Visualizing Robot Intent for Object Handovers with Augmented Reality

Rhys Newbury, Akansel Cosgun, Tysha Crowley-Davis, Wesley P. Chan, Tom Drummond, Elizabeth A. Croft
Monash University, Australia

***Abstract*— Humans are highly skilled in communicating their intent for when and where a handover would occur. However, even the state-of-the-art robotic implementations for handovers display a general lack of communication skills. This study aims to visualize the internal state and intent of robots for Human-to-Robot Handovers using Augmented Reality. Specifically, we aim to visualize 3D models of the object and the robotic gripper to communicate the robot's estimation of where the object is and the pose in which the robot intends to grasp the object. We tested this design via a user study with 16 participants, in which each participant handed over a cube-shaped object to the robot 12 times. Results show that visualizing robot intent using augmented reality substantially improves the subjective experience of the users for handovers. Results also indicate that the effectiveness of augmented reality is even more pronounced for the perceived safety and fluency of the interaction when the robot makes errors in localizing the object.**

I. Introduction

Handing over objects is a ubiquitous interaction type between humans and an important skill for robots that interact with people. Humans often communicate their intent for when and where a handover will occur using several modalities, including speech, gaze, or body gestures. This observation indicates that robots also require such communication skills. Recent years have seen a proliferation of research in human-robot handovers [1]–[4]. Even though it has been found that communication cues such as speech and gaze positively impact human-robot handovers [5], a recent survey found that a majority of robotic systems that perform handovers did not use any communication cues at all [6]. These findings suggest that there is room for innovation in how to communicate intent for human-robot handovers.

Recent advancements in graphics and hardware technology have led to increased popularity of Augmented Reality (AR). AR presents new opportunities for robotics applications, and it is especially promising for Human-Robot Interaction [7]. Using AR, robots can communicate their intent to users through visual displays, allowing users to consider the robot's plans and act accordingly. This idea has found some success in previous studies where AR was used to communicate the future motion of robotic arms [8], [9].

In this paper, we explore the use of AR for a specific application: to communicate the robot's internal state and intent in Human-to-Robot handovers. We propose visualizing two 3D models through AR: 1) The detected object pose, visualized as a wireframe of the object model, and 2) The planned grasp pose, visualized as a virtual, low-opacity 3D model of the robotic gripper. A single object, which is tracked with the help of artificial markers, is used for the experiments. Figure 1 shows the experimental setup and example views from the AR Headset. We conduct a user study to compare the perceived safety, trust and fluency of the handovers using our proposed AR interface. Half of the conditions for each user involved adding a random but controlled **faked error** to the robot's estimation of the object pose to simulate markerless vision-based pose tracking in which accurate pose estimation is challenging.

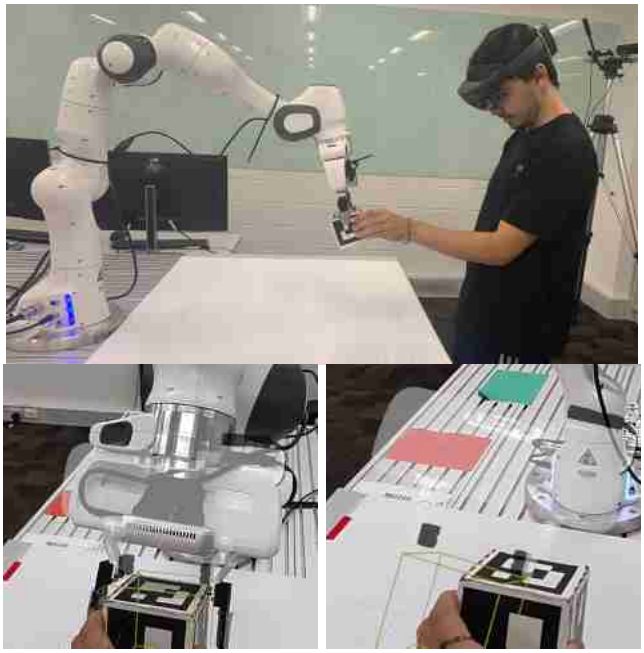


Fig. 1: Top: The robot picking up an object from the user's hand. The user is wearing an Augmented Reality headset. Bottom Left: The detected pose of the object, and AR visualization of how the robot is planning to grasp the object. Bottom Right: A "faked" error is introduced to object pose estimation, to test how the system would perform when the robot makes errors.

