To cite this version :

Anjela MAYER, Augustin RUNGEARD, Jean-Rémy CHARDONNET, Polina HÄFNER, Jivka OVTCHAROVA - Immersive Hand Instructions in AR – Insights for Asynchronous Remote Collaboration on Spatio-Temporal Manual Tasks - In: 2023 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Tunisie, 2023-06-12 - 2023 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) - 2023

Immersive Hand Instructions in AR – Insights for Asynchronous Remote Collaboration on Spatio-Temporal Manual Tasks

Anjela Mayer [*]
Institute for Information Management in Engineering
Karlsruhe Institute of Technology

Augustin Rungeard
Karlsruhe Institute of Technology

Jean-Rémy Chardonnet [†]
Arts et Metiers Institute of Technology
LISPEN, HESAM Université, UBFC

Polina Häfner[‡]
Institute for Information Management in Engineering
Karlsruhe Institute of Technology

Jivka Ovtcharova[§]
Institute for Information Management in Engineering
Karlsruhe Institute of Technology

ABSTRACT

Collaborative Virtual Environments (CVEs) have recently become more prominent in academic research and industrial applications. Physical tasks often rely on sharing spatial information between the collaborators, which are difficult to communicate verbally or by 2D media. Immersive technologies, like Mixed Reality (MR), can support more natural and intuitive interaction between the collaborators, share important non-verbal cues, like gaze and deictic gestures, and improve the performance time of remote guidance tasks. In this work we present HoloHands - a concept for immersive hand instructions which can be used for asynchronous remote collaboration on spatial tasks. A Hololens device is utilized to capture hand motions and interactions with virtual representations of physical objects to create a guidance by simply demonstrating the instruction steps. The recordings can be consumed by other collaborators at a later time to perform the physical task which we validated in a preliminary user experiment. Since it allows the creation and retention of digital information for a later consumption, it provides the possibility to revisit previous work sessions at any time and thus opens new possibilities to collaborate asynchronously on spatio-temporal tasks. In this work we share valuable insights from our experience with the HoloHands concept and the utilization of the Hololens headset.

Index Terms: Asynchronous interaction, Virtual and augmented reality, Collaborative learning, Computer-supported cooperative work

1 INTRODUCTION

Collaborative work is continuously transforming and in front of modern challenges like globalization, digitization, pandemics and climate change the transformation process tends towards remote collaboration. With the changing requirements and technological innovation, new possibilities arise for the design of more flexible collaboration forms less restricted by the time and space boundaries. Remote collaboration can be divided in two categories - synchronous and asynchronous. While the simultaneous presence of all stakeholders is required in synchronous collaboration, asynchronous collaborative work can be conducted independently, at different times.

The focus of this work is on asynchronous collaboration during remote manual tasks in the engineering domain. For instance, the maintenance of an on-site production machine can involve manual repair tasks through which a local worker can be guided by a remote expert for this specific machine.

For synchronous remote collaboration, many methods are recently researched, for instance video-guided systems [6, 12] and Augmented Reality (AR) labels [2, 11] in combination with speech and hand gestures [16, 24]. In global collaboration scenarios with different time-zones, it is required to support asynchronous remote collaboration. For future systems, it is recommended to provide all collaboration modes [5], which would be a revolutionary step towards more effective and efficient workflows [4].

Physical tasks in the engineering domain often rely on sharing spatial information between the collaborators and are difficult to communicate by speech, 2D images or videos [15]. In fact, using speech to communicate spatial locations and actions can be ambiguous or vague and cause confusion and error [18]. AR and Virtual Reality (VR), can help to develop new approaches for synchronous and asynchronous remote collaboration since they immerse the user into the virtual 3D space where spatial references can be communicated naturally [1, 23]. Visualization of hands and hand gestures is beneficial for remote collaboration in physical tasks, since it enhances task awareness while reducing the cognitive workload [7].

2 RELATED WORK

2.1 Asynchronous Remote Collaboration with AR

Asynchronous collaboration in Mixed Reality (MR) is a slowly growing research topic, since most of the research in the CVE area is focusing on synchronous collaboration [4, 5]. Ens et al. [5] have reviewed 110 papers about collaboration in MR published between 1995 and 2018. They found that the vast majority of papers (106, or 95%) focus on synchronous collaboration. Their findings are also backed up in the literature review of de Belen et al. [4] where a total of 259 papers between 2013 and 2018 were reviewed. They also point out that an investigation in asynchronous MR collaboration can be beneficial especially for the Architecture, Engineering, Construction, Operations and Industrial (AECOI) application areas. Since it allows the creation and retention of digital information for a later consumption, it provides the possibility to revisit previous meetings and work sessions at any time [9].

