Exploring the Visual Space to Improve Depth Perception in Robot Teleoperation Using Augmented Reality: The Role of Distance and Target's Pose in Time, Success, and Certainty

Stephanie Arévalo Arboleda, Tim Dierks, Franziska Rücker, Jens Gerken

Exploring the Visual Space to Improve Depth Perception in Robot Teleoperation using Augmented Reality: The Role of Distance and Target's Pose in Time, Success, and Certainty

Stephanie Arévalo Arboleda[0000−0001−5577−4407], Tim Dierks, Franziska Rücker, and Jens Gerken

Westphalian University of Applied Sciences, Gelsenkirchen, Germany
{stephanie.arevalo,tim.dierks, franziska.ruecker, jens.gerken}@w-hs.de

Abstract. Accurate depth perception in co-located teleoperation has the potential to improve task performance in manipulation and grasping tasks. We thus explore the operator's visual space and design visual cues using augmented reality. Our goal is to facilitate the positioning of the gripper above a target object before attempting to grasp it. The designs we propose include a virtual circle (Circle), virtual extensions (Extensions) from the gripper's fingers, and a color matching design using a real colormap with matching colored virtual circles (Colors). We conducted an experiment to evaluate these designs and the influence of distance from the operator to the workspace and the target object's pose. We report on time, success, and perceived certainty in a grasping task. Our results show that a shorter distance leads to higher success, faster grasping time, and higher certainty. Concerning the target object's pose, a clear pose leads to higher success and certainty but interestingly slower task times. Regarding the design of cues, our results reveal that the simplicity of the Circle cue leads to the highest success and outperforms the most complex cue Colors also for task time, while the level of certainty seems to be depending more on the distance than the type of cue. We consider that our results can serve as an initial analysis to further explore these factors both when designing to improve depth perception and within the context of co-located teleoperation.

Keywords: Human-robot interaction · depth perception · augmented reality · robot teleoperation · visual cues · certainty.

1 Introduction

One of the most common tasks in Human-Robot Interaction (HRI) is pick-and-place since it presents a basic interaction for teleoperation of robotic arms. These tasks are common in non-standardized assembly workspaces, where target objects often change in shape and position, which may require the human operator to fine-control (semi-autonomous) robotic arms to succeed in the tasks.

Picking an object can be further segmented into **1) (Coarse) Pointing**, which requires the user to identify the location of a target object and consequently move the gripper of the robotic arm roughly to that location; **2) Positioning**, which requires to align the gripper with the target object to successfully grasp it; and **3) Grasping** then requires the user to move the gripper to the point in space where the object can finally be acquired. While this, at first sight, might sound straightforward, it often requires many trial-and-error attempts from the operator due to misjudgments regarding the position in space of either the gripper or the object. This can lead to precarious situations when handling hazardous material.

The crucial point in this process takes place while positioning the gripper. It might not always be possible for the operator to accurately determine the gripper's position relative to the target object due to the visual perception of distance and object's pose, both factors related to depth perception. Even in co-located scenarios (robot and operator located in the same physical space) operators could have a fixed viewpoint due to physical mobility restrictions, e.g., people with disabilities, or due to environmental limitations, e.g., safety measures due to the manipulation of hazardous materials. As a result, spatial abilities are impeded, hindering the operator to correctly identify the shape, pose, and distance of a target object relative to the gripper and other objects within the workspace.

Augmented Reality (AR) creates opportunities to improve depth perception of real-world objects by enhancing the visual space to provide awareness of their position in the workspace. Some studies present evidence that AR counteracts visual feedback limitations in teleoperation [39], [53],[19]. Some of these limitations are related to the manner in which humans perceive distance and judge depth. Humans perceive visually the environment through different visual cues (monocular, binocular, dynamic) [14]. These cues can be enhanced through AR and thus refine the visual space. Here, it is important that these cues provide a message that can be clearly interpreted by the observer, avoiding unnecessary extra mental processing due to conflicts among cues [29].

In this paper, we contribute to the exploration of the design space to improve depth perception. Therefore, we first propose three designs of visual cues in AR that systematically build on each other. Each one considers different design elements that could foster depth perception in co-located robot teleoperation and provide a foundation for future HRI designs. Second, we present a systematic experiment investigating the relationship between the design of visual cues, the operator's distance, and the target object pose to evaluate the advantages and downsides of each visualization concept. Third, we include the operator's perceived certainty during picking as a novel measure to better understand if and how objective measures of performance such as success (certainty accuracy or effectiveness) and time (efficiency) correlate with the subjective certainty of the operator.

2 Background

We first review theoretical foundations of visual perception regarding distance and depth, the use of AR to improve depth perception in robot teleoperation, and the role of time, success, and certainty in visual perception.

2.1 Visual Perception of Depth

Visual perception can be defined as a complex representation of the visual world, where features of objects in the environment are collected visually to then be interpreted through computations in multiple areas of the brain [49]. In order to visually determine the depth of objects, factors such as size, shape, and distance are determinants to gain an accurate perception of objects in the environment [24]. Brenner & Smeets [9] present the relation between size and distance as straightforward, e.g., smaller sizes provide hints about how distant an object might be, while shape and distance present a more complex relation. Distance can compromise determining the shape of objects, e.g., problems in accurately determining the shape of far-away objects. Further, perceived objects' shape varies depending on the manner that they are positioned (pose). Here, having different perspectives acquired by motion assists with accurately determining an object's shape.

Within the visual space, it is relevant to explain the perceived location. This is understood as the perception of direction and distance from the observer's viewpoint (egocentric distance) or the distance between two external points (exocentric distance) [34]. Related to perceived location, the space layout of a person is determinant for evaluating distance and depth as farther distances compromise the manner how objects in the surroundings are perceived. Cutting& Vishton [12], divide the layout of space around a person into three: personal space (within 2m), typically a stationary working space; action space (from 2m to 30m), which is the moving and public action space; and vista space ($>$ 30m). Different authors agree that humans underestimate egocentric distances (distance relative to the observer) in the real-world [33],[34]. Distance can be evaluated through different measures, e.g., verbal reports, walking around a target, visually matching distances to a familiar one, and blind walking [47]. It can also be estimated through bisection or fractionation, which consists of determining the midpoint of a distance from the observer's perspective to a target, and specify that bisection does not provide absolute measures of egocentric distances [7].

Humans perceive and interpret distance through a set of visual cues: monocular (perceived using one eye), binocular (perceived using both eyes), and dynamic (perceived by movement) [14]. Nonetheless, interpreting all the cues perceived under some environmental conditions is not a straightforward process. Laramee & Ware [31] describe it as an ambiguity solving process, where different cues need to be primed over others to get an accurate picture of the environment (cue dominance). Further, Howards & Rogers [25] highlight that the reliability of cues is determined by cue average, cue dominance, cue specialization, range extension, and probabilistic models.

Rolland et al. [46] already found a relation between object size and distance in virtual objects displayed in ARHMD. In Mixed Reality (MR) environments, this problem can be not only inherited from the real-world but can be exacerbated [26], [40]. Further, El Jaimy & Marsh [14] present a survey on depth perception in head-mounted displays (HMDs) and highlight the importance of evaluating not only depth but also distance perception in virtual and augmented reality environments.

