018.06.094.

Gromov, B., G. Abbate, L. M. Gambardella, and A. Giusti. 2019. "Proximity human-robot interaction using pointing gestures and a wrist-mounted IMU." In *Proc., 2019 Int. Conf. on Robotics and Automation (ICRA)*, 8084–8091. New York: IEEE.

Gromov, B., L. Gambardella, and A. Giusti. 2020. "Guiding quadrotor landing with pointing gestures." In *Springer proceedings in advanced robotics*, 1–14. Cham, Switzerland: Springer.

Gromov, B., L. M. Gambardella, and A. Giusti. 2018. "Robot identification and localization with pointing gestures." In *Proc., 2018 IEEE RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 3921–3928. New York: IEEE.

Gustavsson, P., M. Holm, A. Syberfeldt, and L. Wang. 2018. "Human-robot collaboration—Towards new metrics for selection of communication technologies." *Procedia CIRP* 72 (Jan): 123–128. https://doi.org/10.1016/j.procir.2018.03.156.

Hentout, A., M. Aouache, A. Maoudj, and I. Akli. 2019. "Human–robot interaction in industrial collaborative robotics: A literature review of the decade 2008–2017." *Adv. Rob.* 33 (15–16): 764–799. https://doi.org/10.1080/01691864.2019.1636714.

Heydaryan, S., J. S. Bedolla, and G. Belingardi. 2018. "Safety design and development of a human-robot collaboration assembly process in the automotive industry." *Appl. Sci.* 8 (3): 344. https://doi.org/10.3390/app8030344.

Inkulu, A. K., M. R. Bahubalendruni, and A. Dara. 2022. "Challenges and opportunities in human robot collaboration context of Industry 4.0: A state of the art review." *Ind. Robot* 49 (2): 226–239. https://doi.org/10.1108/IR-04-2021-0077.

Intel. 2022. "Intel RealSense Camera 400 series product family datasheet." Accessed March 17, 2022. https://www.intelrealsense.com/depth-camera-d435/.

ISO. 2021. *Robotics: Vocabulary.* ISO 8373:2021. Geneva: ISO.

Jevtić, A., G. Doisy, Y. Parmet, and Y. Edan. 2015. "Comparison of interaction modalities for mobile indoor robot guidance: Direct physical interaction, person following, and pointing control." *IEEE Trans. Hum.-Mach. Syst.* 45 (6): 653–663. https://doi.org/10.1109/THMS.2015.2461683.

Jevtić, A., A. F. Valle, G. Alenyà, G. Chance, P. Caleb-Solly, S. Dogramadzi, and C. Torras. 2019. "Personalized robot assistant for support in dressing." *IEEE Trans. Cognit. Dev. Syst.* 11 (3): 363–374. https://doi.org/10.1109/TCDS.2018.2817283.

Jirak, D., D. Biertimpel, M. Kerzel, and S. Wermter. 2021. "Solving visual object ambiguities when pointing: An unsupervised learning approach." *Neural Comput. Appl.* 33 (7): 2297–2319. https://doi.org/10.1007/s00521-020-05109-w.

Keskin, C., A. Erkan, and L. Akarun. 2003. "Real time hand tracking and 3D gesture recognition for interactive interfaces using hmm." In *Proc., Joint Int. Conf. ICANN/ICONIP.* Berlin: Springer-Verlag.

Kim, K., and Y. K. Cho. 2015. "Construction-specific spatial information reasoning in building information models." *Adv. Eng. Inf.* 29 (4): 1013–1027. https://doi.org/10.1016/j.aei.2015.08.004.

Kim, S. W., B. Qin, Z. J. Chong, X. Shen, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus. 2015. "Multivehicle cooperative driving using cooperative perception: Design and experimental validation." *IEEE Trans. Intell. Transp. Syst.* 16 (2): 663–680. https://doi.org/10.1109/TITS.2014.2337316.

Kim, Y., H. Kim, R. Murphy, S. Lee, and C. R. Ahn. 2022. "Delegation or collaboration: Understanding different construction stakeholders' perceptions of robotization." *J. Manage. Eng.* 38 (1): 04021084. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000994.

Koh, K. H., M. Farhan, K. P. C. Yeung, D. C. H. Tang, M. P. Y. Lau, P. K. Cheung, and K. W. C. Lai. 2021. "Teleoperated service robotic system for on-site surface rust removal and protection of high-rise exterior gas pipes." *Autom. Constr.* 125 (May): 103609. https://doi.org/10.1016/j.autcon.2021.103609.

