Interaction with Virtual Content using Augmented Reality: a User Study in Assembly Procedures

BERNARDO MARQUES, JOÃO ALVES, MIGUEL NEVES, RAFAEL MAIO, INÊS JUSTO, ANDRÉ SANTOS, RAQUEL RAINHO, DANY COSTA, CARLOS FERREIRA, PAULO DIAS, BEATRIZ SOUSA SANTOS, IEETA, DETI, University of Aveiro

Assembly procedures are a common task in several domains of application. Augmented Reality (AR) has been considered as having great potential in assisting users while performing such tasks. However, poor interaction design and lack of studies, often results in complex and hard to use AR systems. This paper considers three different interaction methods for assembly procedures (Touch gestures in a mobile device; Mobile Device movements; 3D Controllers and See-through HMD). It also describes a controlled experiment aimed at comparing acceptance and usability between these methods in an assembly task using Lego blocks. The main conclusions are that participants were faster using the 3D controllers and Video see-through HMD. Participants also preferred the HMD condition, even though some reported light symptoms of nausea, sickness and/or disorientation, probably due to limited resolution of the HMD cameras used in the video see-through setting and some latency issues. In addition, although some research claims that manipulation of virtual objects with movements of the mobile device can be considered as natural, this condition was the least preferred by the participants.

CCS Concepts: • **Human-centered computing → Mixed / augmented reality**; *User studies*; *Usability testing*; • **Applied computing** → Computer-aided manufacturing;

Keywords: Augmented Reality; Assembly Guidance; Interaction; Touch Gestures; Device movement; Mid-air Gestures - Controllers, User Study.

ACM Reference Format:
Bernardo Marques, João Alves, Miguel Neves, Rafael Maio, Inês Justo, André Santos, Raquel Rainho, Dany Costa, Carlos Ferreira, Paulo Dias, Beatriz Sousa Santos. 2020. Interaction with Virtual Content using Augmented Reality: a User Study in Assembly Procedures. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4, ISS, Article 196 (November 2020). ACM, New York, NY. 17 pages. https://doi.org/10.1145/3427324

1 INTRODUCTION

Augmented Reality (AR) is a promising technology for guidance in assembly procedures. Through AR, it is possible to display relevant information [3] in the field of view of users while performing assembly procedures. This includes step-by-step instructions, 3D illustrations, or other relevant data for the on-going task. [26]. Numerous studies recognize that AR promotes learning and facilitate effective training [25, 32, 38, 39, 46]. Moreover, AR has also been shown to outperform Virtual Reality (VR), which in turn outperform conventional paper instructions by a great margin [29].

Author's address: Bernardo Marques, João Alves, Miguel Neves, Rafael Maio, Inês Justo, André Santos, Raquel Rainho, Dany Costa, Carlos Ferreira, Paulo Dias, Beatriz Sousa Santos, bernardo.marques@ua.pt,paulo.dias@ua.pt,bss@ua.pt, IEETA, DETI, University of Aveiro, Aveiro, 3810-193.

2573-0142/2020/November-ART196 $15.00
https://doi.org/10.1145/3427324

These advantages, combined with the emergence of more affordable and powerful devices [12], make AR particularly suitable for assembly guidance in different areas of application [4, 27, 50].

The usefulness of AR is directly related to the user interfaces and interaction methods AR can provide. In this context, the study of innovative interfaces and interaction methods is of utmost importance since it directly affects the effectiveness and efficiency of users [20, 21, 27]. Currently, it is possible to identify at least six different categories of AR-based interfaces: traditional 3D user interfaces, natural user interfaces, tangible AR interfaces, emerging multimodal AR interfaces, collaborative AR interfaces and finally hybrid AR interfaces [6, 27, 50].

Traditional 3D User Interfaces

Most traditional 3D User interfaces require the use of input devices like motion tracking sensors (with 6 DOF). These allow tracking of various physical objects including the user's body motion, and let users point, select or manipulate virtual objects. One of the most prominent problems is that the methods used for interacting with virtual objects are different from interacting with physical objects, where users mainly use their hands for direct manipulation [6].

