Challenges in Deictic Gesture-Based Spatial Referencing for Human-Robot Interaction in Construction

Sungboo Yoon[a], YeSeul Kim[b], Changbum R. Ahn[a], and Moonseo Park[a]

[a]Department of Architecture and Architectural Engineering, Seoul National University, South Korea
[b]Department of Multidisciplinary Engineering, Texas A&M University, College Station, TX 77843
E-mail: yoonsb24@snu.ac.kr, leslieyskim@tamu.edu, cbahn@snu.ac.kr, mspark@snu.ac.kr

Abstract –

As robots are envisioned to be deployed in construction job sites to work with humans, there is an increasing need for developing intuitive and natural communication between robots and humans. In particular, spatial information exchange is critical to navigating or delegating tasks to collaborative robots. However, such deictic gestures are inherently imprecise and ambiguous. Thus, it is challenging for robots to reason about the exact region of interest, especially in a cluttered large-scale construction environment. To address this limitation, this study evaluates the performance of spatial information exchange through the experiments based on pointing targets on the wall and ceiling, which are the most common workspaces in construction. We observed that the current deictic gesture-based method can estimate the pointed position on the wall and ceiling with a mean distance error of 0.767m, while the error tends to increase by 0.715m in the ceiling and 0.115m in the side panels. Our experimental results indicate that the deictic gesture-based method has some challenges in ceiling and side panel conditions, while the overall panel recognition shows acceptable performance. The findings of this study will help novice construction workers naturally and effectively communicate with robots by delivering spatial information on specific objects or regions in the shared workspace.

Keywords –

Deictic Gestures; Spatial Referencing; Human-Robot Interaction

1 Introduction

As robotic technologies have advanced, the focus of robot adoption has shifted from large-scale robotic platforms to small-scaled task robots [1]. These robots are designed for various applications and have shown possibilities of the collaboration of robots and humans in construction sites. As they interact with people during task execution, there is a growing need for a more intuitive and natural human-robot communication interface. In particular, spatial information exchange (e.g., target objects or regions of interest) is critical to navigating or delegating tasks to collaborative robots. Potential human-robot collaborative applications, for example, include controlling a robot to change its position [2], referring to a target object [3], and indicating target ceiling panel for installation [4] or target wall area for painting [5]. In human-human interactions, people often utilize deictic gestures to deliver spatial information, which are effective means to develop a mutual understanding of a referent with others [6]. Deictic gestures are especially beneficial for construction applications because they require no additional devices, and therefore are intuitive for novice users.

However, deictic gestures are known to be inherently imprecise and ambiguous for both humans and robots. It is especially challenging for a robot to reason about the exact region of interest in an unstructured and cluttered construction environment. Construction tasks are carried out in a large-scale 3-dimensional (3D) environment; thus, it is necessary to share various dimensional and scaled spatial information (e.g., floor vs. wall vs. ceiling, centered vs. angled). Although several previous works showed the performance of gesture-based spatial information exchange for short-distance applications (e.g., table, wall, and floor [7]), it has not been evaluated in a comparatively large environment. Thus, this study aims to identify the challenges associated with the deictic gesture-based spatial referencing in a large-scale environment. This was done by evaluating the performance of spatial information exchange through the experiments based on pointing targets on the wall and ceiling, which are the most common workspaces in construction. The findings can guide and inform our approaches to developing collaborative construction robots supported with a natural human-robot interface.

2 Background

2.1 Deictic Gesture-Based Spatial Referencing

Deictic gestures are often referred to as "pointing gestures", typically performed by extending the arm and the index finger [8,9]. In general, people often use pointing gestures to deliver spatial information to others. In other words, deictic gestures are fundamental to direct others' attention to objects and help develop a mutual understanding of objects in space [10,11].

Deictic gesture-based spatial referencing has been explored substantially in previous works for developing and evaluating various spatial referencing models according to task requirements. This large body of work shares the same purpose: to solve the problem of interpreting deictic gestures in order to map the referent in the environment that the user wants to indicate [2]. Tölgyessy et al. [12] presented a spatial referencing method navigating a mobile robot to an endpoint marker on the ground floor defined by a pointing gesture of a human operator. The suggested method shows the precise positioning of all the entities included in the interaction in 3D space. Jevtić et al. [13] employed pointing recognition for selecting shoes placed on the platform in front of the user. Furthermore, they exploited the concept of multi-modality to develop personalized interaction with a robot assistant for assisted dressing. Mayer et al. [14] evaluated humans' referencing accuracy when interpreting deictic gestures for pointing the targets positioned horizontally on the wall. However, they only measured the performance in a collaborative virtual environment (CVE).