2.2 Designing to improve depth perception in AR Environments

We consider the work of Park & Ha [41] as a keystone to improving depth perception in virtual and mixed reality environments. They provide a classification of visual enhancements techniques that offer spatial information as follows: geometric scaling, understood as the variation of size to provide a distance relationship; symbolic enhancements, which are visual representations that allow creating associations, e.g., a grid surface or ground plane to transfer spatial information; visual cues, which are a combination of monocular cues to improve the perception of depth; a frame of reference that provides a mental model of the environment; and visual momentum, referring to providing perceptual landmarks to reduce visual inconsistency among different displays/scenes. Following this line of research, Cipiloglu et al. [11] grouped a series of methods for depth enhancement in 3D scenes focusing on perspectives, focus, shading and shadows, among others.

Heinrich et al. [21] provided a state-of-the-art overview of visualization techniques using AR for depth perception. They presented different visualization concepts that have been applied to improve depth perception with an emphasis on projective AR without using HMDs. Also, Diaz et al. [13] explored different factors that influence depth perception in AR such as aerial perspective, cast shadows, shading, billboarding, dimensionality, texture, and the interaction of cues. They presented two experiments where they evaluated the effect of these factors in participants' perception of virtual objects' depth relative to real targets. Their results showed that among all their designs, the use of cast shadows improved depth estimation the most. This aligns with other studies [56] that used casting shadows through AR as depth cues. They used pictorial cues (color encoded markers) to provide information about egocentric distance together with shadows and aerial perspective, showing similar results in distance perception.

When aiming to enhance the visual space through AR the work of Kruijff et al. [29] is of particular importance. They identified and classified a series of factors that affect the augmentations related to the environment, capturing, augmentations, display devices, and users. These provide a framework to identify the potential factors that influence depth perception.

2.3 AR for Depth perception in Robot Teleoperation

Previous studies present evidence that AR diminishes visual feedback limitations in teleoperation by providing additional information to the operator [39],

[52],[19]. Presenting cues through AR could enhance the perception of depth and distance in the real world. Choi et al. [10] present reinforcement of the user's cognitive abilities as the purpose of AR. Moreover, the use of AR for robot control has been found to reduce the mental load in robot programmers [48].

Depth perception is still an issue in MR and teleoperation [10] that invites further exploration. Casting shadows using AR has been used in the teleoperation of aerial robots [60], [53] since it has proven to support aerial navigation by improving the spatial relationships between the environment and the aerial robot. Zollman et al. [60] used different techniques that aim to maintain or replicate natural depth cues to design visual hints (waypoints and pathlists) that provide flight-relevant information. In pick-and-place tasks, AR has also been used to present a projection-based AR interface that provides task instructions [18]. Here, shadows have been used to highlight intended target positions and thus provide instructions to operators.

2.4 Certainty, Time, and Accuracy in Visual Perception

Certainty can be defined as a sense of conviction and is considered as a foundation of people's beliefs [42]. In order to achieve "good visual certainty", observers need to consider information that goes beyond monocular, binocular, or motion parallax cues and acknowledge sources of uncertainty, being thus able to predict an outcome [35]. In order to measure the degree of certainty, people need information derived from evidence and time, wherein evidence, in turn, contributes to accuracy [27].

Evidence can be acquired from experience, where certainty plays a pre-and-post-decision-confidence role. Pre-decision confidence relates to the current incoming information and post-decision confidence is the information derived from experience [20]. Also, evidence can be acquired through perceptual evaluation, wherein visual perception has a relevant role. Here, findings of [17] show that the amount of evidence has a positive correlation with accuracy.

Time plays a significant role in relation to attitude certainty. This relation has received special attention in neuroscience. Willis & Todorov [55] shown that a longer period of time allows to form greater impressions of certainty, yet it does not necessarily improve the impression of accuracy. In line with this view, Barden & Petty [4] provided evidence that greater thoughtfulness leads to greater certainty. However, there are controversial results about the influence of time on certainty. Some studies [3],[2] associate longer times with lower certainty. Further, Kiani et al. [27] showed that time alone cannot explain fluctuations in certainty but time together with difficulty have a critical role in certainty.

In visual perception, accuracy is bound to the ability to make a good judgment of the visual stimulus, while certainty relates to the ability to make a good judgment on the validity of a perceptual decision [35]. Accuracy has been suggested to be connected to the amount of evidence from the environment that can be collected [44]. Perceived visual certainty can be dissociated from accuracy [35], and adding to that, a line of research suggests that time can also

be disassociated from accuracy [27]. Visual certainty is a topic with controversial findings that has several factors that influence it. Barthelme & Mamassian [5] found that visual uncertainty predicts objective uncertainty. Interestingly, Gardelle & Mamassian [16] showed that subjective uncertainty is abstract and task-independent. They compared identical and perceptual different task-trials in succession in a visual discrimination task, e.g. the orientation of a bar, and found no difference between conditions.

3 Exploration of Visual Cues for Co-located Teleoperation

Building on the research mentioned in Sections 2.2 and 2.3, we present three designs of visual cues to help operators to evaluate distance and depth which in turn could allow to better estimate the position of the gripper relative to a target object for successful grasping. In particular, our designs of visual cues capitalize on previous findings of the effectiveness of cast and drop shadows [13], symbolic enhancements [41], and matching physical and virtual landmarks [60] to provide better awareness of the position of the gripper on the workspace. While some related work has focused on improving depth perception of virtual objects and their interaction in the real-world, we focus our designs on using virtual elements to improve the depth perception of real-world objects. Consequently, all our designs augment the environment through AR using the Microsoft HoloLens. As can be depicted in Fig. 1, each of our designs builds upon the previous one. Thereby, the Extensions include the Circle visual cue and the Colors include both, the Extensions and Circle.

3.1 Virtual Circle using Cast Shadows (Circle)

In this design, we provide a virtual cue derived from the real-world (the gripper) which in turn acts on the physical world (workspace). We cast the real grip region as a virtual representation of it through a virtual circle. This virtual circle's diameter matches the width of the grip region and is shown right under the real gripper on the workspace, see Fig. 1a. We base our design on previous findings of the efficacy of cast shadows to improve depth perception [13]. Cast shadows can be described as the shadow of an object that is reflected on a different surface [59]. This first design of visual cue is minimalistic and intends to provide a simple and comprehensible hint of the location of the gripper on the workspace.