The contributions of this paper are two-fold:

- The first use of Augmented Reality for human-robot handovers, to convey the robot's belief on where the handover would occur, as well as its plan on how it would grasp the object.
- User studies that confirm the effectiveness of the proposed AR interface.

The organization of this paper is as follows. Related literature is reviewed in Section II and the system is described in Section III. The hypotheses are described in Section IV and the user study design is presented in Section V. Results are then presented in Section VI. We discuss the results in Section VII before concluding in Section VIII.

II. Related Work

In this section, we review the literature in Human-to-Robot handovers and communication of robot intent.

A. Human-to-Robot Handovers

As the need for collaborative manipulation increases, the importance of researching object handovers between robots and humans continues to grow. A recent survey paper by Ortenzi [6] reviews the progress made in this field. They summarize different capabilities that enable handovers in the robotic system, including communication, grasp planning, perception, and error handling. Among the papers surveyed, most focus on Robot-to-Human handovers, while only a small fraction of the papers focus on Human-to-Robot handovers. The authors observe that communication during handovers is often overlooked, with very few papers addressing the communication of intent as the primary focus. This highlights a gap in the research regarding how to communicate the robot's intent to the user effectively.

Literature in the field has predominantly focused on specific aspects of handovers, such as the trajectories of the arm either learning from human-data [10], using dynamic primitives [11] or predicting the transfer point of the object [12]. Two recent works focus on creating a human-to-robot handover system that can generalize to a series of objects [1], [4]. These approaches use a deep learning-based grasp planner and skin segmentation to find a safe grasp pose for the robot. Yang [4] also made the system reactive, allowing the grasp pose to change through the duration of the handover.

B. Communication of Robot Intent

Norman [13] proposed the idea of the *Gulf of Evaluation* as the ability for the user to directly interpret the state of the system. In this context, during human-robot interaction, a lack of communication of goals and intent from the robot can lead to a large *gulf of evaluation*. Effective communication from the robot can enhance the user's perception of the reliability of the system and make the human feel more comfortable around the robot [14].

One option to achieve a more natural communication between the human and the robot is using a head-mounted display (HMD) to visualize the robot's future motion from the point of view of the human [15], [16]. This idea has been utilized for both the motion of wheeled robots [17] and robotic arms [8], [9]. Walker [17] utilized AR to communicate the motion intent of a robot. They used different visualization markers to show the robot's intended path and directions, allowing the humans to know the robot's motions in advance. They found that AR could improve task efficiency over a baseline in which users only saw physically embodied orientation cues. Rosen [8] utilizes AR to visualize the path of the robotic arm to allow better collaboration between human and robot. They show an increase in accuracy and decrease in the time taken to label a trajectory as either collision or collision-free with blocks on a table. Gruenefeld [9] shows that using AR to visualize the robot's future motion can allow humans and robots to work in small shared spaces with a decreased likelihood of shutdowns.

C. Communication of Robot Intent in Handovers

Research has previously explored the use of nonverbal cues in human-robot handovers. The use of gaze results in faster object reaching and more natural perception of the interaction by human receivers [5]. Admoni [18] further showed that modulating the speed of the handovers, such that the object released is delayed until the human gaze is drawn back to the robot, can increase the conscious perception of the robot's communication. Furthermore, integrating the use of both head orientation and eye gaze into the decision-making of the robot significantly increases the success rate of robot-to-human handovers [19]. Other communication modalities explored include body gestures, such as an extended arm which presents the object to the receiver such that the free part of the object is directed towards the receiver can convey intent to initiate a handover [20]. Pan [21] showed that the initial pose of the robot could inform the giver about the geometry of the handover and improve the fluency of the handover.

To our knowledge, AR has previously not been utilized for signaling robot intent in human-robot handovers. Our work aims to fill the research gap in investigating the communication the robot intent through AR.

III. Approach

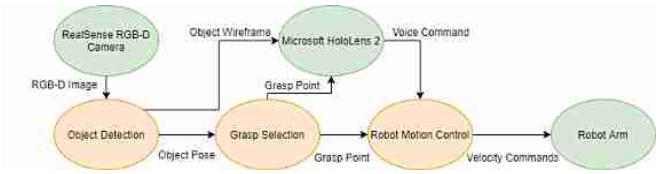


Fig. 2: The system diagram. Hardware components are shown in green. The orange software blocks were developed in this system..