According to the reviews [4, 22], most of the papers about asynchronous collaboration in MR allow the creation and consumption of annotations, like virtual graffiti and photos which are placed at certain locations within the immersive environment and can be viewed and interacted with by other collaborators at another time [2, 11]. Irlitti et al. are researching combination methods for tangible markers and augmented annotations, which can be left for the next worker [10]. However, tasks in the AECOI area often in-

---

[*] e-mail: anjela.mayer@kit.edu
[†] e-mail: jean-remy.chardonnet@ensam.eu
[‡] e-mail: polina.haefner@kit.edu
[§] e-mail: jivka.ovtcharova@kit.edu

volve continuous spatial information, which is hard to communicate using static annotations and images.

Tseng et al. [20] present a system which not only preserves respectively correct annotations, but additionally visualizes the continuous position and orientation of the recorded user's head and hands. This provides the minimum of continuous information to perceive the movements of the user's head and hands over time. In the work of Tsang et al. [19] an AR system is developed which can record multimodal streams of annotation data, including viewpoints, voice and gesture information. After a recording is complete, users can save or replay the annotation session.

While the majority of the literature focus on general concepts providing proof-of-concept prototypes, others show how to apply asynchronous collaboration methods to specific domains. Poelman et al. [16] present a system for crime scene investigators with remote support from experts. Although their main focus is on synchronous collaboration during the investigation, the authors also discuss a recording option to support a later review of the investigation research by judges. Marques et al. [13] present a collaboration system for maintenance tasks for the industrial area. Their system enables remote experts to support on-site technicians with augmented annotations during synchronous as well as asynchronous sessions.

2.2 Manual Task Instructions in Remote Collaboration with AR

Since there are not many publications approaching manual tasks in asynchronous remote collaboration, literature addressing asynchronous and synchronous collaboration is considered in this section.

Oda et al. [15] address the challenge of spatial referencing during remote collaboration in industrial tasks utilizing virtual replicas of physical objects in a work environment. The remote expert can place annotations relative to the virtual replica and demonstrate manipulations to the object in VR. Local workers can use AR to see the annotations which are matched accordingly on the physical counterpart of the virtual replica, as well as the expert's demonstration. Although hand gestures are not utilized in the work, the authors emphasize that traditional approaches using voice or video streams are not sufficient for remote guidance, especially for operations that require spatial referencing and demonstrated actions [15].

Yound et al. [24] present an approach to share hand gestures during video calls. A local user sends their camera stream and device orientation to a remote user's VR Head Mounted Display (HMD). The remote user's hand gestures and their spatial orientation are tracked and can be shared with the local user overlaying the video. The pilot study showed an increased sense of presence compared to live video calling without sharing spatial information. However, their system does not support hand interaction with virtual or physical objects like it is required in industrial tasks. Similar approaches with shared hand gestures within a shared video stream is presented in the work of Huang et al. [7, 8], which conclude that sharing hand gestures seem to be a great improvement over previous 2D based gesture systems [7]. Sasikumar et al. [17] further found that shared gestures improve the feeling of co-presence and reduce physical workload during remote collaboration tasks.

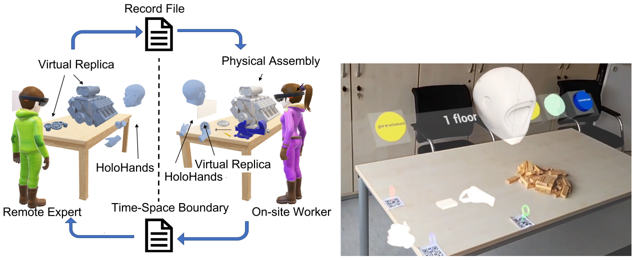
Remote collaboration approaches utilizing AR have been researched by Wang et al. [21, 23]. Their user studies have shown that gestures alone may be sufficient for simple tasks, while complex physical tasks involving small objects benefit from a combination of shared gestures and annotations. Furthermore, sharing hand gestures of a remote expert to a local user can improve the interaction, copresence and task awareness during assembly tasks [23]. In a later work, the remote user can interact with a virtual replica of the physical assembly parts and thus share interaction information with the objects additional to the hand gestures. Their results show a great potential of the shared gestures combined with Computer Aided Design (CAD) models for assembly training applications [21].