3.2 Virtual Extensions as a Symbolic Enhancement (Extensions)

In this design, in addition to the Circle, we virtually extend an object from the physical world (the gripper). Based on previous work about symbolic enhancements [41] and considering Walker et al.'s framework for AR in HRI [52], we augment the robot's end-effector, through a virtual elongation of the gripper's

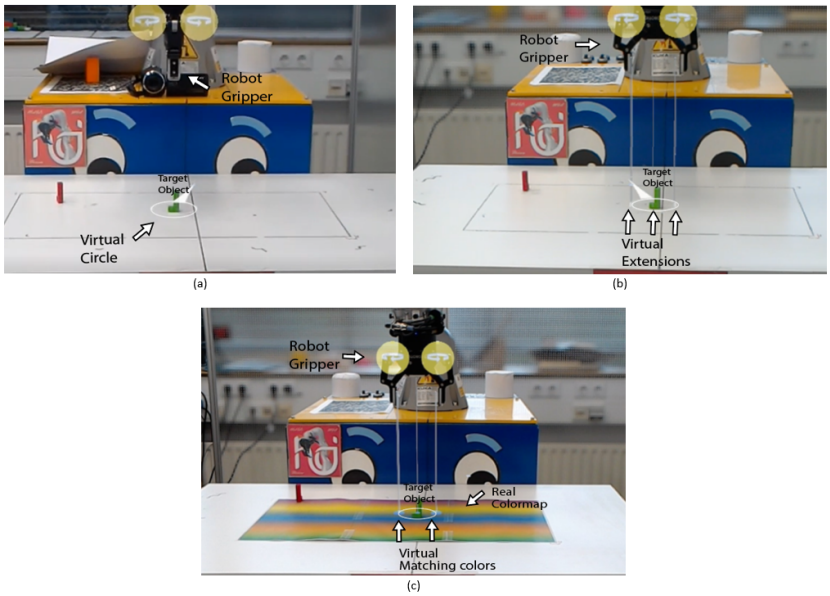

Fig. 1. (a) Circle. (b) Extensions. (c) Colors: Physical color map with matching virtual circles at the end of the Extensions.

fingers. We present three virtual elongations, one at the tip of each finger, and another one rendered at the center of the grip region, see Fig. 1b. These virtual extensions are designed to allow for a better visualization of the exact gripper position on the workspace. Through these visualizations, the operator can determine if the target object is within reach or if there are any potential collisions, e.g., between the fingers and some surface of the target object or other elements in the environment.

3.3 Mixed Reality Color Gradient using Physical Landmarks (Colors)

Building on the Circle and Extensions, the Colors cues aim to provide a connection to the real-world by matching physical landmarks with virtual ones. This could potentially improve the mapping and alignment between the augmentation and the real-world objects as it provides more information about the spatial arrangement of the environment. Here, we take into account the work of Zollman et al. [60], who explored this connection in flying aerial robots. We considered these findings and applied them to our scenario by providing a real physical landmark that connect the gripper and the workspace.

Our landmark is a physical colormap. It presents 5 different color gradients (6 cm per stripe) that cover the workspace, signaling incremental depth, see Fig. 1c. Our reasoning behind the use of a colormap comes from the problems

that have been found of determining with precision a position in monochromatic surfaces, e.g. previous research showed that the human eye loses depth cues on uniform surfaces [29]. Further, the use of colors has proven to be effective to signal depth [56]. Through a colormap, the operator can identify a color in the workspace where a target object is located and then position the gripper with greater precision above it. To facilitate pointing at the target area of the workspace, the cursor also adopts the color of the area that is being pointed at. Additionally, we added small colored circles at the end of the virtual extensions, which adopt the color of the current area.

3.4 Interaction Design for Teleoperation

This research is part of a larger project that takes a closer look at hands-free multimodal interaction. Our interaction design consists of head orientation to point, head yaw for positioning, and voice commands to commit an action. Previous studies have shown that the use of speech and head movements are an intuitive interaction concept for human-robot collaboration in pick-and-place tasks [30]. Further, these modalities are natively supported by the HoloLens, and are common for MR environments. The initial and resting position of the robotic arm can be seen in Fig. 1 and the interaction is explained as follows:

Pointing. The operator sees a white head pointer which moves with the operator's head movements. Once an intended position has been determined, the operator gazes at that position and commits the action with the command "Move". After the command has been recognized, the head pointer turns green in a "pie timer" manner for one second. During this time, the operator can decide to keep (remain still) or modify (change head position) the pointer's position before the gripper starts moving to the selected position on the x,y plane. This helps to counter slight inaccuracies that can result from unintended head movements which can happen while uttering the voice command.

Positioning. Once the gripper is located over the desired position. We placed a set of virtual buttons anchored to the real gripper. These virtual buttons rotate the gripper to the left or right and allow the operator to ensure that the fingers' grasping points match the affordances of the object, see Fig. 1. In addition, the operator can activate a fine control mode through the command "Precision". This fine control is executed through the operator's head yaw, where the gripper moves slowly following the head's yaw. In this mode, we display (on the edges of the circle) an arrow to indicate the direction that the gripper is moving towards.

Grasping. When the operator has determined that the gripper is located at an adequate grasping position, the command "Pick" commits the action of grasping, i.e., the gripper moves with the fingers opened along the z-axis towards the object, stops at 0.5 cm above the tabletop, closes the fingers, and moves back up to the resting position.

4 Study

We present a study with 24 participants that aims at exploring the relationship between the different designs of cues (Circle, Extensions, Colors) and depth-related variables distance (2m and 3m) and object pose (clear, ambiguous) when teleoperating a co-located robotic arm. In particular, we looked at measures of effectiveness (success), efficiency (time), and perceived certainty. In the study, participants wore the Microsoft HoloLens to visualize our designs of visual cues while teleoperating a robotic arm.

4.1 Hypotheses

Distance perception of the bare eye has an imminent effect on depth perception. Based on this, we hypothesize that our designs of visual cues will reduce the impact of distance (H1). Similarly, objects' pose can reduce depth perception, we thus hypothesize that our visual cues will reduce the impact of pose (H2). Based on the fact that the individual designs of each visual cue builds systematically on each other, subsequently adding more depth information, we expect a step-wise rise between each of them. This means that the performance with Circle will be exceeded by Extensions and Extensions by Colors for the evaluated metrics (H3). Further, building upon visual perception and certainty, we designed our visual cues to provide more information (evidence) about the workspace, which will in turn show a correlation between these two metrics (H4). Specifically, we hypothesize that:

H1. Our designs of visual cues will lead to similar results in grasping time, success, and certainty at 2m and 3m.

H2. Our design of visual cues will leading to similar results in success and time but a higher degree of certainty for the clear pose.

H3. Colors will lead to a higher success rate, shorter execution time, and higher perceived certainty compared to Extensions and Circle. Similarly, the Extensions would perform better on those metrics compared to the Circle alone.

H4. Our results will point to a positive correlation between certainty and success rate.

4.2 Participants

We recruited 24 participants among students and university staff with an average age of 28.67 (SD = 7.72). The pre-test questionnaire revealed that 6 participants had previous experience using the Microsoft HoloLens 1 and 11 participants had some experience with robots (not necessarily a robotic arm). One participant reported having red-green colorblindness but did not experience problems in distinguishing the colors used. Participants received 7 euros for their participation.

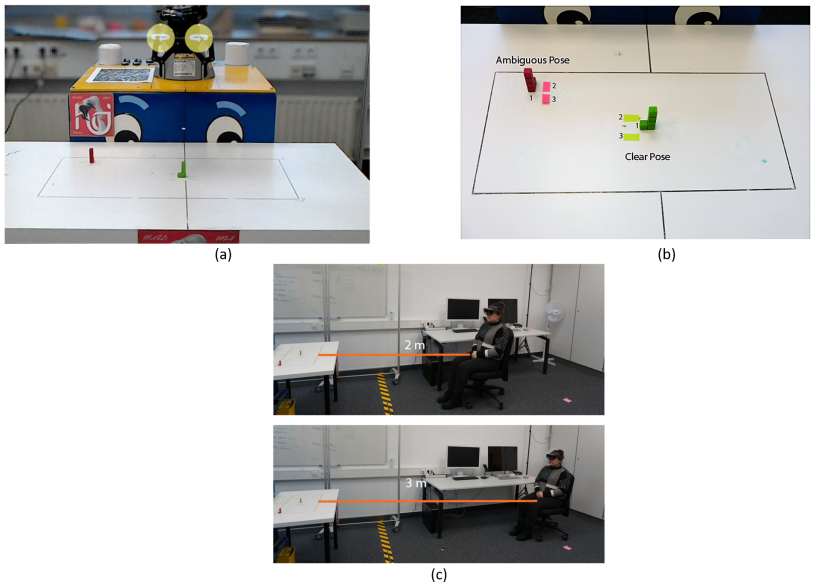

Fig. 2. (a) Experiment set-up for Circle and Extensions. (b) Upper view from the workspace with the target objects. (c) An operator located at different distances from the workspace.