Our system consists of 3 modules, as shown in Fig. 2. For simplicity, we use a generic cube object for handovers. We adopt a fixed world frame, with the origin located on the robotic arm's base, where gravity acts in the $-z$ direction. We estimate the pose of the cube by using the artificial markers located on each face of the cube. To pick up the object from the user's hand, a feasible grasp pose is selected among several predefined grasp poses with respect to the object frame. A voice command initiates the handover detected through an AR Headset. A simple servo controller is used to drive the robot end-effector on a linear path that connects the starting end-effector pose to the selected grasp pose. The robot's estimation of the object pose and the selected grasp pose is visualized using the AR Headset to give users a sense of the robot's intention.

A. Hardware Setup

We use a table-mounted Franka Emika Panda robotic arm. The robot has 7 degrees of freedom and a two-finger parallel gripper. To increase the robustness of grasps, we use custom-made gripper fingers made of silicon rubber, as suggested

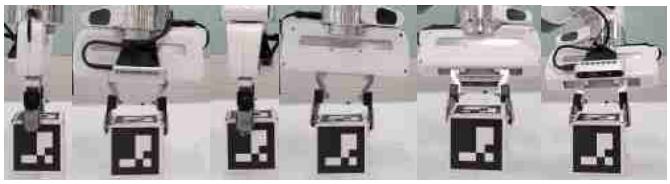


Fig. 3: Predefined grasp poses for one face and one edge of the cube used in the experiments. To grasp a face of the cube, we consider 4 possible rotations of the end-effector that are all 90° apart. To grasp the edge of a cube, we consider 2 possible rotations of the end-effector that are 180° apart.

by [22]. Our system uses a single Intel RealSense D435 RGB-D camera that is mounted to the end-effector. The system is implemented over two computers, one equipped with a GPU (Nvidia GTX 1080) and the other with a real-time kernel and high-performance network card for real-time control of the robotic arm. Both computers use 64 bit Ubuntu 18.04 LTS and are connected via LAN. Robot Operating System (ROS) is used for inter-node messaging.

We use a Microsoft HoloLens 2 as the designated AR Headset, which provides accurate positional tracking of the user's head, allowing for accurate visualizations in 3D space. The coordinate frame transformation between the robot and the HoloLens is initially established using an artificial marker placed in a predefined location. After the initial calibration, the HoloLens continually updates its pose with respect to the robot base frame. Communication of data to the HoloLens is achieved using WiFi over a local area network.

B. Object Detection

An 8cm edge cube object was used for our experiments. We placed unique arUco markers on each side of the cube, with each arUco marker oriented such that the z-axis of the marker is pointing outwards. The pose of the cube can be calculated by detecting a single arUco marker and projecting a fixed distance in the $-z$ direction of the detected marker. When multiple markers are detected, the best fit for the pose is used. A low-pass filter is applied to the resulting pose to smooth any sudden changes in object pose.

C. Grasp Selection

We consider four predefined grasp poses for each face and two grasp poses on each edge of the cube for a total of 48 possible grasp poses on the cube object. Each grasp pose is defined in the coordinate frame of the cube object. The grasp poses for one face and one edge of the cube is shown in Fig 3. These poses are repeated for each face/edge. We heuristically choose between the defined grasp poses according to the rules below.

1) The angle (α) between the vector pointing in the direction of the center of the cube from a grasp pose (V_c) and the forward vector of the end-effector mounted camera must be less than $120°$. This is shown in Fig. 4.
2) We choose the face/side of the cube with the largest z coordinate in the world frame.
3) We filter between the two or four grasp poses (depending on if grasping on a face or side), removing kinematically infeasible grasp poses. If no grasp poses are kinematically feasible, we choose the next highest face/side and repeat.
4) From the remaining grasp poses, we choose the pose with the minimum angular difference between the current pose and the respective grasp pose.

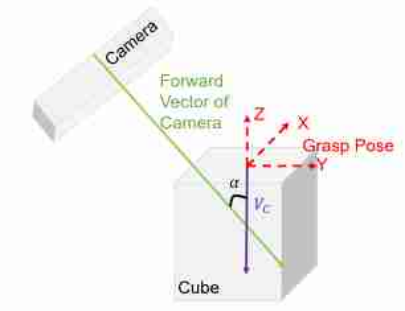


Fig. 4: The angle (α) between the forward vector of the camera and V_c must be less $120°$.