While previous works showed acceptable performance of the deictic gesture-based spatial referencing for short-distance applications, limited applications in large-scale environments need to be further evaluated.

2.2 Deictic Gesture Recognition

Deictic gesture-based spatial referencing aims to exchange accurate spatial information through deictic gestures. Therefore, deictic gesture recognition has a significant impact on the final referencing results.

Two main approaches for deictic gesture recognition have been proposed in the literature. One is a wearable sensor-based approach. This approach attempts to recognize deictic gesture by analyzing the electrical muscle stimulation (EMS) from electromyography (EMG) generated during the muscle activity [15,16], the change in measures from inertial measurement units (IMUs) [2,17], and the posture and motion data from data gloves [18]. However, although wearable sensors have the benefit of direct acquisition of the spatial posture of the pointing arm, they often require connection to a data acquisition (DAQ) device, thus restricting the applicability of this method outside of a controlled environment [19,20].

Meanwhile, recent advances in computer vision technologies have brought vision-based approaches to mainstream deictic gesture recognition. Vision-based deictic gesture recognition does not require users any additional devices and only employs their pointing arms within the camera angle. Earlier approaches detected gestures through the visual features (i.e., skin-color blobs) collected from monocular cameras (e.g., RGB or infrared camera) [21] and binocular cameras [22].

Recent works on vision-based approaches have focused on the implementation of RGB-D cameras. Owing to the ability to augment the RGB image with depth information, RGB-D cameras are frequently being adopted in vision-based approaches.

In a vision-based approach, the deictic gesture is defined based on the relationships among the body joints. Three main models for estimating the pointing direction were developed [12,23]:

- *Elbow-wrist model* assumes that the pointing direction is defined by a vector connecting the elbow and the wrist (hand) of the pointing arm.
- *Head-wrist model* assumes that the pointing direction is defined by a vector connecting the head and the wrist (hand) of the pointing arm.
- *Shoulder-wrist model* assumes that the pointing direction is defined by a vector connecting the shoulder and the wrist (hand) of the pointing arm.

The choice of a particular model mainly depends on the task and on the technology available for sensing the subject's posture [2]. This work evaluates the performance of the spatial referencing method using a shoulder-wrist model, because the elbow-wrist model gives lower accuracy in large-scale environments and the head-wrist model has potential problems associated with the occlusion in pose estimation (i.e., safety helmets) [24].

3 Methodology

3.1 Deictic Gesture Detection

The detection of the deictic gesture is performed based on the 3D human skeletal data extracted from the RGB and depth images. To estimate the human skeletal data, we employ OpenPose [25] library, a real-time human pose estimation system. The library (BODY-25 model) detects 25 human body joints from each RGB image frame in 2D coordinates. The 2D coordinates are then projected to corresponding 3D points using the depth information [26].

In particular, we focus on the position of the shoulder $\mathbf{p}_s = (x_s, y_s, z_s)$, elbow $\mathbf{p}_e = (x_e, y_e, z_e)$, and the wrist

$\mathbf{p}_w = (x_w, y_w, z_w)$ joints for deictic gesture detection, which is required for the selected shoulder-wrist model. We use wrist position instead of fingers, considering the computation efficiency for further on-site applications. Given the position of the three body joints, the elbow joint angle θ is defined by:

$$\cos\theta = \frac{\mathbf{v}_{se} \cdot \mathbf{v}_{sw}}{|\mathbf{v}_{se}||\mathbf{v}_{sw}|} \tag{1}$$

where $\mathbf{v}_{se} = \mathbf{p}_e - \mathbf{p}_s$ is the vector from the shoulder to the elbow joint and $\mathbf{v}_{sw} = \mathbf{p}_w - \mathbf{p}_s$ is the 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" and performs the panel estimation.