With this work, we narrow the research gap in asynchronous remote collaboration in virtual environments and present HoloHands, as a concept to share immersive hand instructions in physical tasks which are relying on spatio-temporal information. HoloHands combines the strengths of immersive hand instructions on virtual objects with benefits from asynchronous collaboration. In contrast to immersive hand instructions from the presented related work, our concept provides continuous 3D information of hand interactions with virtual replicas of physical objects. This allows not only to comprehend what should happen to the objects, but also which hand actions are required to achieve this result. To extend this concept towards asynchronous work, HoloHands can capture the physical hand interactions with virtual objects, store the recordings, and playback them with virtual hands interacting with virtual objects. This allows following users to be guided by the virtual hands' playback instead of a present collaborator as in synchronous collaboration.

Furthermore, we contribute our lessons learned from our work with the Hololens and provide helpful insights for future research.

3 CONCEPT

In this work, HoloHands - a concept for asynchronous collaboration in remote manual tasks is presented. Our concept can be used with AR HMDs for the collaboration between a remote expert providing continuous spatial instructions and a local user working on a physical task while viewing the instructions as depicted in Fig. 1. HoloHands allows for the remote expert and local user to conduct their work at different times, since the spatial instructions can be saved and exported into files which can be loaded and viewed at any time. Thus, this concept can support asynchronous collaboration on tasks involving spatio-temporal hand instructions. As depicted in Fig. 1a, the remote expert is able to see the virtual CAD representation of the physical object for which he wants to create instructions through the AR device. HoloHands supports hand tracking, thus the user can simply use hand gestures to interact with the CAD parts. While the expert is performing the maintenance task on the virtual object, his motions and interactions are captured and stored into a record. A record contains all required information to reconstruct the motions of head and hands, gestures and the interacted objects as well as their continuous position and orientation.



(a) Concept (b) AR view

Figure 1: a) Asynchronous collaboration concept using HoloHands and b) AR view on the HoloHands instructions.

The local worker, as can be seen in Fig. 1b and Fig. 1a, can open the record files with the AR application. By pressing the virtual play button, the previously recorded instructions by the remote expert can be viewed in the AR HMD. Fig. 1b shows the first person view of the local worker during replay, viewing the head movements and hand gestures as well as the interacted virtual objects of the remote expert. While watching the immersive instructions, the local user can simultaneously work on his physical task. Furthermore, the local user can also create a record while processing his task and send the record file back to the remote expert for review.

4 IMPLEMENTATION

HoloHands is based on the asynchronous capture and replay from our previous work [14], in which the concept was implemented for a VR assembly training application. In this work, the adaptation and extension is presented, which was necessary to apply the concept in AR applications capturing hand and finger motions. The implementation will be published on GitHub within an opensource template for Asynchronous Collabroation in Immersive Environments (ACIE) when a more mature state is reached.

4.1 Underlying System

The implementation of the CVE for spatio-temporal capture and replay was made in Unreal Engine 4.26, where a VR player class is mapping VR hardware sensor data to the virtual avatar and camera motions. The VR hardware consists of an HDM, which is tracking the head motions for the virtual camera rendering, and VR controllers to track hand motions and user input actions. The avatar consists of virtual hands, including three hand stances (*open*, *want to grab*, *grab*) for interaction with virtual objects. CAD models can be imported using the Unreal Engine Datasmith plugin and converted to VR-interactable objects using the generate function of the *AssemblyManager* class. The capture and replay of user movements and interactions with the virtual objects, is handled in the *ReplayManager* class, which can capture and play tracked data frame-by-frame. The data is held in record data structures at runtime and contains all required information to restore the recorded interaction, like the position, orientation and state of the hands and of the interacted objects. The *RecordExporter* class exports the records to text-based files and import files to restore the record structure. The record and replay functions of the *ReplayManager* can be triggered by pressing the corresponding 3D Buttons with the virtual hands. During replay, a ghost avatar is instantiated from the *GhostActor* class, consisting of a virtual head and hands, which are animated by the record playback. If an interaction with objects was recorded, the *ReplayManager* also creates and animates replicas of the objects. Created objects during replay have a transparent *ghost* material.

A first prototype was successfully tested in user experiments on a standalone Oculus Quest 2 VR HMD [14].

4.2 Adaptations for HoloHands

To facilitate collaborative work in AR, several adaptations had to be done. First, support for the Microsoft Hololens 2 HMD, was enabled by adding the Microsoft OpenXR plugin to the project. A more significant adaptation to the concept was enabling of the hand tracking feature. Since the Hololens does not use controllers for user input, this was a necessary step to re-enable user interaction with virtual objects. As a consequence, the recording data structure was extended by hand tracking data, including the positions and orientations of the 26 bone joints defining each hand. For replay visualization of the hand tracking, a hand mesh based on the skeletal bones was added to visualize the hand and finger motion. At last, the transparent *ghost* material was replaced by an opaque material for the head, hands and virtual objects to increase the visibility on the passthrough AR Hololens device.