The grasp pose is updated in real-time until the user starts the handover. The grasp pose will then be fixed, and the robot arm will move towards the selected grasp pose. These rules are designed such that we are likely to choose feasible grasp positions for the arm. If the user holds the object flat, four edges and a face of a cube, will all have approximately the same z value, with only noise determining the grasp pose. This leads to the pose jumping between the grasp poses. To mitigate this, we bias the z coordinate of the face/side in which the previous grasp point exists by $2cm$. We found that this bias prevents the grasp pose to jump unless there is a big change of the pose on the cube. Finally, the selected grasp pose is filtered over a moving window of 5 frames, which reduces the likelihood of a false positive detection due to sensor noise.

D. Robot Motion Control

We use resolved-rate motion control to drive the robot towards the desired grasp pose on a linear path. The joint velocity commands are sent to the low-level controller of the Panda robot. Once the robot reaches the grasp pose, the gripper is closed until a force threshold is reached, the end-effector is moved towards a pre-defined dropping location, and the gripper is opened.

We do not consider the human hand while servoing to the object, however, for safety, we slow the arm down significantly for the last 7cm of the handover. This ensures there is plenty of time for both the human interaction partner and experiment supervisor to react to a potentially unsafe situation. Additional safety considerations, similar to [1] could be added, however, this was not the primary focus of this work.

E. Visualization in Augmented Reality

Two 3D models are displayed through AR, which is intended to inform the user as to where the robot thinks the object is located and which pose the robot will attempt to grasp the object from. The two models are:

1) Detected Object Pose: The wireframe of the object is visualized at its estimated 6D pose.
2) Planned Grasp Pose: The 3D model of the robot gripper is visualized as grasping the object wireframe, with the selected grasp pose the robot is planning to execute. The gripper model is visualized with low opacity in order not to block the user's view.

Once the user initiates the handover with a voice command, the robot's chosen grasp pose and visualizations are fixed in place until the object is picked up from the human.

IV. Hypotheses

Following our previous work in human-to-robot handovers [1], we anticipate that humans will feel safer and more trusting of the robot with the ability to visualize the robot's intent. We also expect it to decrease the number of failed handovers, with less reliance on the experimenter manually stopping the robot. We formulate the following hypotheses to test on a user study with a robot:

H1 The use of **AR** will have a positive effect on the subjective metrics related to **fluency** when completing human-to-robot handovers.

H2 The use of **AR** will have a positive effect on the subjective metrics related to **trust** when completing human-to-robot handovers.

H3 The use of **AR** will have a positive effect on the subjective metrics relating to the **predictability** between the human and the robot.

H4 The use of **AR** will have a positive effect on the subjective metrics related to **safety** when completing human-to-robot handovers.

H5 The use of **AR** will decrease the **mental load** to complete the task.

H6 **AR** have will have a larger positive effect on all subjective metrics when the fake error artifact is present.

V. User Study Design

A user study was conducted to test the effect of AR on human-to-robot handovers. The methodology for this user study is inspired by our previous work on human-robot collaboration [23].

A. Independent Variables

We manipulate two independent variables:

1) Visualization Mode: One of the two following conditions is chosen.
   a) **AR**: The participant wore the AR headset. The object's 6D pose and the robot's planned grasp pose is visualized, as described in Sec. III-E.
   b) **No AR**: The participant still wore the AR headset, but no visualizations were displayed.
2) Presence of Perception Error Artifact: One of the two following conditions is chosen.
   a) **Without "Faked Error"**: Object pose estimation result along with the selected grasp pose is visualized as is.
   b) **With "Faked Error"**: We introduce purposeful additions of error to the object pose estimation. The robot servos towards the erroneous grasp pose and the user is expected to compensate for the error by moving the object inside the robot gripper. Random noise is added to the object position, parameterized as an angle (sampled uniformly) and a distance which is sampled from a normal distribution (μ=10cm, σ=1cm) at the start of a handover. The positional noise is added along the plane perpendicular to the axis emanating from the camera's forward direction. Random noise is also added to each Euler angle, sampled from a normal distribution (μ=10°,σ=3°). In this study, we are only modeling object position error, however, during more realistic scenarios, we would expect many different sources of error. We expect the handover partner to compensate for these errors by moving the object to the final position of the robot.

   The use of faked error serves two purposes. First, since we use an idealized pose estimator using artificial markers, the faked error mimics a marker-less pose estimation system by adding a controlled artificial error. Second, we are interested in understanding whether AR would help participants compensate for the robot's errors as well as whether AR would still bring value in the presence of possible robot perception errors.