4.3 Study Design

We designed a full factorial 2x2x3 within-subject experiment (2 distances x 2 poses x 3 designs of visual cues) where participants executed 3 trials per condition, yielding 36 trials per participant and 864 trials in total. The conditions were counter-balanced through a Latin-square design.

Our independent variables were the type of visual cue (Circle, Extensions, Colors), distance (2m, 3m), and pose (clear, ambiguous). For distance, we varied the egocentric space between the workspace and the stationary position of the operator (Fig. 2c). We chose 2m and 3m as distances to represent realistic co-located scenarios covering the egocentric personal (2m) and action space (3m) respectively. We did not vary the size of the object due to kinetic invariance––when an object varies in size it affects how the observer perceives the size and distance of the object [22].

Different poses were achieved by manipulating the position and orientation of an L-shaped target object. To differentiate the poses, we used two objects of the exact same size and shape but with different colors. A green object was deemed as the "clear" pose. It was located roughly in the center of the workspace and oriented in such a way that the shape was fully visible. A red object was deemed as the "ambiguous" pose. It was located in the left corner, almost at the border of the workspace and oriented in such a way that participants could only

see one side of the object. We considered this pose ambiguous since it required participants to perform mental rotations (see Fig. 2).

We collected objective and subjective measures. As objective measures, we considered time, measured in seconds, and success, measured as a binary value. Our measure of time relates to the amount of time that participants took to position and align the gripper above the target object until invoking the picking operation—referring to positioning, the second step of the picking interaction. For each trial, we measured whether or not participants succeeded in grasping the target object (success). As a subjective measure, we prompted participants with the following question directly in the HoloLens whenever they invoked the picking operation: "In this position, how certain (in %) are you that you can grasp the object". The percentages were provided as choices on a 7-point scale. Tormala [50] argues for a 7-point scale to measure attitude certainty, as in certain circumstances extreme attitudes can be held with less certainty. Also, qualitative data were collected through a short interview at the end of the study.

We used a mixed-methods approach to evaluate our data. The data were analyzed using RM-ANOVA for time and certainty, as a normal distribution of data could be assumed through the means of Shapiro-Wilk test. For pairwise comparisons (through post-hoc Estimated Marginal Means), we applied Bonferroni corrections to control for Type I errors. For the analysis of the binary variable grasping success, we could not assume a normal distribution and therefore opted for GEE (Generalized Estimating Equations). GEE accounts for correlations within-participants in repeated measures designs and has been commonly used for binary outcome variables [57], [32]. To investigate potential associations between our dependent variables (certainty and success/time), we applied Spearman's partial correlation controlling for the effect of our independent variables.

4.4 Task

Our main task simulated a grasping task inspired from manufacturing workspaces. The task comprised of teleoperating the robotic arm to grasp an "L-shaped" object (4x1x2)cm placed at two different positions and with the different poses on the workspace. The operator performed this task at 2m and 3m from the workspace, see Fig. 2c. Participants were instructed to remain seated in a comfortable position and avoid movements to the left or right or tilt their heads to change their visual perspective. This instruction helped to evaluate the effects of distance and pose using our designs of visual cues on depth perception.

We performed a pilot test with 3 users to identify potential factors that may affect our experiment. We discovered learning effects due to the position invariance of the target objects during trials. Thus, we moved each object 1 cm apart from the last position for each trial for the experiment, see Fig. 2b.

4.5 Procedure

The experiment took approximately 60 minutes divided into (1) introduction, (2) calibration, (3) training (4) task, and (5) post-test questionnaires and interview.

(1) During the introduction participants were handed a standardized consent form and pre-test questionnaire. They were briefed about the devices, the objects that they will be interacting with, and the interaction modalities. Also, we showed an explanatory video depicting the interaction techniques, interface, and visual cues. (2) Next, participants calibrated the HoloLens, where the interpupillary distance was recorded for each participant. (3) Then, they proceeded to perform a training task, consisting of teleoperating an industrial robot arm to pick an L-shaped object, similar to the one to be used in the real task but bigger in size, at 1m from the robot. (4) Following, participants executed the experiment. Half of the participants started the task at 2m and after finishing all designs of cues and pose combinations repeated the same procedure at 3m. The other half started at 3m and then at 2m. For pose and designs of cues, we followed a Latin square distribution. (5) Finally, we carried out a short interview about their experiences.

4.6 Apparatus & Communication

We used a Kuka iiwa 7 R800 lightweight robotic arm with an attached Robotiq adaptive 2-Finger Gripper. Both are controlled via the Kuka iiwa's control unit. For our visual cues and the user interface, we used the Microsoft HoloLens 1, which is equipped with inside-out tracking, an HD video camera, and microphones allowing the use of head movement, speech, and gestures as input options [36]. The HoloLens has a field of view (FOV) of $30^\circ \times 17.5^\circ$ with a resolution of 1268×720 px per eye. The application running on the HoloLens was programmed with Unity 2018.1.2 and the HoloToolkit Plugin [38].

The Kuka iiwa and the HoloLens communicate via the User Datagram Protocol in a local network. The control unit of the Kuka iiwa runs a specially designed back-end program that processes the received messages, moves the robot according to the receiving data, and returns its current status. To communicate position data between devices, we first converted the pose from Unity's left-handed coordinate system to the robot's right-handed one and used common length units (cm). Then, this cartesian position, rotation and velocity is sent to the control unit. This recognizes the command and allows the robot to plan the movement via internal inverse kinematics to then physically move the robotic arm.

A calibration process is important to achieve accuracy when using AR. The HoloLens adjust the hologram display according to the interpupillary distance. When it is not accurate, holograms may appear unstable or at an incorrect distance [37]. Thus, we ran the Microsoft calibration application for each user. Also, in order for the virtual tracking to be transferred to the real world, the position and alignment of the robot must be known in the virtual-world. For this, we use a 2D marker and the camera-based marker detection from Vuforia [45] when starting the application. Since the position of the 2D marker in relation to the robot is known, the position of the robot in the virtual'world can be determined and set as world anchors. The stability of world anchors has been previously evaluated and found a mean displacement error of 5.83 ± 0.51mm [51]. This is precise enough to handle grasping tasks.

5 Results

5.1 Objective Measures

Time (Efficiency). Analyzing the time spent to perform the tasks, we opted for a RM-ANOVA with the design of cues, distance, and object pose, as independent variables and time as a dependent variable. For time, we used the average across the three trials for each condition. Testing the assumptions with Shapiro-Wilk, the distribution of some residuals showed a slight deviation from the normal distribution. The inspection of QQ-plots as well as skewness and kurtosis analysis, revealed that all residuals in question were skewed in the same positive direction and within an acceptable range [28]. In addition, the literature [6] suggests that a slight deviation from a normal distribution can be handled by ANOVA procedures, which is why we kept this approach. For pairwise comparisons, we refer to the estimated marginal means and report p-values and standard errors.