Tangible AR Interfaces

To minimize the virtual/real object manipulation of traditional 3D User Interfaces, a possible solution is the use of tangible interfaces in which virtual objects are assigned to physical ones to ease interaction. Virtual objects can then be interacted with by manipulating the corresponding tangible objects in an intuitive, natural and seamless way. Billinghurst et al. developed a set of design principles that tangible AR interfaces should follow to be intuitive to use and provide seamless interaction with virtual content [5, 8]:

- Support 3D interaction, allowing to move, rotate and approximate virtual objects;
- Support the use of physical objects for manipulating virtual content;
- Match the physical constraints of the object to the requirements of the task;
- Support manipulation of multiple objects at the same time.

Natural User Interfaces in AR

More recently, support to natural user interfaces in AR, using bare hand input (mid-air interaction) has been actively investigated. Advance in computer vision technology enable recognizing users body motion and gestures in real time without requiring the user to wear any sensors. Despite its simplicity, mid-air interaction is considered tiring for long-term use and is prominently used in the field of motor dysfunction assessment and rehabilitation [6, 27].

Multimodal AR Interfaces

To provide richer interactivity in AR, there have been efforts to combine different modalities of input into multimodal interfaces. Multimodal interfaces are often preferred by users [45] and support natural human language and behavior such as speech, touch or natural hand gestures as interaction methods. Users might need to use additional Hardware (like gloves or special input tool in their hands or body) or not (in mid-air interaction for example). Some research also studied the use of small cameras (incorporated into eye-wear) to determine gaze from eye monitoring, an interesting alternative that require additional calibration and filtering to be robust to involuntary eye movements [28]. Multimodal interfaces are interesting as they offer the freedom to select the preferred mode of interaction depending on the context, and provide intuitive ways of interaction with AR (visual and audio) cues superimposed on the real world [6, 45].

Collaborative AR Interfaces
Collaborative interfaces are used for both face-to-face and remote collaboration and aim to support communication between users. In face-to-face collaboration, these interfaces are used to enhance the shared physical world and create an interface for computer supported cooperative work. Similarly, in remote collaboration, collaborative interfaces may enhance communication using gaze, visual, audio and other non-verbal cues like gestures, and/or manipulation of virtual objects [5, 7, 27].

Hybrid AR Interfaces
Hybrid interfaces allow the use of multiple displays and different interaction devices, depending on the user's current needs and real-world context. Hybrid user interfaces should be built upon a flexible infrastructure that allows the use of several input devices and interaction techniques [5, 41, 50].

Although most AR user interfaces focus on selection and manipulation of virtual content, poor interaction design and lack of studies make most AR systems difficult to use [21, 31] requiring further research on the topic [16, 27].

This paper presents a user study comparing three interaction methods in AR for assembly scenarios. We selected two interaction methods based on Handheld AR (touch gesture and device movement) as mobile devices present very interesting characteristics: availability, low cost, ease of use, user acceptance based on daily use and presence of multiple sensors. The other method evaluated is based on a video see-through HMD and 3D controllers, a configuration offering hands-free capabilities and a more immersive experience. To compare these methods, we performed a user study with 27 participants performing an assembly task using Lego brick pieces, evaluating performance, ease of use and acceptance. The rest of the paper describes related works, the experimental design and setup, and a discussion of the results, as well as conclusions and ideas for future work.

2 AUGMENTED REALITY IN ASSEMBLY GUIDANCE

Several research works have been performed using different AR-based methods to assist assembly procedures. To structure our literature review, we identify three main components: the AR experience provided, the interaction methods used and the feedback given to the users (Figure 1).

2.1 AR experiences

Different assembly experiences can be provided to the users by seamlessly combining real world with various computer-generated contents. These experiences can be provided though the use of See-through AR, Indirect AR, Handheld AR or Spatial AR. These experiences can benefit from tracking mechanisms to detect assembly errors in real-time and monitor the current state of the assembly sequence, even though only a reduce number of works has explored such mechanisms [2].

Tang et al. (2003) explored See-through AR HMD to present assembly instructions. A user study with 75 participants was conducted using brick blocks. Participants had to complete an assembly task following instructions in different formats: 1- Printed media, 2- Instructions on a monitor, 3- Instructions on a See-through HMD, 4- Spatially registered AR instructions on a See-through HMD. Results showed that AR-based systems can improve task performance (lowest average time and number of errors) and reduced mental workload on assembly tasks compared to other media [44].

Loch et al. (2016) studied Indirect AR integrated in workstation using a camera to track the user's workflow, following the assembly procedures automatically. The study involved 17 students using brick blocks comparing the system with video-based assistance regarding performance, user acceptance and mental workload. Results showed improvements in accuracy, task performance and reduction in the number of errors and task time when AR was used [34].