We adopt a 2 by 2 design, therefore, each participant experienced 4 different conditions:

1) **AR, Without Faked Error**
2) **No AR, Without Faked Error**
3) **AR, With Faked Error**
4) **No AR, With Faked Error**

To reduce order effects, the order of visualization mode is counterbalanced between participants. The order of **With Faked Error** and **Without Faked Error** conditions were fixed, however, with the **With Faked Error** condition always following the **Without Faked Error** condition. This is because we are interested in capturing participants' actual opinions on our proposed system first. Exposing participants to the erroneous robot behavior first would have likely affected their subjective opinion of the overall system.

B. Participant Allocation

We recruited 16 participants (13 male, 3 female) from within Monash University[1], aged $21-27$ (M = 23.5, SD = 1.82). The participants were not compensated for their time. 13 of the participants had some prior experience with robots whereas 3 had not seen a robot before. 6 of the participants had previous experience with AR, while 7 only heard about AR through popular media.

C. Procedure

The experiment took place at a university laboratory under the supervision of an experimenter. Participants stood at a

[1]Due to the ongoing COVID-19 pandemic, no external participants could be recruited. This study has been approved by the Monash University Human Research Ethics Committee (Application ID: 27499)

designated location in front of the robot arm to handover objects. Users first read the explanatory statement and signed a consent form. Next, the experimenter explained the experiment by reading from a script. Users then completed a small demographic survey, before completing a training phase without the use of AR. During training, the supervisor initiated the handover when the user was ready. This was repeated as many times as needed for the user to feel comfortable with the robot's behavior.

For the actual experiments, the user was required to successfully handover the cube 3 times for each condition, for a total of 12 successful handovers during the 4 trials. The user completed two types of surveys: a survey after each of the 4 test conditions, and one post-experiment survey for additional comments on the study. The survey questions are shown in Table I. The mean experiment duration was approximately 30 minutes.

Human-Robot Fluency
Q1: The robot contributed to the fluency of the interaction
Trust in Robot
Q2: I trusted the robot to do the right thing at the right time.
Predictability
Q3: I understand what the robot's goals are
Safety
Q4: I felt safe completing the handovers
Mental Load
Q5: How mentally demanding was the task? (R)

TABLE I: User Study Survey Questions

D. Dependent Variables

Subjective Measures: We adopt the metrics proposed by [6] and use a subset of questions that were relevant to the study. Question 5 was added about the cognitive load, adapted from NASA TLX [24]. These questions were designed to measure **H1** - **H5**. All questions were measured on a 5 point Likert scale. For Q1-Q4, 1 represents *Strongly disagree* and 5 represents *Strongly agree*. For Q5, 1 represents *Very easy* and 5 represents *Very difficult*.

Objective Measures: We count how many handover failures are experienced by the user.

VI. Results

We study the effect of the independent variables on the dependent variables. In total we analyze $(N = 16) * 4 = 64$ conditions. The results of the user study are shown in Fig. 5.

A. Objective Measures: Handover Failures

The only failure mode observed during this study was the robot being unsuccessful in grasping the object. This was primarily due to the person not compensating for the slight noise in the object pose estimation for the **Without Faked Error** condition or the additional error in the **With Faked Error** condition. The number of handover failures are shown in Tab. II. As expected, the most number of failures occurred during the **With Faked Error** and **No AR** case. This was due to users being unable to correctly estimate the grasp pose quickly enough. Overall, we observe that **AR** led to fewer number of handover failures.

Number of Failures	**No AR**	**AR**
Without Faked Error	4	4
With Faked Error	10	4

TABLE II: The distribution of failed handovers in all experiments. Each condition had a total of 48 successful handovers.

B. Subjective Measures: Survey Questions

For each question, we perform an Aligned Rank Transform [25], which allows for nonparametric testing of interactions and main effects of ordinal data, such as a Likert Scale, using standard ANOVA techniques. Interaction effects occur when the effects of an independent variable depend on the other variable. In this study, none of the questions had a significant interaction between the independent variables, which allowed us to examine the significance of an independent variable and collapsing the other independent variable. The comparison between **AR** and **No AR** cases is shown in Tab. III, and the comparison between **With Faked Error** and **Without Faked Error** is shown in Tab. IV.

	AR		No AR		F(1,63)	p
	μ	σ	μ	σ		
Q1 (Fluency)	3.66	1.15	3.06	1.22	2.710	0.105
Q2 (Trust)	3.97	1.00	3.03	1.15	11.078	**0.001**
Q3 (Predictability)	4.63	0.71	3.47	1.46	6.083	**0.017**
Q4 (Safety)	4.38	0.71	3.72	1.08	5.157	**0.027**
Q5 (Mental Load)	1.69	0.69	2.06	0.72	5.488	**0.022**

TABLE III: Results from the user study comparing AR and No AR and collapsing the other variable. The significance was calculated using Aligned Rank Transform ANOVA. Q2 - Q5 (bold) all have significant results.