Results show that there were no significant three-way ($F(2, 46)=0.225$, $p= .8$) or two-way interaction effects for our independent variables (design of cues per distance $F(2, 46)=2.31$, $p= .11$; pose per distance $F(1, 23)=0.979$, $p= .333$; pose per design of cues, Greenhouse-Geiser corrected due to sphericity violation $F(1.562, 35.93)=2.036$, $p= .154$).

We found a significant main effect for each of our independent variables. For distance ($F(1, 23)=7.124$, $p = .014$, $\eta p^2= .236$), grasping objects at 2m (M=26.25s, SE=1.67) was significantly faster than at 3m (M=30.95s, SE=2.12), Fig. 3b. For pose ($F(1, 23)=7.145$, $p= .014$, $\eta p^2= .237$) participants completed the ambiguous pose (M=26.46s, SE=1.94) significantly faster than the clear pose (M=30.74s, SE=1.79), Fig. 3d. Finally, we also found significant differences in grasping times for the three design of cues ($F(2, 46)=13.029$, $p < .001$, $\eta p^2=.362$). Post-hoc pairwise comparisons (Bonferroni adjusted) showed significant differences for Colors (M=33.14s, SE=1.78) compared to Circle (M=26.12s, SE=2.04, p=.001) and Extensions (M=26.54s, SE=1.91, $p= .003$).

Success (Efficacy). For the dichotomous variable success, we applied a GEE model for which we used the GENLIN procedure in SPSS. As the working correlation matrix, we applied an exchangeable structure, and we used a binary logistic response model. For pairwise comparisons, we refer to the estimated marginal means and report p-values and standard errors.

Our results show no significant three-way or two-way interaction effects for our independent variables (design of cues per distance Wald $\chi^2(2, N=864) = 3.74$, $p= .154$; distance per pose Wald $\chi^2(1, N=864) = 3.59$, $p= .058$; pose per design of cues Wald $\chi^2(2, N=864) = 3.59$, $p= .166$).

Again, we found significant main effects for each of our independent variables. For distance (Wald $\chi^2(1, N=864) = 28.35$ $p < .001$), grasping objects at 2m (M=0.71, SD=0.46) was significantly more successful compared to 3m (M=0.43, SD=0.495). For pose (Wald $\chi^2(1, N=864)=11.526$, $p= .001$), the clear pose (M=0.66, SD=0.475) shows a significantly higher success compared to the ambiguous pose (M=0.48, SD=0.5). Finally, we also see a significant difference

between the design of cues (Wald $\chi^2(2, N=864) = 42.96$, $p < .001$). Post-hoc pairwise comparisons (Bonferroni adjusted) revealed that Circle (M=0.72, SE=0.03) showed a significantly higher success compared to both Extensions (M=0.58, SE=0.05, p= .004) and Colors (M=0.45, SE=0.04, $p < .001$). Also, there is a significant difference between Extensions and Colors (p= .02).

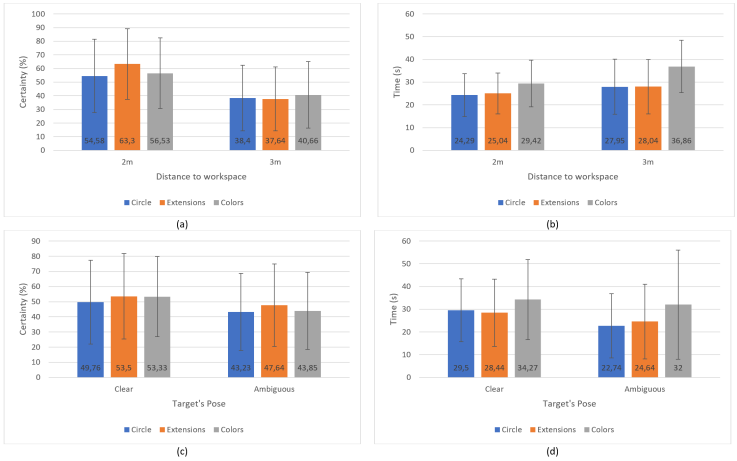

Fig. 3. Summary of M and SD with each design of visual cue (a) Shows certainty vs distance. (b) Shows time in seconds vs distance. (c) Shows certainty vs pose. (d) Shows time vs pose.

5.2 Subjective Measures

Certainty. We calculated the mean certainty rating across the three trials for each participant in each condition. Shapiro-Wilk tests and inspection of QQ plots showed that for the residuals normality could be assumed, we thus applied a RM-ANOVA. We found no significant three-way interaction ($F(2, 46)=0.65$, p= .94) but a significant two-way interaction for distance per design of cues ($F(2, 46)=4.01$, p= .025). Looking at the simple main effects for this interaction (post-hoc Estimated Marginal Means, Bonferroni adjusted), we found that for the 2m distance, the Extensions led to significantly higher certainty (M=63.30%, SE=4.27) compared to both Colors (M=56.53%, SE=4.11, p= .01) and Circle (M=54.58%, SE=4.83, p= .015). However, this is not the case for the 3m distance, where Extensions reached the lowest perceived certainty (Extensions M=37.64%, SE=4.06; Colors M=40.66%, SE=4.32; Circle M=38.40%, SE=4.01; differences not significant). We found a significant main effect both for pose ($F(1, 23)=7.41$, p= .012, $\eta p^2 = .24$) and distance ($F(1, 23)=32.96$, $p < .01$, $\eta p^2 = .589$). For pose, the clear pose led to a higher certainty (M=52.13%, SD=27.42) compared to the ambiguous pose (M=44.91%, SD=26.02), Fig. 3c.

For distance, the closer distance of 2m led to a higher certainty (M=58.14%, SD=26.43) compared to 3m (M=38.90%, SD=23.89), Fig. 3a.

To understand the relationship between perceived certainty and measures success and time we calculated a Spearman Partial Correlation, controlling for design of cues, distance, and pose. Results show that there is a significant positive partial correlation between certainty and success ($rs(859)=0.144$, $p < .001$) but not between certainty and time ($rs(859)=0.021$, $p= .53$).

6 Discussion

First **(H1)**, we hypothesized that our design of visual cues would perform similarly regardless of the different egocentric distances, improving hence depth perception. Our results could not support this hypothesis. A smaller distance prompted better results in terms of time, certainty, and success compared to 3m. This aligns with previous findings of depth perception decaying at greater distances in the real-world and in AR [43]. Further, Microsoft recommends a distance of 2m for an MR environment when using the HoloLens 1 as further distances can induce perceptual problems. Specifically, problems derived from capturing (flares and calibration) that can provoke scene distortion and problems in environment abstraction [29]. This might have influenced the results and the experience of using our visual cues at 3m. Additionally, all participants mentioned that performing the grasping task at 3m was harder than at 2m, e.g., P3, "The farther distance was exponentially more difficult." P19, "The farther away, the harder it is to understand depth independently of the virtual supporters." These results lead us to think that teleoperation in co-located spaces should be performed within the personal space (up to 2m) for better depth perception and thus performance. Despite that, we highlight the importance of further exploring distances beyond 2m to better understand the dynamics of teleoperation within the operator's action space.