5 VALIDATION

The implemented HoloHands prototype was validated in a pilot experiment including a spatio-temporal assembly task. While the instructions were preliminarily created by the experimenters, the experiment participants used these instructions to perform the assembly. The experiment supposed to show whether an assembly can be succeeded using the HoloHands instructions:

Hypothesis: *HoloHands can be used in asynchronous collaboration to communicate spatio-temporal hand instructions to succeed a manual task.*

5.1 User Experiment

For the experiment, 22 participants were recruited (10 female, 12 male) who were primarily students. Upon arrival, the subjects filled out a questionnaire, containing general questions about the participants. The vast majority (18) of the participants were 20-30 years old, two 30-40 and one younger than 20. Regarding their experience, 12 had already experimented with AR, 9 were trying AR for the first time and one was using AR on a daily basis.

Preliminarily to the experiments, the HoloHands and PDF instructions were created by the experimenter. For the HoloHands instructions, the experimenter was performing the task while wearing the Hololens 2 HMD. Therefore, the record button was pressed, and then the assembly steps were simply demonstrated on the virtual objects. The PDF instructions consisted of 2D photos of each assembly step, which the experimenter had to capture after each step and reorganize them into one instruction document in the post-processing.

During the experiments, participants were using the instructions to perform two assembly tasks. Each task consisted from 57 wood pieces of a Jenga game and had to be processed under both conditions, one with HoloHands and the other with the PDF instructions.

To see the virtual representations of the real wood pieces which were manipulated by the virtual hands in the HoloHands condition, participants had to wear the Hololens 2 HMD. The PDF instructions could be viewed on a desktop in the physical environment.

5.2 Preliminary Results

In the user experiments, the creation of the instructions was succeeded by the experimenter as well as the assembly tasks were succeeded by every experiment participant for both modalities: HoloHands and PDF. Naturally, HoloHands required more time compared to PDF, due to its playback duration. Although, the HoloHands completion times ($M = 11.78, SD = 2.05$) of the subjects are distributed in a wider range than in the PDF condition ($M = 5.17, SD = 0.99$), no significant correlation could be found between completion time and AR experience of the users.

To measure the accuracy of both conditions, the number of correctly assembled parts was counted for each task. The subjects were able to achieve maximum accuracy by placing all pieces ($N = 57$) into the correct position and orientation by the end of the task. A statistical significant difference ($Z = 2.69, p = 0.01$) with a moderate effect ($r = 0.41$) was found after carrying out a Wilcoxon signed rank test. The accuracy of the HoloHands condition was more scattered and in average lower compared to the PDF condition ($M_{HoloHands} = 45.82, M_{PDF} = 52.95$). Nevertheless, it was possible to reach maximum accuracy with both conditions during the experiment, for 2 subjects with the AR condition and 5 subjects with the PDF condition. Both subjects, who completed the tasks correctly with HoloHands, also had no mistakes in the PDF condition.

After each task the subjects were asked to fill out questionnaires to measure their subjective feedback. The included questionnaires were: System Usability Scale (SUS), NASA Task Load IndeX (NASA TLX) and three questions regarding the satisfaction with the conditions in the After-Scenario Questionnaire (ASQ). Wilcoxon signed rank test was used to compare the total scores of the HoloHands and PDF conditions. Due to the preliminary character of our work, no significant results could be found regarding the total scores, while significant results could be found for a few single questions. The detailed testing revealed a statistically significant difference with a moderate effect for SUS_4 ($Z = 2.91, p =< 0.01, r = 0.44$): *"I think that I would need the support of a technical person to be able to use this guide"* and SUS_8 ($Z = 2.54, p = 0.01, r = 0.38$): *"I found the guide very cumbersome/awkward to use"* (SUS). Furthermore, a significant difference was found in the Performance dimension of NASA TLX *"How successful were you in performing the task? How satisfied were you with your performance?"* ($Z = 3.04, p < 0.01$) of a moderate effect ($r = 0.46$) as well as the ASQ_2: *"Overall, I am* *satisfied with the amount of time it took to complete the tasks in this scenario"*, with a significant difference ($Z = 2.3, p = 0.02$) and a moderate effect ($r = 0.35$).

Additionally, the qualitative questions of our custom questionnaire revealed interesting subjective opinions. Regarding the conditions, 60% ($N = 13$) of the subjects would prefer to work with HoloHands over the PDF condition. The replay speed was "fine" for 45% ($N = 10$) of the subjects, "fast" for 9% ($N = 2$), "slow" for 23% ($N = 5$) or "too slow" for 23% ($N = 5$). Many subjects (55%, $N = 12$) had difficulties to see the virtual objects with the Hololens device. Finally, one subject had a feeling of discomfort during the task in the HoloHands condition.