	Without Error		With Error		F(1,63)	p
	μ	σ	μ	σ		
Q1 (Fluency)	3.81	0.93	2.91	1.30	8.530	**0.005**
Q2 (Trust)	3.69	1.06	3.31	1.26	1.955	0.156
Q3 (Predictability)	4.22	1.26	3.88	1.29	1.143	0.289
Q4 (Safety)	4.06	0.88	4.03	1.06	0.001	0.979
Q5 (Mental Load)	1.66	0.55	2.09	0.82	5.947	**0.017**

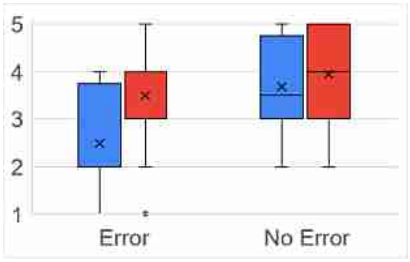
TABLE IV: Results from the user study comparing With Faked Error and Without Faked Error and collapsing the other variable. The significance was calculated using Aligned Rank Transform ANOVA. Only Q1 has significant results.

Fluency (Q1, H1): **AR** has a positive effect on the subjective measure of fluency, increasing the mean rating from 3.06 to 3.66, however the difference was not statistically significant. The presence of **Faked Error**, however, significantly reduced the fluency of the interaction. Therefore, **H1** was not affirmed.

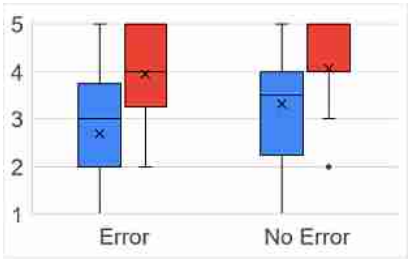
Trust (Q2, H2): The mean subjective perception of trust increases from 3.03 to 3.97 with the use of **AR** and the difference was statistically significant. This affirms **H2**. We do not observe a significant difference between the **With Fake Error** and **Without Fake Error** conditions.

Predictability (Q3, H3): The subjective level of predictability significantly increases from 3.47 to 4.63 with the use of **AR**. This affirms **H3**. Moreover, most participants noted that during the **Without Faked Error** and **AR** condition that they fully understood the robot's goals.

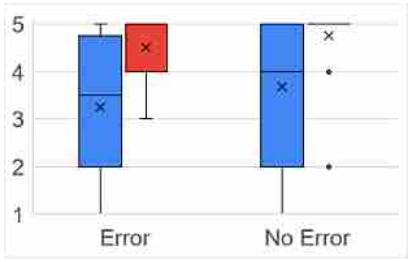
Safety (Q4, H4): Use of AR significantly increases the subjective level of perceived safety, increasing the mean



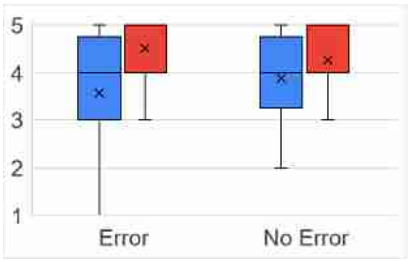
(a) Q1: The robot contributed to the fluency of the interaction



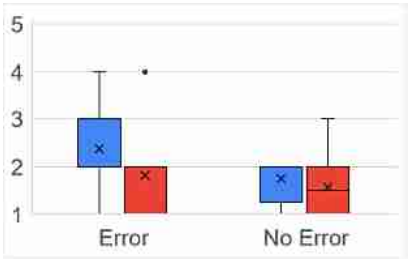
(b) Q2: I trusted the robot to do the right thing at the right time.



(c) Q3: I understand what the robot's goals are.



(d) Q4: I felt safe completing the handovers



(e) Q5: How mentally demanding was the task?

Fig. 5: User study results. Answers are on 5-Likert scale and Q5 is reverse scale

rating from 3.72 to 4.38. This affirms **H4**.

Mental Load (Q5, H5): The subjective mental load significantly decreases through the use of **AR**, from 2.06 to 1.69. This affirms **H5**. Fig 5e shows that the mental load is similar in all conditions apart from **With Faked Error** and **No AR**, where the user experienced an increase in the mental load. Furthermore, we found that the introduction of **Faked Error** caused the mental loads of the participants to increase significantly. The mean subjective rating on mental load decreased from 2.03 to 1.72 with the use of **AR**.