3.2 Pointed Panel Estimation

To estimate the pointed panel, we first compute the pointed position. The pointing direction is defined by a straight line starting from the shoulder to the wrist joint:

$$\mathbf{s} = \mathbf{p}_s + \lambda(\mathbf{p}_w - \mathbf{p}_s), \lambda \in \mathbb{R} \tag{2}$$

For a ceiling pointing task, panels are parallel to floor at a constant height of the ceiling h. Therefore, the pointed position $\mathbf{p}_p = (x_p, y_p, z_p)$ is calculated as follows:

$$x_p = x_s + \frac{h - z_s}{z_w - z_s}(x_w - x_s) \tag{3}$$

$$y_p = y_s + \frac{h - z_s}{z_w - z_s}(y_w - y_s) \tag{4}$$

$$z_p = h \tag{5}$$

In a wall pointing task, panels are parallel to wall at a constant distance d. Thus, akin to ceiling, the pointed position in this case is computed as:

$$x_p = d \tag{6}$$

$$y_p = y_s + \frac{d - x_s}{x_w - x_s}(y_w - y_s) \tag{7}$$

$$z_p = z_s + \frac{d - x_s}{x_w - x_s}(z_w - z_s) \tag{8}$$

The pointed target is then estimated using the pointed position. Let $\mathbf{p}_{t,i}$ be the center point of the target panel index $i \in \{1, 2, 3, \ldots, n\}$, where $n \in \mathbb{N}$. A panel with the closest Euclidean distance from the center point is selected as a pointed panel i_p.

$$i_p = \arg\min_i(|\mathbf{p}_p - \mathbf{p}_{t,i}|) \tag{9}$$

4 Experiment

4.1 Experimental Setup

The experimental setup is depicted in Figure 1. The RGB and depth images are simultaneously captured by Intel RealSense™ Depth Camera D435 at a frame rate of up to 30 fps and with an image resolution of 640 x 480 pixels and 840 x 480 pixels, respectively. The camera has an operating range of 0.11-10m. It is installed at position $\mathbf{p}_c$ facing participants, at the height of 0.7m and the pointing subject is located at position $\mathbf{p}_h$, 3.0m away from the camera. Five target panels with an equal size of 0.7 x 0.7m are located side by side on both ceiling and wall.

Four participants (two males and two females) were recruited to perform the pointing tasks. Each participant performed two experiments, 15 iterations for each experiment. A single iteration consists of 10 pointing trials: five ceiling panels (from C1 to C5) and five wall panels (from W1 to W5), in sequential order. In sum, we obtained 2 x 15 x 10 = 300 trials for each participant. A single experiment took approximately 10 minutes per participant.

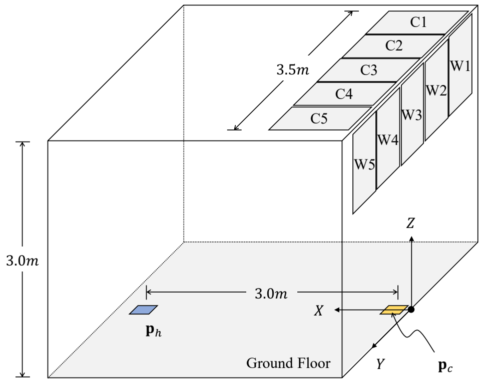

Figure 1. Illustration of the experimental environment.

4.2 Performance Metrics

We use the following metrics to evaluate the performance of the deictic gesture-based spatial referencing.

Distance error. Euclidean distance between the pointed position $\mathbf{p}_p$ and the center point of the target panel $\mathbf{p}_t$.

$$\varepsilon = |\mathbf{p}_p - \mathbf{p}_t|. \tag{10}$$

F1 score. We also refer to this measure as panel recognition rate (see Section 5.2). F1 score is defined by:

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

For each pointed position, we consider it as true positive (TP) if it is classified as a correct target panel, false negative (FN) if it is classified as other target panels, false positive (FP) if the subject is pointing at other target panels, and true negative (TN) if the subject is pointing at other target panels but classified correctly.

5 Results

A total of 1,200 pointing trials of four pointing subjects were evaluated in an offline setting in order to validate the performance of the spatial referencing method. The main results are shown in Table 1 and Table 2.