Kumar, P., J. Verma, and S. Prasad. 2012. "Hand data glove: A wearable real-time device for human-computer interaction." *Int. J. Adv. Sci. Technol.* 43 (Jun): 15–26.

Kyjanek, O., B. Al Bahar, L. Vasey, B. Wannemacher, and A. Menges. 2019. "Implementation of an augmented reality AR workflow for human robot collaboration in timber prefabrication." In *Proc., 36th Int. Symp. on Automation and Robotics in Construction, ISARC*, 1223–1230. Cambridge, UK: International Association for Automation and Robotics in Construction.

Lai, Y., C. Wang, Y. Li, S. S. Ge, and D. Huang. 2016. "3D pointing gesture recognition for human-robot interaction." In *Proc., 28th Chinese Control and Decision Conf., CCDC 2016*, 4959–4964. New York: IEEE.

Lamon, E., M. Leonori, W. Kim, and A. Ajoudani. 2020. "Towards an intelligent collaborative robotic system for mixed case palletizing; towards an intelligent collaborative robotic system for mixed case palletizing." In *Proc., 2020 IEEE Int. Conf. on Robotics and Automation (ICRA)*. New York: IEEE.

Lei, T., Y. Rong, H. Wang, Y. Huang, and M. Li. 2020. "A review of vision-aided robotic welding." *Comput. Ind.* 123 (Dec): 103326. https://doi.org/10.1016/j.compind.2020.103326.

Li, X. 2020. "Human–robot interaction based on gesture and movement recognition." *Signal Process. Image Commun.* 81 (Feb): 115686. https://doi.org/10.1016/j.image.2019.115686.

Li, Y., J. Huang, F. Tian, H. A. Wang, and G. Z. Dai. 2019. "Gesture interaction in virtual reality." *Virtual Reality Intell. Hardware* 1 (1): 84–112. https://doi.org/10.3724/SP.J.2096-5796.2018.0006.

Liang, C. J., V. R. Kamat, and C. C. Menassa. 2020. "Teaching robots to perform quasi-repetitive construction tasks through human demonstration." *Autom. Constr.* 120 (Dec): 103370. https://doi.org/10.1016/j.autcon.2020.103370.

Liang, C.-J., V. R. Kamat, C. C. Menassa, and W. Mcgee. 2021a. "Trajectory-based skill learning for overhead construction robots using generalized cylinders with orientation." *J. Comput. Civ. Eng.* 36 (2): 04021036. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001004.

Liang, C.-J., X. Wang, V. R. Kamat, and C. C. Menassa. 2021b. "Human–robot collaboration in construction: Classification and research trends." *J. Constr. Eng. Manage.* 147 (10): 03121006. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002154.

Liu, Y., M. Habibnezhad, and H. Jebelli. 2021. "Brain-computer interface for hands-free teleoperation of construction robots." *Autom. Constr.* 123 (Mar): 103523. https://doi.org/10.1016/j.autcon.2020.103523.

Lundeen, K. M., V. R. Kamat, C. C. Menassa, and W. McGee. 2017. "Scene understanding for adaptive manipulation in robotized construction work." *Autom. Constr.* 82 (Oct): 16–30. https://doi.org/10.1016/j.autcon.2017.06.022.

Lundeen, K. M., V. R. Kamat, C. C. Menassa, and W. McGee. 2019. "Autonomous motion planning and task execution in geometrically adaptive robotized construction work." *Autom. Constr.* 100 (Apr): 24–45. https://doi.org/10.1016/j.autcon.2018.12.020.

Makris, S., P. Karagiannis, S. Koukas, and A. S. Matthaiakis. 2016. "Augmented reality system for operator support in human–robot collaborative assembly." *CIRP Ann.* 65 (1): 61–64. https://doi.org/10.1016/j.cirp.2016.04.038.

Malik, A. A., and A. Bilberg. 2019. "Developing a reference model for human–robot interaction." *Int. J. Interact. Des. Manuf.* 13 (4): 1541–1547. https://doi.org/10.1007/s12008-019-00591-6.

Mayer, S., J. Reinhardt, R. Schweigert, B. Jelke, V. Schwind, K. Wolf, and N. Henze. 2020. "Improving humans' ability to interpret deictic gestures in virtual reality." In *Proc., Conf. on Human Factors in Computing Systems*. New York: Association for Computing Machinery.