More qualitative feedback, expressed by the participants, can be found in section Sect. 7 along with the implications for future work.

6 DISCUSSION ON THE EXPERIMENT RESULTS

In related work, immersive instructions were found to be more effective and efficient compared to non-immersive instructions [7,21]. Although the goal of our research is to add to their findings, at the current state of our implementation, we cannot prove nor disprove the hypotheses. Nevertheless, our prototype could successfully be used in the preliminary experiments on asynchronous assembly tasks, validating our hypothesis.

To evaluate the efficiency, for instance the completion time, in future experiments HoloHands should be compared to time-based instructions, like videos, instead to static PDF or photo instructions.

Regarding effectiveness, both conditions offer the possibility to solve the assembly tasks with maximum accuracy, which confirms that our HoloHands concept can be used for collaboration to perform on-site assembly tasks with asynchronous remote instructions. Nevertheless, the PDF instructions seem to lead to better results, although the PDF condition cannot be considered as a collaborative condition. Users who could use the HoloHands instructions with maximum accuracy were likely to achieve similar accuracy with PDF instructions. Seldom, users achieved better accuracy with HoloHands than with PDF instructions. The results regarding NASA TLX Performance confirm this conclusion, the lower the accuracy, the less satisfied the users felt with their own performance while using the HoloHands instructions.

In related work, subjects using shared gestures in remote physical tasks perceived less physical workload compared to a remote collaboration without hand gestures [17]. Since we did not find a significant difference in the Physical Demand dimension of NASA TLX between the conditions, we can neither confirm nor deny this conclusion with our experiment results.

Furthermore, the SUS questionnaire revealed that HoloHands are more "awkward/cumbersome" to use and convey the desire for technical support. Thus, we conclude that our concept is not intuitive to use at this development state. This again contradicts the findings in the related work, where the combination of 3D CAD models and hand gestures was perceived as more intuitive and expressive [21].

During the experiments, some participants expressed a "feeling of beeing bored" and didn't see the necessity of using HoloHands instructions on "such a task", "although AR was a great experience". We conclude that the chosen tasks were too easy to require and benefit from spatio-temporal instructions, like HoloHands, since each assembly action was relatively obvious to comprehend just by seeing the result, like in the PDF instructions.

7 INSIGHTS FOR THE USE OF AR IN ASYNCHRONOUS COLLABORATION

During our research, we collected valuable qualitative data which will help to improve our concept and prototype, and which we also want to share with fellow researchers. Based on comments given during the experiments and insights from our preliminary experiment, our lessons learned and implications for future work are summarized in this section.

7.1 Insights on the HoloHands Concept

In our validation, (64%) of the users would prefer to have both modalities, with their own benefits for assembly tasks. Ideally, having the flexibility of the PDF instructions, where the users are in control over the assembly pace, combined with the ability in AR to visualize the final location of each part in space without having to rethink the orientation of the plan and objects. Additionally, some participants disagreed whether the position and orientation of the pieces could be better perceived in the PDF or AR modality, leading to our recommendation to provide static simplified instructions for an assembly preview combined with spatio-temporal instructions showing the process and hand actions to get to the previewed result.

Occlusion and Visibility of the Instructions

A vast majority of users (82%, N = 18) who tested our application reported difficulties to see the virtual objects. One reason was occlusion by other virtual objects. Since the avatar instructions were recorded and replayed from the same position, it was likely that the avatar head would move into the user's view in front of the virtual objects on the table. Due to the transparent appearance of all virtual objects in the Hololens, objects cannot be distinguished from each other during occlusion. In particular, this effect is more likely to occur for users who are smaller compared to the user who recorded the avatar. Due to the transparency, working with the physical environment was still possible, but also cumbersome. Furthermore, the interaction with occluded virtual objects was nearly impossible, and the instructions could not be followed.

Some users repositioned themselves to unergonomic poses to be able to follow the hand instructions. As a result, 82% of the users would rather not see the head at all. On the other hand, users who were smaller than the avatar did not report any difficulty. Anyhow, they were likely not to see the head of the avatar since it was located above their field of view.

Similar issues leading to proxemic violations were also reported in [3]. As a prevention, the user is not allowed to teleport to the same position as the replayed avatar. We furthermore suggest, to spawn the avatar beside the user during replay to avoid proxemic violation and occlusion by the avatar.