5.1 Distance Error

The distance error by the pointing subject is shown in Figure 2. We observed a similar tendency between all four participants: on average, the ceiling pointing task yielded a higher mean distance error and standard deviation (1.125 ± 0.263m) compared to the wall pointing task (0.410 ± 0.174m).

Among the subjects, Subject 3 reached the highest mean distance error for the ceiling pointing task with 0.482m of distance gap between the lowest, Subject 1. For the wall, Subject 2 showed the highest distance error with 0.207m of distance gap between the lowest, Subject 1.

Moreover, the mean distance error was higher when pointing the side panels (0.836 ± 0.241 for C1/C5 and W1/W5) than the panels near the center (0.721 ± 0.210 for C2-C4 and W2-W4). This phenomenon will be expanded up in Section 5.

5.2 Panel Recognition Rate

We found that the ceiling pointing task shows a lower panel recognition rate. The mean F1 score of the ceiling pointing task was 0.815, while the wall pointing task was 0.896.

Higher error distance increases the probability of inferring the wrong panels located nearby, which in turn lowers the panel recognition rate. Therefore, the F1 score tends to follow the reverse order of the distance error. This tendency is especially salient in the panels near the center.

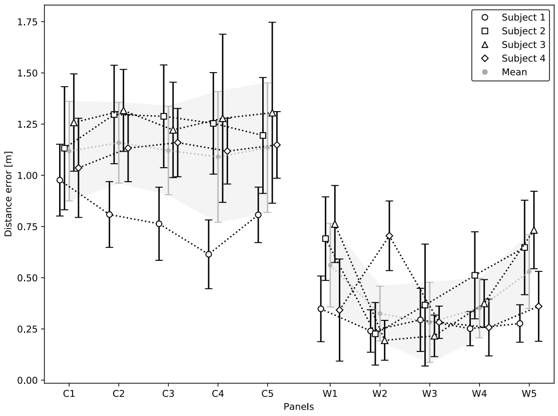

Figure 2. Distance error by the pointing subject. (C1-C5 and W1-W5 refers to target ceiling and wall panels from left to right, respectively.)

Table 1. Evaluation results of the ceiling pointing task: Distance error (Mean ± SD) and F1 score.

Metrics	C1	C2	C3	C4	C5
Distance Error (m)	1.118 ± 0.243	1.159 ± 0.197	1.121 ± 0.216	1.090 ± 0.319	1.135 ± 0.316
F1 Score	0.849	0.826	0.837	0.756	0.808

Table 2. Evaluation results of the wall pointing task: Distance error (Mean ± SD) and F1 score.

Metrics	W1	W2	W3	W4	W5
Distance Error (m)	0.561 ± 0.204	0.325 ± 0.134	0.282 ± 0.196	0.352 ± 0.145	0.530 ± 0.180
F1 Score	0.948	0.802	0.946	0.861	0.921

6 Discussion

Our experimental results present that the current deictic gesture-based method can estimate the pointed position on the wall and ceiling with a mean distance error of 0.767m. We observed worse performance in the ceiling with a mean distance error of 1.125m, which was 0.715m higher than the mean distance error of the wall. In the estimation of the panel, the mean F1 score dropped at a rate of 8.98% compared to the wall. These measures indicate that variation in plane causes a performance gap in the dimensional information exchange regarding the perception accuracy and target recognition rate.

Furthermore, the mean distance error tends to increase by 15.91% when the target changes from the center to side panels. In this situation, a subject mainly delivers the scaled spatial information to a robot.

In general, these results can be explained by the pinhole projection model (Figure 3). Human eyes see the world via pinhole projection. The 3D world (on the world coordinate system) is projected onto a flat projection plane: this plane is focal length d away from the projective center along the Z_h axis (on the human coordinate system), the gaze direction [27]. Thus, a 3D point $\mathbf{p} = (x_h, y_h, z_h)$ in the human coordinate system is projected to 2D coordinates on the projection plane at a rate of d/z_h: $\mathbf{p}' = (x_h, y_h)d/z_h$. Therefore, the area of the target panel is also projected to the projection plane, affecting the visible area of the panel with respect to the rate of d/z_h.