Recently, Lai et al. (2020) introduced a system consisting of multi-modal AR-based instructions, with the support of deep learning for tool detection. The multi-modal AR rendering may provide various on-site instructions as texts, videos or 3D animations. The detector is developed using a Faster R-CNN model trained on a CAD-based synthetic tool dataset, which detects real physical tools with an average precision of 84.7%. To demonstrate its ability in assisting human operators to perform complex assembly tasks, a case study was conducted focusing on a mechanical assembly of a CNC carving machine. Results showed that the system helped reduce the time and errors of the given assembly tasks by 33.2% and 32.4%, respectively [30].

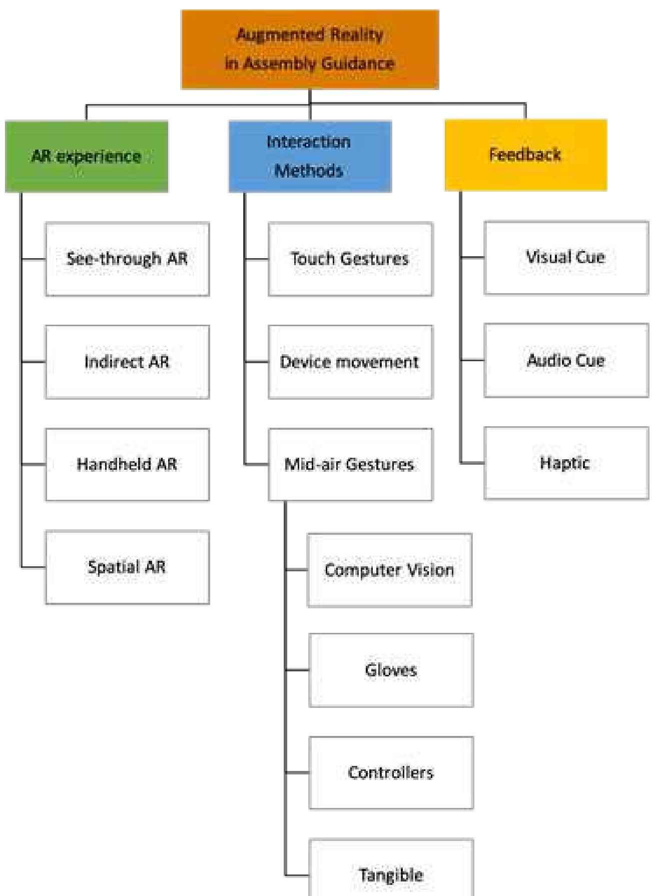

Fig. 1. Main components of Augmented Reality in assembly guidance.

Nishihara and Okamoto (2015) presented one of the first solutions using Handheld AR (instead of the traditional AR-enabled See-through HMD) with image analysis and processing for object recognition and display of assembly instructions. The solution was tested using a Pentamino Puzzle. The work focused mainly on the image processing algorithm that obtained a 98% success rate using a recognition based on the combination of the canny edge detection algorithm, morphology closing and border following algorithm. However, no formal evaluation with users was performed [37].

Funk et al. (2017) conducted a long-term study using Spatial AR in an industrial assembly hall, considering expert and untrained workers. Their results indicate a decrease in performance for expert workers, despite an increase in the perceived cognitive load. Regarding the untrained workers, the use of Spatial AR in-situ instructions was considered useful during the learning phase [18].

More recently, Alves et al. (2019) also explored the use of Spatial AR for assembly guidance using a validation process. A study to compare different AR solution regarding user performance, ease of use, and acceptance was conducted with 15 students using brick blocks. Results showed participants made fewer errors and were significantly faster using the Spatial AR condition. Moreover, a Nasa TLX rating also showed that the Spatial AR condition had a slightly lower cognitive load on the participants [2].

2.2 Interaction methods

Another important component of any AR guidance system is the type of interaction used to complete the assembly procedures. Interaction can be provided through: touch gestures, device movement, mid-air gestures, controllers, gloves or tangible mechanisms [20].

The widespread use of touch screens propelled by the democratization of mobile devices makes touch gestures one of the most popular input method for virtual object manipulation. These gestures can be mapped from a 2D space (screen) to 3D world transformations as depicted in Figure 2 - (1) [20, 33, 35].