77% (N=17) of the participants expressed their desire to be able to distinguish objects being manipulated by the ghost from other objects within the virtual environment. In our experiment, all assembly pieces had the same geometry and texture, and thus grabbed pieces were difficult to distinguish from others. Furthermore, similar to the avatar head, overlapping or occluded pieces were hard to recognize.

In [3] highlighting of what others see was recommended to display relevant objects with a higher visual salience. To enhance the visibility of virtual objects in passthrough AR, we also suggest three improvements: First, the outlining of object edges should improve the visual differentiation of overlapping objects. Based on the user feedback, we recommend to additionally highlight grabbed objects, for instant by a different colour, to indicate the interaction. Finally, a location indicator for the hand instructions should be added when they are out of the user's field of view.

Control and Preview Options

While the PDF instructions were described as easy to control, the AR instructions were experienced as "intuitive" to use and "more ergonomic". However, three participants were unsatisfied with the speed of the AR instructions, while one participant perceived the replay speed as "good". However, we recommend adding more playback options to provide more control to the users during the instruction replay, like the selection of an individual playback speed as well as rewind and skip functions. Furthermore, since it was helpful for two participants to preview next steps in the PDF instructions and see the final assembly, we also suggest considering preview options when designing immersive AR instructions. Chow et al. [3] have a similar conclusion and recommend extending the playback timeline by visual or semantic delimiters additionally.

Accuracy of the gesture-based interaction

For the interaction with virtual objects, gesture-based interaction was implemented utilizing the Hololens hand tracking. In our preliminary tests, accuracy issues were encountered involving the placement of virtual objects. Due to the frame-based gesture recognition, it was significantly difficult to precisely superimpose virtual objects with their physical counterparts. Grabbed virtual objects are attached to the virtual hand and can be moved by moving the hands. To release a grabbed object and place it in the VE a gesture (opening the hand) must be recognized by the system. This results in a delay between the user's intention to release the virtual object until it is actually released. During this delay, the virtual object is still moved by the hand, leading to an offset to the intended destination. The gesture-based placing accuracy was tested by trying to superimpose two identical objects as depicted in Fig. 2a.

Overall, 40 measures were conducted for three different geometries by the same user. During the tests, no object could be placed to precisely superimpose the reference object. Furthermore, accuracy seem not to correlate with the time spent for a test run. As expected, the sphere objects required less time for superimposition, since the orientation didn't have to be considered. With increasing geometrical complexity, superimposition required more time. The results obtained during the tests were spread over a range of 4 cm for positioning and 10° for orientation of the virtual objects. Regarding the rotations, the deviations are mostly between 0° and 2.5° For the position, the average positioning difference is around 1 cm. We conclude that the gesture-based interaction is not satisfying for applications relying on precise interaction with virtual objects, like in industrial assembly applications with complex shapes, like gears, casings or nuts. In [8] video-streamed hand interactions were not sufficient for precise interaction on small objects, we can confirm that gesture-based interaction is also struggling with issues. In future work, other interaction concepts for precise placing of virtual objects should be explored. A possibility is to add precise collision volumes to the virtual objects, enabling precise push interactions by the virtual hands. However, the creation of precise collision volumes requires more time and computing power.

7.2 Insights on the Prototype Implementation

In the prototype implementation, recording of tracking data is frame-based, which caused several issues during our experiments. Different devices may support different framerates and thus the replay speed of the recordings will vary accordingly. In our case, a cross-platform test was conducted, including an Oculus Quest 1 VR and the Hololens 2 AR devices. On the Hololens the recording which was previously created with an Oculus Quest was played back in a much slower pace caused by the lower framerate of the AR device. To enable a replay in real time, the delta time between frames should be recorded and considered during playback. This would also enhance the cross-platform capability of the system.

Another issue is the blind spot between frames, which results in inaccurate object placing. Missing data during gesture recognition, especially when releasing a grabbed object, leads to a small delay of the release action. As a result, the object will continue moving until it is eventually released. Consequently, a precise placing of a virtual object is nearly impossible, as discussed in Sect. 7.1.

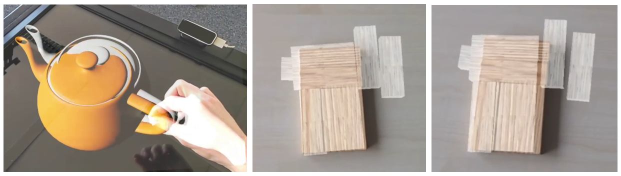
As workaround, recording can be toggled by grab and release interactions, which significantly adds to the complexity of the system.

7.3 Insights on the Hololens Device for AR Collaboration

During the user experiments, several problems occurred related to the Hololens 2 device, in particular during the placement of virtual objects and issues with specific models of glasses.