Second **(H2)**, we hypothesized that our design of visual cues would reduce the effect of the target's pose, which would be reflected in similar results of success and time. However, we expected a higher degree of certainty for the clear pose compared to the ambiguous pose. Our results partially support this hypothesis. The effect of pose proved to be still present in spite of our designs of visual cues. Our results show that the clear pose prompts indeed higher certainty but also a significantly higher success with a compromise in terms of time (longer time) compared to the ambiguous pose. These results align with the findings of Barden & Petty [4] in terms of a direct relation between certainty and time. During our interview, we asked participants if they deemed one pose harder than the other for grasping, while most of them found the ambiguous pose harder, many also stated that they did not find differences among them, which could possibly be an effect of our visual cues. For instance, P7, "Both objects were equally hard," P9, "There was no difference between the 2 objects," "I did not find any object harder than another." Another factor that might have influenced these results is the type of task. Grasping tasks in a workspace with no other objects in vicinity

that partially or completely occlude the target objects, may not highlight the potential benefits of our designs of cues.

Third **(H3)**, we hypothesized about potential differences between our design of cues with respect to time, success, and certainty. We assumed that Colors would perform best, followed by Extensions, and then Circle. Our results indeed showed that our designs of visual cues had different effects on our dependent variables, but the effects were different from what we hypothesized.

Regarding time, we found that our participants were faster using both the Circle and Extensions compared to using Colors. Additionally, no significant differences were found with respect to time between Circle and Extensions. When analyzing success, we found higher success rates when using Circle compared to using Extensions and Colors. Also, using Extensions showed a higher significant success compared to Colors.

Concerning certainty, we found a significant interaction effect for distance per design of cues. These point to differences at 2m, where Extensions prompted the highest level of certainty, which in turn was significantly higher compared to both Circle and Colors. Since our designs were incremental, Extensions, which included the Circle, seemed to provide the necessary information about depth without the visual complexity of Colors. The absence of this effect at 3m might be due to the fact that the effect of distance alone overshadowed any potential difference among our designs of cues. While Audley [2] remarked that the amount of information (evidence) that can be collected from the environment influences the amount of perceived certainty, e.g., the more evidence that is collected, the higher certainty that is evoked. Our results show that one must be careful when designing to provide more information about the environment. While the Colors certainly provided the most depth cues, our results point towards the fact that this information was not always usable to our participants. This is also reflected by our participants' comments, who mostly pointed out the simpler cues, Circle and Extensions, and deemed them as helpful: P1, "The visual extensions were good for the further distance." P10, "The most helpful thing was the circle. I used it the most to align the gripper." P11, "The circle was what helped the most." We attribute the preferences towards the Circle, to the fact that it did not fully occlude the real object, and when it did it was an indicator of misalignment of the gripper over the target object. Further, participants' subjective preferences for Extensions and Colors together with higher efficiency and effectiveness lead us to recommend a simpler design of visual cues.

In H3, we did not expect Colors to perform the worst across conditions. In consequence, we further explored the reasons behind the low scores in time and success by analyzing the participants' comments. During our interview, participants expressed confusion when using the Colors due to lack of color opponency. Specifically, they expressed problems with the color gradient on the physical colormap, e.g., P13, "The gradient of the colormap made it a little bit confusing. I could not tell if it was already yellow or still blue?" P15, "I would prefer stripes, not gradient colors, that would have made it easier to distinguish where I was on the tabletop." Added to that, participants expressed difficulties in distinguishing

the real and physical colors, which is related to focal rivalry—the human visual system cannot focus on two elements at the same time, e.g., P14, "I had to concentrate to match the colors;" P15, "It was hard to see the real and the virtual colors because they were right above each other." Therefore, we believe that this lack of color opponency added to focal rivalry worsened depth perception. This accords to the observations of Ellis & Menges [15], who determined that physical surfaces influence depth perception misestimation.

Fourth **(H4)**, we hypothesized about finding a correlation between certainty and success as a possible pointer to better depth perception. Our results confirmed the correlation between certainty and success, even when controlling for our independent variables (distance, pose, and design of cues). We still found a positive and significant, albeit rather weak, correlation. This aligns with a line of research suggesting that subjective certainty correlates closely with objective success [8], [58]. We consider that all our design cues provided additional evidence about the position of the gripper in the workspace, which in turn influenced certainty. A note of caution is due here since, as mentioned in Section 2.4, there have been controversial findings related to the influence of time and success in certainty. We acknowledge that other factors can influence certainty and were not considered in our experiment such as fatigue and changes in attention, and these have proven to influence perceived certainty [20]. Further, when evaluating depth perception in 3D environments, it is necessary to separate "the amount of depth that an object is seen to have (mind independent property) and the realism of the experience (mind dependent property)" [23]. In fact, we consider that this construction can also be applied to AR environments and is related to certainty. This in consequence might have influenced the perceived certainty during our experiment.

7 Limitations

A limitation of our work relates to the multimodal interaction technique used. While a joystick or a control pad are commonly used in robot teleoperation, we aim to explore hands-free multimodal interaction. This type of interaction has raised interest in the research community, especially in HRI. Our previous experiences have shown that using speech and head movements is simpler to learn than using a control pad or even a joystick. Additionally, these modalities are natively supported by the technology used (Microsoft HoloLens 1). We further stress that the modalities were kept stable among conditions to avoid them causing a major influence on the evaluation of the other factors in our experiment.

This work may be also limited to the technology used. For instance, our design of the Colors cue, which combines a real-world colormap with virtual colors displayed over it, could have contributed to focal rivalry problems. However, this problem is present not only in the Microsoft HoloLens 1 but in the current generation of MR headsets [31]. Additionally, current ARHMDs have a limited

field of view which do not cover human's peripheral vision. This presents a disadvantage when using AR, as mentioned by Williams et al. [54].

Our experimental design also presents certain limitations. We did not consider a condition without cues since a previous study [1] already evaluated the use of visual cues versus the absence of them, suggesting potential improvements in certainty. This study thus builds upon those findings and further explores the influence of distance and pose. We did not consider grasping accuracy, defined by the distance to an ideal grasping position, since we realized that the effect of the displacement of the virtual visual cues at 3m influences greatly this measure.

Our main goal is to capture first experiences with certain designs of visual cues that can provide direction for a better design that improves depth perception. Furthermore, people can perceive distance and depth differently due to individual differences in visual acuity, eye dominance, color vision, and spatial abilities [29]. We considered different color visions for the colors used in the colormap but left aside the other factors which might have influenced on how each participant experienced not only the types of cues but depth perception.

8 Conclusions

In this paper, we evaluated how egocentric distances and target objects' pose affect co-located teleoperation when using certain designs of visual cues presented through AR. To this end, we performed an experiment with 24 participants. Our results align with previous studies about depth perception within the observer's personal space. Teleoperating a robot at 2m with our designs of visual cues leads to a higher success rate, shorter time, and a higher degree of certainty compared to 3m. A clearer pose leads to higher success and certainty but requires longer time. As time and certainty are closely tied, i.e., greater thoughtfulness is required to achieve higher certainty and success. We also found a positive correlation between success and certainty, but we are careful with these findings since other factors, e.g., attention and fatigue, have an effect on certainty and were not considered in our experiment. Additionally, we found differences of our designs of visual cues. The Circle and Extensions cues prompted shorter times and higher success compared to Colors, wherein Circle showed the highest success. These findings suggest that simplicity of design leads to higher efficiency and effectiveness.