Value of AR under Pose Estimation Errors (H6): We are interested in investigating if the the benefit of using AR (defined as the difference of ratings between **AR** and **No AR**) is more pronounced when there is **Faked Error** imposed on the object pose. For this analysis, for each question, an additional Kruskal-Wallis test was performed on the difference between **AR** and **No AR** for both **With Faked Error** and **Without Faked Error** conditions. This formed two additional variables for each question. We then used the Kruskal-Wallis nonparametric test, which can be used for skewed Likert data [26]. This test aims to see if **AR** has a larger effect on the performance of the system in the **With Faked Error** condition compared to the **Without Faked Error** condition. The results of the Kruskal-Wallis test are summarized in Tab. V. Q1 (Fluency) and Q4 (Safety) have significant results, which partially affirms **H6**. Therefore, we observe that the advantage of using AR is more pronounced in metrics related to perceived fluency and safety when the robot's object pose estimation is imperfect.

	H(1)	P
Q1 (Fluency)	**6.960**	**0.003**
Q2 (Trust)	1.114	0.264
Q3 (Predictability)	0.051	0.808
Q4 (Safety)	**3.767**	**0.037**
Q5 (Mental Load)	1.114	0.244

TABLE V: We check if AR had a greater improvement on results in the presence of the Faked Error Artifact. A Kruskal-Wallis test was performed for each question. Significant results are indicated in bold text.

VII. Discussion

The user study results suggest that hypotheses **H2**, **H3**, **H4**, **H5** were affirmed. No significant results were observed for **H1**. **H6** was partially affirmed only for the subjective metrics of perceived fluency and safety.

The users' perceived level of safety was similar between both **Without Faked Error** and **With Faked Error** conditions, as shown in Fig. 5d, with a rating of 4.03 and 4.06 respectively. This suggests that overall, users felt safe interacting with the system.

Participants discussed the usefulness of seeing the robot's goal as it allowed them to adjust the object transfer point to account for errors in the robot's goal. Participant 3 noted that *"Visualizing the gripper was probably the best thing, since I was able to adjust my position to have the robot grasp the object and handover."*. Further, Participant 4 said that *"it was far more confusing when I did not have any grasp information"*. This is further shown by a large difference (1.16) in the subjective rating of the predictability between the **No AR** and **AR** conditions. Participants also noted that this visualization increased their feeling of trust and fluency in their handover.

For the **AR** and **With Faked Error** condition, it was observed that users tend to adjust their hand position relatively quickly to the final grasp point of the robot. Once this was completed, the users tended to stay quite still until the robot had completed the handover. Contrasting to this, during the

No AR and **With Faked Error** condition, the users, would continuously adjust their position as the robot moved. This takes more effort from the user and is the most likely cause for the increased mental load to complete the task without **AR**. Further qualitative studies could track the position of the user's wrist to prove this observation.

Some users noted that the visualization could be distracting, especially if it had a slight misalignment with the object for the **Without Faked Error** condition. Participant 9 said that *"The offset in the wireframe and the mismatching of visualization to the gripper was distracting at times. But once I got used to it, the visualization was helpful for my understanding of what was happening."*. This misalignment was caused by errors in calibration between the HoloLens and the robot and tracking errors of the HoloLens. Future work could include using a more accurate method to calibrate to ensure the visualization aligns throughout the handovers.

VIII. Conclusion and Future Work

Conveying the robot's intent is often disregarded in human-to-robot handovers. We propose a novel AR-based interface to communicate the robot's internal state to the user by visualizing the estimated pose of the object and where the robot is planning to grasp the object. User studies demonstrate that conducting the proposed interaction through AR significantly improves the subjective experience of the participants in terms of fluency of interaction, trust towards the robot, perceived safety, mental load, and predictability. Our proof-of-concept is achieved for a single object, in which its pose is tracked using artificial markers. The proposed method is subjectively perceived as safe, fluent and trustworthy even when random artificial noise is introduced to the grasp pose. This suggests that the approach would be well-suited to more realistic scenarios where accurate detection of the object may not be possible and that humans are willing to compensate for errors in robotic vision if robots can communicate their intent to their human partners.

This paper serves as a foray into communication methods for object handovers. What information should be visualized, and how remains an open research question. Different visualizations can be explored, such as the robot's future trajectory or an approximate bounding box of the object instead of the full 6D pose. Future directions include the possible use of interactive markers, and using hand gestures to allow the user to possibly move the robot's grasp pose to compensate for errors or grasp objects in a semantically preferred way.