Shifting position of virtual objects

The Hololens relies on the continuous mapping of its environment via depth sensors, which are affected by factors, such as light intensity, reflections and plain environments. Thus, the localization quality of the HMD within the physical environment varies, which can lead to inaccurate cartography and therefore result in inaccurate localization of the holograms. While the Hololens is constantly improving its orientation within the physical space, it is not uncommon that depending on the user's point of view adopted the virtual objects can be shifted, which we observed during our experiments as depicted in Fig. 2b and Fig. 2c. For tasks which rely on spatial accuracy, this characteristic is very problematic because the user never knows which is the correct position of the object. In applications relying on pose precision of virtual objects, we recommend using active localization on the Hololens. However, this either adds complexity to the system when implemented with optical markers or requires more computing power if object detection algorithms are utilized. Finally, if the HMD is correctly localized, it must also be correctly calibrated to the user's interpupillary distance (IPD). The IPD is highly individual and is also affected by the orientation of the HMD on the user's head. It is recommended to calibrate the Hololens before each use to ensure its orientation and IPD are correct for this instance. Furthermore, to utilize image recognition, Microsoft recommends using the same angle for the images to limit errors. During our work, marker detection was used to localize the VE, as well as the IPD calibration for every use. For tasks relying on spatial accuracy, the use of Hololens requires precise prerequisites, making it troublesome to use in everyday work.



(a) Superimposition (b) Frame t (c) Frame $t+1$

Figure 2: a) Accuracy tests of gesture-based grabbing and position shift of holograms between two render frames, b) and c).

Issues with specific glasses

Officially, Hololens supports being used with glasses, however only glasses with a single focal length seem to be correctly supported [55]. During our experiments, several issues were reported by subjects wearing glasses (45%, N=10) although being calibrated for each participant. The participants reported a narrow field of view, which was probably additionally narrowed by some glasses. Depending on the volume of the glasses frame, it was not possible to optimally position the HMD on the subject's head, which can affect the spatial localization of the virtual objects, as was previously discussed. One participant reported a "slight tingling in the eyes". Another participant was not able to see the correct colors, reporting that the "green and purple colors were irritating" and assumed a connection to the coating of the glass lenses. Some glasses integrate light filters, to reduce reflections, consequently that affects the color perception of the virtual objects in the Hololens. Finally, the Hololens uses the IPD in combination with the eye gaze to optimize the positioning of the virtual objects on the display. Glass lenses may induce deviations in the measurement which leads to a misplacement of the projection and thus a blurred or double vision for the user. A workaround to mitigate these effects while allowing people with visual impairments to continue using the Hololens is switching to contact-lenses. However, this is not always possible nor desirable by the users. We recommend considering problems that can occur due to glasses of the users, like more narrowed field of view, color deviations, shift of objects due to a mismeasurement of the IPD during the design of virtual experiences with the Hololens.

8 CONCLUSION AND FUTURE WORK

The presented HoloHands concept for asynchronous collaboration in remote manual tasks facilitates the creation of immersive instructions for manual tasks using AR HMDs and can be used as a communication method in asynchronous collaboration. The concept was validated in a preliminary experiment, comparing it to a traditional method for static 2D assembly instructions. In the experiment, all participants were able to finish the tasks using HoloHands, approving that the concept can successfully be used for asynchronous communication of hand instructions. In this paper, we outline the lessons learned during our research and make important implications for future work with spatio-temporal instructions and the Hololens device. Further evaluation is needed concerning the creation process of the HoloHands instructions, as well as the whole collaboration cycle, with a focus on asynchronous interaction between the participants. Since immersive instructions can outperform traditional instructions in spatial tasks, we believe that HoloHands can become an efficient and effective approach for asynchronous remote collaboration after the implementation of the outlined improvements.

The authors would like to thank the Ministry of Science, Research and Arts of the Federal State of Baden-Württemberg for the financial support of the projects within the InnovationsCampus Future Mobility (ICM).

REFERENCES

[1] M. Alibali. Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation - SPAT COGN COMPUT*, 5:307–331, 12 2005. doi: 10.1207/s15427633scc0504_2

[2] H. Chen, A. S. Lee, M. Swift, and J. C. Tang. 3d collaboration method over hololens™ and skype™ end points. In *Proceedings of the 3rd International Workshop on Immersive Media Experiences*, ImmersiveME '15, p. 27–30. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2814347.2814350

[3] K. Chow, C. Coyiuto, C. Nguyen, and D. Yoon. Challenges and design considerations for multimodal asynchronous collaboration in VR. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019. doi: 10.1145/3359142