We consider that our findings are thought-provoking and present a detailed analysis of distance and target pose in co-located teleoperation. These, further contemplate the role of certainty as a factor that can shed some light about depth perception. Besides, our designs of visual cues suggest advantages and downsides of using cast shadows, symbolic enhancements, and combining real and virtual landmarks to enhance the visual space when teleoperating a robotic arm in co-located spaces. Although our study focuses on evaluating specific factors related to depth perception, our findings may well have a bearing on designing cues using AR to improve perception of real objects and facilitate teleoperation of robotic arms.

Acknowledgements

This research is supported by the German Federal Ministry of Education and Research (BMBF, FKZ: 13FH011IX6).

References

1. Arévalo Arboleda, S., Dierks, T., Rücker, F., Gerken, J.: There's more than meets the eye. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. pp. 104–106. ACM, [S.l.] (2020). https://doi.org/10.1145/3371382.3378240
2. Audley, R.J.: A stochastic model for individual choice behavior, vol. 67 (1960). https://doi.org/10.1037/h0046438
3. Baranski, J.V., Petrusic, W.M.: Probing the locus of confidence judgments: Experiments on the time to determine confidence. Journal of experimental psychology. Human perception and performance **24**(3), 929–945 (1998). https://doi.org/10.1037//0096-1523.24.3.929
4. Barden, J., Petty, R.E.: The mere perception of elaboration creates attitude certainty: exploring the thoughtfulness heuristic. Journal of Personality and Social Psychology **95**(3), 489–509 (2008). https://doi.org/10.1037/a0012559
5. Barthelmé, S., Mamassian, P.: Evaluation of objective uncertainty in the visual system. PLoS Computational Biology **5**(9), e1000504 (2009). https://doi.org/10.1371/journal.pcbi.1000504
6. Blanca, M.J., Alarcón, R., Arnau, J., Bono, R., Bendayan, R.: Non-normal data: Is anova still a valid option? Psicothema **29**(4), 552–557 (2017). https://doi.org/10.7334/psicothema2016.383
7. Bodenheimer, B., Meng, J., Wu, H., Narasimham, G., Rump, B., McNamara, T.P., Carr, T.H., Rieser, J.J.: Distance estimation in virtual and real environments using bisection. In: Proceedings, APGV 2007. p. 35. ACM (2007). https://doi.org/10.1145/1272582.1272589
8. Boldt, A., Yeung, N.: Shared neural markers of decision confidence and error detection. Journal of Neuroscience **35**(8), 3478–3484 (2015). https://doi.org/10.1523/JNEUROSCI.0797-14.2015
9. Brenner, E., Smeets, J.B.J.: Depth perception. In: Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, pp. 1–30. John Wiley & Sons, Inc (2018). https://doi.org/10.1002/9781119170174.epcn209
10. Choi, H., Cho, B., Masamune, K., Hashizume, M., Hong, J.: An effective visualization technique for depth perception in augmented reality-based surgical navigation. The International Journal of Medical Robotics and Computer Assisted Surgery **12**(1), 62–72 (2016). https://doi.org/10.1002/rcs.1657
11. Cipiloglu, Z., Bulbul, A., Capin, T.: A framework for enhancing depth perception in computer graphics. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization. p. 141. ACM (2010). https://doi.org/10.1145/1836248.1836276
12. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances. In: Perception of Space and Motion, pp. 69–117. Elsevier (1995). https://doi.org/10.1016/B978-012240530-3/50005-5

13. Diaz, C., Walker, M., Szafir, D.A., Szafir, D.: Designing for depth perceptions in augmented reality. In: 2017 IEEE International Symposium on Mixed and Augmented Reality. pp. 111–122. IEEE (2017). https://doi.org/10.1109/ISMAR.2017.28
14. El Jamiy, F., Marsh, R.: Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. IET Image Processing **13**(5), 707–712 (2019). https://doi.org/10.1049/iet-ipr.2018.5920
15. Ellis, S.R., Menges, B.M.: Localization of virtual objects in the near visual field. Human Factors: The Journal of the Human Factors and Ergonomics Society **40**(3), 415–431 (1998). https://doi.org/10.1518/001872098779591278
16. de Gardelle, V., Mamassian, P.: Does confidence use a common currency across two visual tasks? Psychological science **25**(6), 1286–1288 (2014). https://doi.org/10.1177/0956797614528956
17. Gherman, S., Philiastides, M.G.: Neural representations of confidence emerge from the process of decision formation during perceptual choices. NeuroImage **106**, 134–143 (2015). https://doi.org/10.1016/j.neuroimage.2014.11.036
18. Gong, L.L., Ong, S.K., Nee, A.Y.C.: Projection-based augmented reality interface for robot grasping tasks. In: Proceedings of the 2019 4th International Conference on Robotics, Control and Automation - ICRCA 2019. pp. 100–104. ACM Press (2019). https://doi.org/10.1145/3351180.3351204
19. Hedayati, H., Walker, M., Szafir, D.: Improving collocated robot teleoperation with augmented reality. In: HRI'18. pp. 78–86. ACM (2018). https://doi.org/10.1145/3171221.3171251
20. Heereman, J., Walter, H., Heekeren, H.R.: A task-independent neural representation of subjective certainty in visual perception. Frontiers in human neuroscience **9**, 551 (2015). https://doi.org/10.3389/fnhum.2015.00551
21. Heinrich, F., Bornemann, K., Lawonn, K., Hansen, C.: Depth perception in projective augmented reality: An evaluation of advanced visualization techniques. In: 25th ACM Symposium on Virtual Reality Software and Technology. pp. 1–11. ACM (11122019). https://doi.org/10.1145/3359996.3364245
22. Hershenson, M.: Size-distance invariance: kinetic invariance is different from static invariance. Perception & psychophysics **51**(6), 541–548 (1992). https://doi.org/10.3758/BF03211651
23. Hibbard, P.B., Haines, A.E., Hornsey, R.L.: Magnitude, precision, and realism of depth perception in stereoscopic vision. Cognitive research: principles and implications **2**(1), 25 (2017). https://doi.org/10.1186/s41235-017-0062-7
24. Howard, I.P.: Depth perception. In: Stevens' handbook of experimental psychology (2002), pp. 77–120
25. Howard, I.P., Rogers, B.J.: Perceiving in depth, vol. 29. Oxford University Press (2012)
26. Jones, A., Swan, J.E., Singh, G., Kolstad, E.: The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In: IEEE virtual reality 2008. pp. 267–268. IEEE (2008). https://doi.org/10.1109/VR.2008.4480794
27. Kiani, R., Corthell, L., Shadlen, M.N.: Choice certainty is informed by both evidence and decision time. Neuron **84**(6), 1329–1342 (2014). https://doi.org/10.1016/j.neuron.2014.12.015
28. Kim, H.Y.: Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restorative dentistry & endodontics **38**(1), 52–54 (2013). https://doi.org/10.5395/rde.2013.38.1.52