Another natural way to interact with virtual objects can be achieved by following the real movements of the user/device in 3D space [20]. Henrysson et al. (2005) proposed this method, where the object position changed while the user moved the device, allowing to manipulate virtual objects as illustrated in Figure 2 - (2) [23].

Mid-air interaction can be a powerful interaction mechanism as shown in Figure 2 - (3). Through the years, several researchers have explored this interaction method using finger, one hand, both hands and body detection approaches with different levels of complexity. The connection between the user mid air gestures and the AR environment can be captured/performed using computer vision, controllers, gloves or tangible mechanisms [20].

Radkowski et al. (2012) explored assembly procedures using hand tracking and hand gesture recognition, without using a graspable device, following the generic conceptualization shown in Figure 2 - (3.1). The mid air gestures allowed 3D object manipulation. Thus, a user can select, manipulate, and assemble 3D models of mechanical systems. The authors also introduced a set of interaction techniques as a direct mode and a precise mode. The direct mode allowed fast translation. The precise mode facilitated a precise placing of virtual parts [40].

Another example was proposed by Buchmann et al. (2004) explored the use of gloves to track the user's fingers gestures and manipulate a virtual object in the augmented environment. More recently, Hayatpur et al. (2019) explore the use of shape constraints to enable quick and precise manipulation of virtual objects through gloves [22]. Both approaches included different types of gestures: pointing, navigating and command gestures, following the generic conceptualization depicted in Figure 2 - (3.2) [9].

Murakami et al. (2013) explored the use of controllers during the performance of assembly tasks, following the generic conceptualization shown in Figure 2 - (3.3). The authors state users can perform assembly activities that need large space through the use of controllers and that haptic feedback can be an effective mechanism to assist them [36]. Likewise, Caputo et al. (2018) proposed the "Smart Pin" approach, allowing users to select, translate, rotate and scale objects relying entirely on the positional tracking of a single hand holding a controller. The method was compared to a two-handed manipulation technique (Handlebar). Results showed that most users preferred the Smart Pin approach for its gestural comfort and ease of use [11].

Yuan et al. (2008) explored tangible interaction through a pen-like object with a certain colour distribution. The authors tracked the pen in real-time and used it to trigger relevant buttons and manipulate virtual content during assembly guidance [49]. A generic conceptualization of this method of interaction is illustrated in Figure 2 - (3.4).

In fact, all the interaction methods can be used individually, but some scenarios might require to combine them in order to address specific requirements, users or use of specific devices.

2.3 Feedback provided

The last component addresses the type of feedback given to the users during assembly procedures, whether to present instructions or provide hints in the case of errors (if validation mechanisms are used). Feedback can be provided in different forms: visual cues, audio cues or haptic devices.

Büttner et al. (2016) presented an in-depth analysis comparing three methods: paper instructions (baseline), See-through AR and Spatial AR. For their study, the authors recruited 13 participants, used brick blocks as case study and visual cues as the feedback mechanism. Spatial AR was used to indicate the box containing the relevant parts. The assembling instructions were displayed as a static image in the center of the users's view in the HMD (using See-through AR) [10].

Webel et al. (2013) presented a solution for assembly procedure based on Handheld AR using visual and audio cues and also haptic feedback through a bracelet. The authors explored virtual post-its as a mechanism to provide instructions and feedback. These could be linked to contextualize information and provide multimedia feedback through video and audio. Likewise, it was also possible to use the bracelet to provide different forces/torques associated to specific stages of the assembly procedure [48].

More recently, Siew et al. (2019) presented an AR-based system to provide adaptive support and feedback to users. The system relies on an image-based step detection module to understand on whether a step has been conducted correctly and uses a wrist-based haptic tracking to provide useful feedback and awareness to the users within the workspace. The system has been demonstrated through a case study with positive results, allowing the user to be more effective and willing to accept guidance information during a maintenance process [42].

In summary, previous works show the potential of different AR characteristics for assisting users during assembly procedures. Although a number of prototype systems using AR have been proposed, limitations still exist when assisting complex assembly procedures, including intuitive user interface, development of proper interaction methods, and lack of user studies [47].