IX. Acknowledgements

We thank Khoa Hoang and Tin Tran for their help during various stages of development.

References

[1] P. Rosenberger, A. Cosgun, R. Newbury, J. Kwan, V. Ortenzi, P. Corke, and M. Grafinger, "Object-independent human-to-robot handovers using real time robotic vision," *IEEE Robotics and Automation Letters*, 2021.

[2] J. Kwan, C. Tan, and A. Cosgun, "Gesture recognition for initiating human-to-robot handovers," *RO-MAN Workshop in Active Vision and Perception in Human-Robot Collaboration*, 2020.

[3] P. Ardón, M. E. Cabrera, È. Pairet, R. Petrick, S. Ramamoorthy, K. S. Lohan, and M. Cakmak, "Affordance-aware handovers with human arm mobility constraints," *arXiv preprint arXiv:2010.15436*, 2020.

[4] W. Yang, C. Paxton, A. Mousavian, Y.-W. Chao, M. Cakmak, and D. Fox, "Reactive human-to-robot handovers of arbitrary objects," *arXiv preprint arXiv:2011.08961*, 2020.

[5] A. Moon, D. M. Troniak, B. Gleeson, M. K. Pan, M. Zheng, B. A. Blumer, K. MacLean, and E. A. Croft, "Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing," in *IEEE International Conference on Human-Robot Interaction*, 2014.

[6] V. Ortenzi, A. Cosgun, T. Pardi, W. Chan, E. Croft, and D. Kulic, "Object handovers: A review for robotics," *arXiv preprint arXiv:2007.12952*, 2020.

[7] Z. Makhataeva and A. Varol, "Augmented reality for robotics: A review," *Robotics*, 2020.

[8] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *IJRR*, 2019.

[9] U. Gruenefeld, L. Prädel, J. Illing, T. Stratmann, S. Drolshagen, and M. Pfingsthorn, "Mind the arm: realtime visualization of robot motion intent in head-mounted augmented reality," in *MuC*, 2020.

[10] K. Yamane, M. Revfi, and T. Asfour, "Synthesizing object receiving motions of humanoid robots with human motion database," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[11] G. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Autonomous Robots*, 2017.

[12] H. Nemlekar, D. Dutia, and Z. Li, "Object transfer point estimation for fluent human-robot handovers," in *IEEE ICRA*, 2019.

[13] D. Norman, *The Psychology of Everyday Things*. Basic Books, 1988.

[14] R. T. Chadalavada, H. Andreasson, R. Krug, and A. J. Lilienthal, "That's on my mind! robot to human intention communication through on-board projection on shared floor space," in *ECMR*, 2015, pp. 1–6.

[15] E. Ruffaldi, F. Brizzi, F. Tecchia, and S. Bacinelli, "Third point of view augmented reality for robot intentions visualization," in *AVR*, 2016.

[16] B. Scassellati and B. Hayes, "Human-robot collaboration," *AI Matters*, vol. 1, pp. 22–23, 12 2014.

[17] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *HRI*, 2018.

[18] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication," in *ACM/IEEE international conference on Human-robot interaction*, 2014.

[19] E. C. Grigore, K. Eder, A. G. Pipe, C. Melhuish, and U. Leonards, "Joint action understanding improves robot-to-human object handover," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[20] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human hand-overs," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2011.

[21] M. K. Pan, E. A. Croft, and G. Niemeyer, "Exploration of geometry and forces occurring within human-to-robot handovers," in *IEEE Haptics Symposium (HAPTICS)*, 2018.

[22] M. Guo, D. V. Gealy, J. Liang, J. Mahler, A. Goncalves, S. McKinley, J. A. Ojea, and K. Goldberg, "Design of parallel-jaw gripper tip surfaces for robust grasping," in *ICRA*, 2017.

[23] S. Bansal, R. Newbury, W. Chan, A. Cosgun, A. Allen, D. Kulić, T. Drummond, and C. Isbell, "Supportive actions for manipulation in human-robot coworker teams," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[24] S. Hart and L. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139–183, 1988.

[25] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *SIGCHI conference on human factors in computing systems*, 2011.

[26] M. Schrum, M. Johnson, M. Ghuy, and M. Gombolay, "Four years in review: Statistical practices of likert scales in human-robot interaction studies," 03 2020, pp. 43–52.