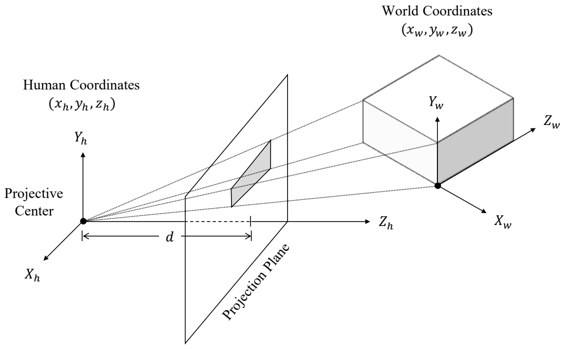

Figure 3. Pinhole projection model [27].

The top and side views of the experiment setup are depicted in Figure 4. $\mathbf{A}_0$ and $\mathbf{A}_1$ refers to the visible area of the panels projected on the projection plane, perpendicular to the gaze direction Z_0 and Z_1 (Here, we assume a person gazes at the center of the target panels when pointing). In both situations, $\mathbf{A}_0$ is larger than $\mathbf{A}_1$ due to the difference in between the angles θ_0 and θ_1, as well as the position of the panels. A smaller visible area hinders the subjects from pointing precisely while maintaining consistency. Thus, compared to the targets with comparatively large visible areas (wall and center panels), the performance degrades in the targets with small visible areas (ceiling and side panels). Overall, it can be noted that the visible area of the target is a crucial factor for human's ability of deictic gesture-based spatial referencing for both wall and ceiling conditions. Therefore, in practice, one can expect lower performance in referencing a distanced and angled regions of interest in overhead operations (e.g., electrical wiring, plumbing, and interior finishing work). In such situations, collaborative robots need mobility for estimation of the workspace geometry through navigating themselves closer to the target.

Considering the results mentioned above, we see three ways to improve the current spatial referencing method for application in collaborative construction robots. First, we could give the robot dimensional and scaled spatial information with interaction modalities (i.e., speech). This allows the robot to reason about the region of interest with additional criteria, thus enhancing perception accuracy. Presenting the spatial information with a form of region could be considered as well. Deictics are often thought of as referring to an object but can also be used to refer to a region of space [8]. This method provides interpretability and predictability to the user intent and has a collateral benefit of correction. Lastly, as suggested by Medeiros et al. [20], visual feedback makes a difference in the accuracy of the pointing task. In particular, we could enhance human's ability to indicate the target with a smaller visible area by receiving visual feedback from robots.

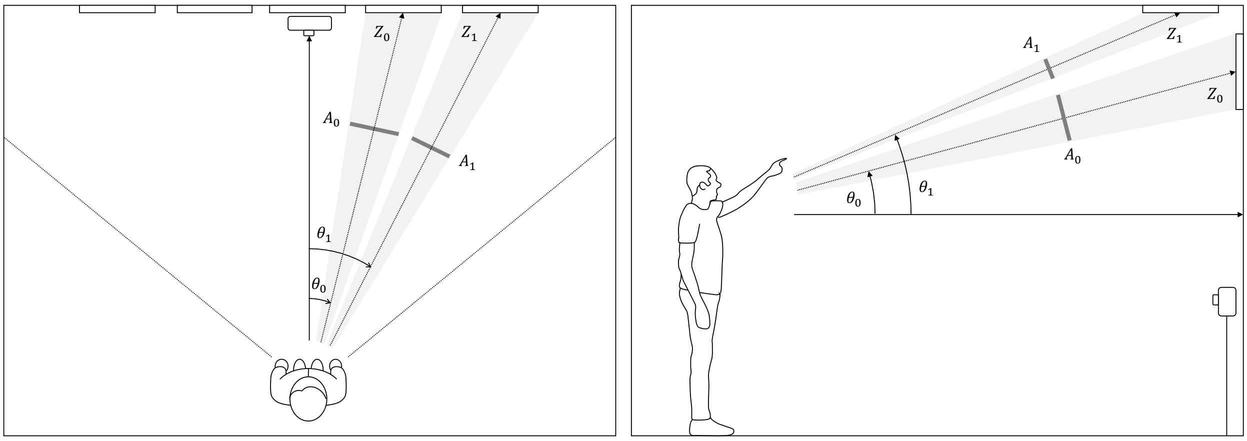

Figure 4. The top (left) and side (right) views of the experimental environment. The panels were spaced for a better understanding.