29. Kruijff, E., Swan, J.E., Feiner, S.: Perceptual issues in augmented reality revisited. In: 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2010. pp. 3–12 (2010). https://doi.org/10.1109/ISMAR.2010.5643530
30. Krupke, D., Steinicke, F., Lubos, P., Jonetzko, Y., Gorner, M., Zhang, J.: Comparison of multimodal heading and pointing gestures for co-located mixed reality human-robot interaction. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1–9. IEEE (2018). https://doi.org/10.1109/IROS.2018.8594043
31. Laramee, R., Ware Colin: Rivalry and interference with a head-mounted display. ACM Transactions on Computer-Human Interaction **9**(3), 238–251 (2002). https://doi.org/10.1145/568513.568516
32. Lee, J.H., Herzog, T.A., Meade, C.D., Webb, M.S., Brandon, T.H.: The use of gee for analyzing longitudinal binomial data: A primer using data from a tobacco intervention. Addictive Behaviors **32**(1), 187 – 193 (2007). https://doi.org/https://doi.org/10.1016/j.addbeh.2006.03.030
33. Livingston, M.A., Zanbaka, C., Swan, J.E., Smallman, H.S.: Objective measures for the effectiveness of augmented reality. In: Virtual reality 2005. pp. 287–288. IEEE (2005). https://doi.org/10.1109/VR.2005.1492798
34. Loomis, J., Knapp, J.: Visual perception of egocentric distance in real and virtual environments. In: Virtual and adaptive environments, pp. 21–46. Lawrence Erlbaum (2003). https://doi.org/10.1201/9781410608888.pt1
35. Mamassian, P.: Visual confidence. Annual review of vision science **2**, 459–481 (2016). https://doi.org/10.1146/annurev-vision-111815-114630
36. Microsoft: Microsoft hololens, https://www.microsoft.com/en-us/hololens
37. Microsoft: Microsoft hololens calibration, https://docs.microsoft.com/en-us/hololens/hololens-calibration
38. Microsoft: Mixed reality toolkit (2017), https://github.com/microsoft/MixedRealityToolkit-Unity/releases
39. Mosiello, G., Kiselev, A., Loutfi, A.: Using augmented reality to improve usability of the user interface for driving a telepresence robot. Paladyn, Journal of Behavioral Robotics **4**(3) (2013). https://doi.org/10.2478/pjbr-2013-0018
40. Paris, R., Joshi, M., He, Q., Narasimham, G., McNamara, T.P., Bodenheimer, B.: Acquisition of survey knowledge using walking in place and resetting methods in immersive virtual environments. In: Proceedings of the ACM Symposium on Applied Perception. pp. 1–8. ACM (2017). https://doi.org/10.1145/3119881.3119889
41. Park, J., Ha, S.: Visual information presentation in continuous control systems using visual enhancements. In: Contact-free Stress Monitoring for User's Divided Attention. INTECH Open Access Publisher (2008). https://doi.org/10.5772/6307
42. Petrocelli, J.V., Tormala, Z.L., Rucker, D.D.: Unpacking attitude certainty: attitude clarity and attitude correctness. Journal of Personality and Social Psychology **92**(1), 30–41 (2007). https://doi.org/10.1037/0022-3514.92.1.30
43. Ping, J., Weng, D., Liu, Y., Wang, Y.: Depth perception in shuffleboard: Depth cues effect on depth perception in virtual and augmented reality system. Journal of the Society for Information Display **28**(2), 164–176 (2020). https://doi.org/10.1002/jsid.840
44. Pleskac, T.J., Busemeyer, J.R.: Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychological review **117**(3), 864–901 (2010). https://doi.org/10.1037/a0019737
45. PTC: Vuforia engine in unity, https://library.vuforia.com/articles/Training/getting-started-with-vuforia-in-unity.html

46. Rolland, J.P., Meyer, C., Arthur, K., Rinalducci, E.: Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. Presence: Teleoperators and Virtual Environments **11**(6), 610–625 (2002). https://doi.org/10.1162/105474602321050730
47. Rosales, C.S., Pointon, G., Adams, H., Stefanucci, J., Creem-Regehr, S., Thompson, W.B., Bodenheimer, B.: Distance judgments to on- and off-ground objects in augmented reality. In: Proceedings, 26th IEEE Conference on Virtual Reality and 3D User Interfaces. pp. 237–243. IEEE (2019). https://doi.org/10.1109/VR.2019.8798095
48. S. Stadler, K. Kain, M. Giuliani, N. Mirnig, G. Stollnberger, M. Tscheligi: Augmented reality for industrial robot programmers: Workload analysis for task-based, augmented reality-supported robot control. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 179–184 (2016). https://doi.org/10.1109/ROMAN.2016.7745108
49. Shimojo, S., Paradiso, M., Fujita, I.: What visual perception tells us about mind and brain. Proceedings of the National Academy of Sciences of the United States of America **98**(22), 12340–12341 (2001). https://doi.org/10.1073/pnas.221383698
50. Tormala, Z.L.: The role of certainty (and uncertainty) in attitudes and persuasion. Current Opinion in Psychology **10**, 6–11 (2016). https://doi.org/10.1016/j.copsyc.2015.10.017
51. Vassallo, R., Rankin, A., Chen, E.C.S., Peters, T.M.: Hologram stability evaluation for microsoft hololens. In: Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment. p. 1013614. SPIE Proceedings, SPIE (2017). https://doi.org/10.1117/12.2255831
52. Walker, M., Hedayati, H., Lee, J., Szafir, D.: Communicating robot motion intent with augmented reality. In: Kanda, T., Ŝabanović, S., Hoffman, G., Tapus, A. (eds.) HRI'18. pp. 316–324. ACM (2018). https://doi.org/10.1145/3171221.3171253
53. Walker, M.E., Hedayati, H., Szafir, D.: Robot teleoperation with augmented reality virtual surrogates. In: HRI'19. pp. 202–210. IEEE (2019). https://doi.org/10.1109/HRI.2019.8673306
54. Williams, T., Hirshfield, L., Tran, N., Grant, T., Woodward, N.: Using augmented reality to better study human-robot interaction. In: Virtual, Augmented and Mixed Reality. Design and Interaction, vol. 12190, pp. 643–654. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-49695-1_43
55. Willis, J., Todorov, A.: First impressions: making up your mind after a 100-ms exposure to a face. Psychological science **17**(7), 592–598 (2006). https://doi.org/10.1111/j.1467-9280.2006.01750.x
56. Wither, J., Hollerer, T.: Pictorial depth cues for outdoor augmented reality. In: ISWC 2005. IEEE (2005). https://doi.org/10.1109/ISWC.2005.41
57. Xie, F., Paik, M.C.: Generalized estimating equation model for binary outcomes with missing covariates. Biometrics **53**(4), 1458 (1997). https://doi.org/10.2307/2533511
58. Yeung, N., Summerfield, C.: Metacognition in human decision-making: confidence and error monitoring. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **367**(1594), 1310–1321 (2012). https://doi.org/10.1098/rstb.2011.0416
59. Yonas, A., Granrud, C.E.: Infants' perception of depth from cast shadows. Perception & psychophysics **68**(1), 154–160 (2006). https://doi.org/10.3758/bf03193665
60. Zollmann, S., Hoppe, C., Langlotz, T., Reitmayr, G.: Flyar: augmented reality supported micro aerial vehicle navigation. IEEE transactions on visualization and computer graphics **20**(4), 560–568 (2014). https://doi.org/10.1109/TVCG.2014.24