[4] R. A. de Belen, H. Nguyen, D. Filonik, D. Favero, and T. Bednarz. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. *AIMS Electronics and Electrical Engineering*, 3:181–223, 2019. doi: 10.3934/ElectrEng.2019.2.181

[5] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131, 2019. doi: 10.1016/j.ijhcs.2019.05.011

[6] S. R. Fussell, L. D. Setlock, and R. E. Kraut. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, p. 513–520. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/642611.642701

[7] W. Huang, L. Alem, F. Tecchia, and H. B. L. Duh. Augmented 3d hands: a gesture-based mixed reality system for distributed collaboration. *Journal on Multimodal User Interfaces*, 12:77–89, 6 2018. doi: 10.1007/s12193-017-0250-2

[8] W. Huang, M. Billinghurst, L. Alem, and S. Kim. Handsintouch: Sharing gestures in remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, OzCHI '18, p. 396–400. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3292147.3292177

[9] A. Irlitti, R. T. Smith, S. V. Itzstein, M. Billinghurst, and B. H. Thomas. Challenges for Asynchronous Collaboration in Augmented Reality. *Adjunct Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2016*, pp. 31–35, 2017. doi: 10.1109/ISMAR-Adjunct.2016.0032

[10] A. Irlitti, S. Von Itzstein, L. Alem, and B. Thomas. Tangible interaction techniques to support asynchronous collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–6, 2013. doi: 10.1109/ISMAR.2013.6671840

[11] S. Kasahara, V. Heun, A. S. Lee, and H. Ishii. Second surface: Multi-user spatial collaboration system based on augmented reality. In *SIGGRAPH Asia 2012 Emerging Technologies*, SA '12, p. 1–4. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2407707.2407727

[12] R. E. Kraut, S. R. Fussell, and J. Siegel. Visual information as a conversational resource in collaborative physical tasks. *Human–Computer Interaction*, 18(1-2):13–49, 2003. doi: 10.1207/S15327051HCI1812_2

[13] B. Marques, S. Silva, A. Rocha, P. Dias, and B. S. Santos. Remote asynchronous collaboration in maintenance scenarios using augmented reality and annotations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 567–568, 2021. doi: 10.1109/VRW52623.2021.00166

[14] A. Mayer, T. Combe, J.-R. Chardonnet, and J. Ovtcharova. Asynchronous manual work in mixed reality remote collaboration. In L. T. De Paolis, P. Arpaia, and M. Sacco, eds., *Extended Reality*, pp. 17–33. Springer Nature Switzerland, Cham, 2022.

[15] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software*, UIST '15, p. 405–415. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2807442.2807497

[16] R. Poelman, O. Akman, S. Lukosch, and P. Jonker. As if being there: Mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, p. 1267–1276. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2145204.2145394

[17] P. Sasikumar, L. Gao, H. Bai, and M. Billinghurst. Wearable remotefusion: A mixed reality remote collaboration system with local eye gaze and remote hand gesture sharing. pp. 393–394. Institute of Electrical and Electronics Engineers Inc., 10 2019. doi: 10.1109/ISMAR-Adjunct.2019.000-3

[18] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci. Bethere: 3d mobile collaboration with spatial input. CHI '13, p. 179–188. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2470679

[19] M. Tsang, G. W. Fitzmzurice, G. Kurtenbach, A. Khan, and B. Buxton. Boom Chameleon: Simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. 2003.

[20] P. Y. Tseng, H. Haraldsson, and S. Belongie. Annotate all ! A Perspective Preserved Asynchronous Annotation System for Collaborative Augmented Reality. pp. 1–3, 2019.

[21] P. Wang, X. Bai, M. Billinghurst, S. Zhang, S. Wei, G. Xu, W. He, X. Zhang, and J. Zhang. 3dgam: using 3d gesture and cad models for training on mixed reality remote collaboration. *Multimedia Tools and Applications*, 80:31059–31084, 8 2021. doi: 10.1007/s11042-020-09731-7

[22] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, and H. Ji. Ar/mr remote collaboration on physical tasks: A review. *Robotics and Computer-Integrated Manufacturing*, 72:102071, 2021. doi: 10.1016/j.rcim.2020.102071

[23] P. Wang, S. Zhang, X. Bai, M. Billinghurst, W. He, M. Sun, Y. Chen, H. Lv, and H. Ji. 2.5dhands: a gesture-based mr remote collaborative platform. *International Journal of Advanced Manufacturing Technology*, 102:1339–1353, 6 2019. doi: 10.1007/s00170-018-03237-1

[24] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE Transactions on Visualization and Computer Graphics*, 25:1908–1918, 5 2019. doi: 10.1109/TVCG.2019.2898737