7 Conclusion

This work explored the challenges of the current spatial referencing method based on deictic gestures. It is challenging for a robot to reason about the exact region of interest in a large-scale 3D construction environment, cluttered in many situations. In this context, we selectively overviewed the performance of deictic gesture-based spatial information exchange in wall and ceiling with various angle conditions and discussed some solutions. Evaluation results show that the performance degrades in exchanging spatial information on the ceiling and side panels pertaining to the perception accuracy and the target recognition rate. The results also imply that the target's visible area is a crucial factor for human's ability of deictic gesture-based spatial information exchange for both wall and ceiling conditions. These findings can guide and inform our approaches to developing collaborative construction robots supported with an intuitive and natural human-robot interface.

In this work, we limited our focus on the performance evaluation to the robot's interpretation ability. Future work will include performance evaluation on the human's interpretation ability in a large-scale environment for a higher level of the collaborative environment such as shared autonomy. Furthermore, while we performed data processing and evaluation offline, we intend to further our research on on-site application of this method.

References

[1] L. Hines, K. Petersen, G.Z. Lum, M. Sitti, Soft Actuators for Small-Scale Robotics, Adv. Mater. 29 (2017) 1603483. https://doi.org/10.1002/adma.201603483.

[2] B. Gromov, L. Gambardella, A. Giusti, Guiding Quadrotor Landing with Pointing Gestures, in: F. Ferraguti, V. Villani, L. Sabattini, M. Bonfè (Eds.), Hum.-Friendly Robot. 2019, Springer International Publishing, Cham, 2020: pp. 1–14. https://doi.org/10.1007/978-3-030-42026-0_1.

[3] A.C.S. Medeiros, P. Ratsamee, Y. Uranishi, T. Mashita, H. Takemura, Human-Drone Interaction: Using Pointing Gesture to Define a Target Object, in: M. Kurosu (Ed.), Hum.-Comput. Interact. Multimodal Nat. Interact., Springer International Publishing, Cham, 2020: pp. 688–705. https://doi.org/10.1007/978-3-030-49062-1_48.

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

[5] E. Asadi, B. Li, I.-M. Chen, Pictobot: A Cooperative Painting Robot for Interior Finishing of Industrial Developments, IEEE Robot. Autom. Mag. 25 (2018) 82–94. https://doi.org/10.1109/MRA.2018.2816972.

[6] T. Obo, R. Kawabata, N. Kubota, Cooperative Human-Robot Interaction Based on Pointing Gesture in Informationally Structured Space, in: 2018 World Autom. Congr. WAC, 2018: pp. 1–5. https://doi.org/10.23919/WAC.2018.8430388.

[7] C.P. Quintero, R.T. Fomena, A. Shademan, N. Wolleb, T. Dick, M. Jagersand, SEPO: Selecting by pointing as an intuitive human-robot command interface, in: 2013 IEEE Int. Conf. Robot. Autom., 2013: pp. 1166–1171. https://doi.org/10.1109/ICRA.2013.6630719.

[8] A. Sauppé, B. Mutlu, Robot deictics: how gesture and context shape referential communication, in: Proc. 2014 ACMIEEE Int. Conf. Hum.-Robot Interact., Association for Computing Machinery, New York, NY, USA, 2014: pp. 342–349. https://doi.org/10.1145/2559636.2559657.

[9] S. Mayer, V. Schwind, R. Schweigert, N. Henze, The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments, in: Proc. 2018 CHI Conf. Hum. Factors Comput. Syst., Association for Computing Machinery, New York, NY, USA, 2018: pp. 1–13. https://doi.org/10.1145/3173574.3174227 (accessed July 28, 2021).

[10] M.W. Alibali, Gesture in Spatial Cognition: Expressing, Communicating, and Thinking About Spatial Information, Spat. Cogn. Comput. 5 (2005) 307–331. https://doi.org/10.1207/s15427633scc0504_2.

[11] G. Butterworth, N. Jarrett, What minds have in common is space : Spatial mechanisms serving joint visual attention in infancy, (1991). https://doi.org/10.1111/J.2044-835X.1991.TB00862.X.