Mayer, S., V. Schwind, R. Schweigert, and N. Henze. 2018. "The effect of offset correction and cursor on mid-air pointing in real and virtual environments." In *Proc., Conf. on Human Factors in Computing Systems*. New York: Association for Computing Machinery.

Mayer, S., K. Wolf, S. Schneegass, and N. Henze. 2015. "Modeling distant pointing for compensating systematic displacements." In *Proc., Conf. on Human Factors in Computing Systems*, 4165–4168. New York: Association for Computing Machinery. https://doi.org/10.1145/2702123.2702332.

Medeiros, A. C. S., P. Ratsamee, J. Orlosky, Y. Uranishi, M. Higashida, and H. Takemura. 2021. "3D pointing gestures as target selection tools: Guiding monocular UAVs during window selection in an outdoor environment." *ROBOMECH J.* 8 (1): 1–19. https://doi.org/10.1186/s40648-021-00200-w.

Mitterberger, D., S. Ercan Jenny, L. Vasey, E. Lloret-Fritschi, P. Aejmelaeus-Lindström, F. Gramazio, and M. Kohler. 2022. "Interactive robotic plastering: Augmented interactive design and fabrication for on-site robotic plastering." In *Proc., Conf. on Human Factors in Computing Systems*. New York: Association for Computing Machinery.

MX3D. 2021. "MX3D." Accessed December 16, 2021. https://mx3d.com/.

Navas Medrano, S., M. Pfeiffer, and C. Kray. 2020. "Remote deictic communication: Simulating deictic pointing gestures across distances using electro muscle stimulation." *Int. J. Hum.-Comput. Interact.* 36 (19): 1867–1882. https://doi.org/10.1080/10447318.2020.1801171.

Nickel, K., and R. Stiefelhagen. 2003. "Pointing gesture recognition based on 3D-tracking of face, hands and head orientation." In *Proc., ICMI'03: 5th Int. Conf. on Multimodal Interfaces*, 140–146. New York: Association for Computing Machinery. https://doi.org/10.1145/958432.958460.

Okishiba, S., R. Fukui, M. Takagi, H. Azumi, S. Warisawa, R. Togashi, H. Kitaoka, and T. Ooi. 2019. "Tablet interface for direct vision teleoperation of an excavator for urban construction work." *Autom. Constr.* 102 (Jun): 17–26. https://doi.org/10.1016/j.autcon.2019.02.003.

Oosterwijk, A. M., M. de Boer, A. Stolk, F. Hartmann, I. Toni, and L. Verhagen. 2017. "Communicative knowledge pervasively influences sensorimotor computations." *Sci. Rep.* 7 (1): 1–12. https://doi.org/10.1038/s41598-017-04442-w.

Park, J., and Y. K. Cho. 2017. "Development and evaluation of a probabilistic local search algorithm for complex dynamic indoor construction sites." *J. Comput. Civ. Eng.* 31 (4): 04017015. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000658.

Roldán, J. J., et al. 2019. "Multi-robot systems, virtual reality and ROS: Developing a new generation of operator interfaces." In Vol. 778 of *Robot Operating System (ROS). Studies in Computational Intelligence*, 29–64. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-91590-6_2.

Sauppé, A., and B. Mutlu. 2014. "Robot deictics: How gesture and context shape referential communication." In *Proc., ACM/IEEE Int. Conf. on Human-Robot Interaction*, 342–349. New York: IEEE.

Sharma, A., R. Nett, and J. Ventura. 2020. "Unsupervised learning of depth and ego-motion from cylindrical panoramic video with applications for virtual reality." *Int. J. Semant. Comput.* 14 (3): 333–356. https://doi.org/10.1142/S1793351X20400139.

Sprute, D., R. Rasch, A. Portner, S. Battermann, and M. Konig. 2018. "Gesture-based object localization for robot applications in intelligent environments." In *Proc., 2018 Int. Conf. on Intelligent Environments, IE 2018*, 48–55. New York: IEEE.

Sprute, D., K. Tönnies, and M. Konig. 2019. "This far, no further: Introducing virtual borders to mobile robots using a laser pointer." In *Proc., 3rd IEEE Int. Conf. on Robotic Computing, IRC 2019*, 403–408. New York: IEEE. https://doi.org/10.1109/IRC.2019.00074.

Steinfeld, A., T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. 2006. "Common metrics for human-robot interaction."

In *Proc., HRI 2006: 2006 ACM Conf. on Human-Robot Interaction*, 33–40. New York: Association for Computing Machinery.