[12] M. Tölgyessy, M. Dekan, F. Duchoň, J. Rodina, P. Hubinský, L. Chovanec, Foundations of Visual Linear Human–Robot Interaction via Pointing Gesture Navigation, Int. J. Soc. Robot. 9 (2017) 509–523. https://doi.org/10.1007/s12369-017-0408-9.

[13] A. Jevtić, A. Flores Valle, G. Alenyà, G. Chance, P. Caleb-Solly, S. Dogramadzi, C. Torras, Personalized Robot Assistant for Support in Dressing, IEEE Trans. Cogn. Dev. Syst. 11 (2019) 363–374. https://doi.org/10.1109/TCDS.2018.2817283.

[14] S. Mayer, J. Reinhardt, R. Schweigert, B. Jelke, V. Schwind, K. Wolf, N. Henze, Improving Humans' Ability to Interpret Deictic Gestures in Virtual

Reality, CHI. (2020). https://doi.org/10.1145/3313831.3376340.

[15] S.N. Medrano, M. Pfeiffer, C. Kray, Remote Deictic Communication: Simulating Deictic Pointing Gestures across Distances Using Electro Muscle Stimulation, Int. J. Human–Computer Interact. 36 (2020) 1867–1882. https://doi.org/10.1080/10447318.2020.1801171.

[16] J. DelPreto, D. Rus, Plug-and-Play Gesture Control Using Muscle and Motion Sensors, in: Proc. 2020 ACMIEEE Int. Conf. Hum.-Robot Interact., Association for Computing Machinery, New York, NY, USA, 2020: pp. 439–448. https://doi.org/10.1145/3319502.3374823 (accessed July 27, 2021).

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

[18] P. Kumar, J. Verma, S. Prasad, Hand Data Glove: A Wearable Real-Time Device for Human-Computer Interaction, Int. J. Adv. Sci. Technol. 43 (2012) 15–26.

[19] Y. Li, J. Huang, F. Tian, H.-A. Wang, G.-Z. Dai, Gesture interaction in virtual reality, Virtual Real. Intell. Hardw. 1 (2019) 84–112. https://doi.org/10.3724/SP.J.2096-5796.2018.0006.

[20] A.C.S. Medeiros, P. Ratsamee, J. Orlosky, Y. Uranishi, M. Higashida, H. Takemura, 3D pointing gestures as target selection tools: guiding monocular UAVs during window selection in an outdoor environment, ROBOMECH J. 8 (2021) 14. https://doi.org/10.1186/s40648-021-00200-w.

[21] C. Malerczyk, Interactive Museum Exhibit Using Pointing Gesture Recognition., in: 2004: pp. 165–172.

[22] K. Nickel, R. Stiefelhagen, Pointing gesture recognition based on 3D-tracking of face, hands and head orientation, in: Proc. 5th Int. Conf. Multimodal Interfaces, Association for Computing Machinery, New York, NY, USA, 2003: pp. 140–146. https://doi.org/10.1145/958432.958460.

[23] S. Abidi, M. Williams, B. Johnston, Human pointing as a robot directive, in: 2013 8th ACMIEEE Int. Conf. Hum.-Robot Interact. HRI, 2013: pp. 67–68. https://doi.org/10.1109/HRI.2013.6483504.

[24] S. Mayer, V. Schwind, R. Schweigert, N. Henze, The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments, Proc. 2018 CHI Conf. Hum. Factors Comput. Syst. (2018). https://doi.org/10.1145/3173574.

[25] Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields, in: 2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, 2017: pp. 1302–1310. https://doi.org/10.1109/CVPR.2017.143.

[26] D. Sprute, R. Rasch, A. Pörtner, S. Battermann, M. König, Gesture-Based Object Localization for Robot Applications in Intelligent Environments, in: 2018 14th Int. Conf. Intell. Environ. IE, 2018: pp. 48–55. https://doi.org/10.1109/IE.2018.00015.

[27] A. Sharma, R. Nett, J. Ventura, Unsupervised Learning of Depth and Ego-Motion from Cylindrical Panoramic Video with Applications for Virtual Reality, 2020.