Tashtoush, T., L. Garcia, G. Landa, F. Amor, A. N. Laborde, D. Oliva, and F. Safar. 2021. "Human-robot interaction and collaboration (HRI-C) utilizing top-view RGB-D camera system." *Int. J. Adv. Comput. Sci. Appl.* 12 (1): 10–17. https://doi.org/10.14569/IJACSA.2021.0120102.

Tavares, P., C. M. Costa, L. Rocha, P. Malaca, P. Costa, A. P. Moreira, A. Sousa, and G. Veiga. 2019. "Collaborative welding system using BIM for robotic reprogramming and spatial augmented reality." *Autom. Constr.* 106 (Oct): 102825. https://doi.org/10.1016/j.autcon.2019.04.020.

Tölgyessy, M., M. Dekan, F. Duchoň, J. Rodina, P. Hubinský, and L. Chovanec. 2017. "Foundations of visual linear human–robot interaction via pointing gesture navigation." *Int. J. Social Rob.* 9 (4): 509–523. https://doi.org/10.1007/s12369-017-0408-9.

Vasilyeva, M., and S. F. Lourenco. 2012. "Development of spatial cognition." *Wiley Interdiscip. Rev. Cognit. Sci.* 3 (3): 349–362. https://doi.org/10.1002/wcs.1171.

Walkowski, S., R. Dörner, M. Lievonen, and D. Rosenberg. 2011. "Using a game controller for relaying deictic gestures in computer-mediated communication." *Int. J. Hum.-Comput. Stud.* 69 (6): 362–374. https://doi.org/10.1016/j.ijhcs.2011.01.002.

Wang, X., C.-J. Liang, C. C. Menassa, and V. R. Kamat. 2021. "Interactive and immersive process-level digital twin for collaborative human–robot construction work." *J. Comput. Civ. Eng.* 35 (6): 04021023. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000988.

Weerakoon, D., V. Subbaraju, N. Karumpulli, T. Tran, Q. Xu, U. X. Tan, J. H. Lim, and A. Misra. 2020. "Gesture enhanced comprehension of ambiguous human-to-robot instructions." *Proc., 2020 Int. Conf. on Multimodal Interaction: ICMI 2020*, 251–259. New York: Association for Computing Machinery.

Weng, T., L. Perlmutter, S. Nikolaidis, S. Srinivasa, and M. Cakmak. 2019. "Robot object referencing through legible situated projections." In *Proc., 2019 Int. Conf. on Robotics and Automation (ICRA)*, 8004–8010. New York: IEEE.

Whitney, D., E. Rosen, J. Macglashan, L. L. S. Wong, and S. Tellex. 2017. "Reducing errors in object-fetching interactions through social feedback." In *Proc., 2017 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 1006–1013. New York: IEEE.

Williams, T., M. Bussing, S. Cabrol, E. Boyle, and N. Tran. 2019. "Mixed reality deictic gesture for multi-modal robot communication." In *Proc., ACM/IEEE Int. Conf. on Human-Robot Interaction*, 191–201. New York: IEEE. https://doi.org/10.1109/HRI.2019.8673275.

Yongda, D., L. Fang, and X. Huang. 2018. "Research on multimodal human-robot interaction based on speech and gesture." *Comput. Electr. Eng.* 72 (Nov): 443–454. https://doi.org/10.1016/j.compeleceng.2018.09.014.

Zamani, M. A., H. Beik-Mohammadi, M. Kerzel, S. Magg, and S. Wermter. 2018. "Learning spatial representation for safe human-robot collaboration in joint manual tasks." In *Proc., ICRA Workshop on the Workplace Is Better with Intelligent, Collaborative, Robot MATEs (WORKMATE)*. New York: IEEE.

Zhou, T., Q. Zhu, and J. Du. 2020. "Intuitive robot teleoperation for civil engineering operations with virtual reality and deep learning scene reconstruction." *Adv. Eng. Inf.* 46 (Oct): 101170. https://doi.org/10.1016/j.aei.2020.101170.

Zwicker, C., and G. Reinhart. 2014. "Human-robot-collaboration system for a universal packaging cell for heavy electronic consumer goods." In *Proc., Enabling Manufacturing Competitiveness and Economic Sustainability: Proc., of the 5th Int. Conf. on Changeable, Agile, Reconfigurable and Virtual Production (CARV 2013)*, 195–199. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-02054-9_33.