3 PROTOTYPE FOR INTERACTION IN AUGMENTED REALITY

To compare different interaction methods using AR, we implemented a prototype to guide users during a virtual assembly procedure, following previous work [1]. The prototype uses a mobile device or an AR-enabled video see-through HMD to display the assembly instructions. We consider these settings as they only require a single device (mobile or HMD) and no additional tracking or hardware capabilities as the glove (3.2) and tangible interaction (3.4) conditions. The perspective in which the instructions are shown is identical in all methods. During the assembly, only visual feedback is provided to the users. The task studied consists in assembling a virtual model using Lego brick block of different shapes and sizes. The system allows to use the three different methods (Touchscreen gestures using a mobile device; Mobile Device movement; See-through HMD and Controllers) to align virtual pieces in a predefined position and orientation indicated by a transparent view of the piece in its final location.

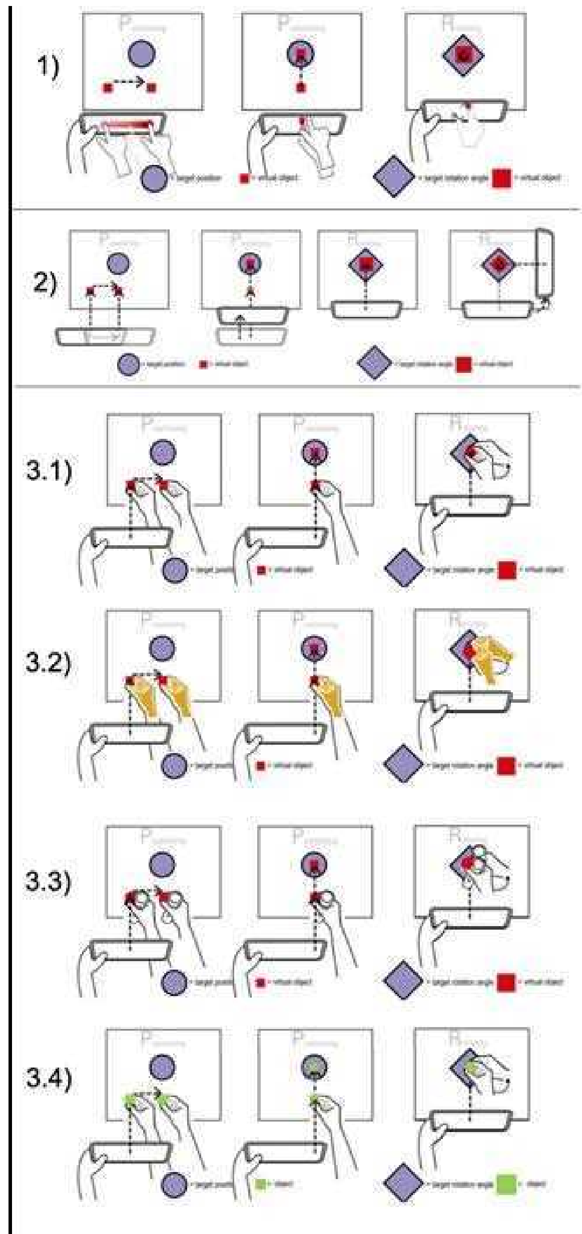

Fig. 2. Interaction example of Handheld AR using touch gestures (1), device movement (2) and mid-air gestures (3) in the form of computer vision (3.1), gloves (3.2), controllers (3.3) and tangible interaction (3.4). Inspired and adapted from: [20].

3.1 Method 1 - interaction using Touchscreen gestures / Touchscreen

The first interaction method uses a mobile device to augment the environment through the detection of a pre-defined marker, as depicted in Figure 3. Users can select the virtual object from a pre-defined set and move it around the environment in two dimensions by dragging the finger in the screen, keeping the depth of the objects locked as well as it's rotation. These properties can be changed using buttons that rotate the object ninety degrees and change the depth of object by a fixed amount (0.50 cm). It is also possible to manipulate several objects at the same time, by selecting different objects successively.

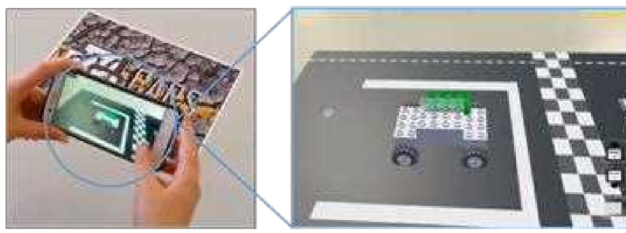

Fig. 3. Method 1 - Touch gesture interaction of virtual content using AR in a mobile device.

3.2 Method 2 - interaction using the Device's Movement

The second method also allows the selection of virtual objects through touch in a mobile device but the object pose is controlled by the device movement, as depicted in Figure 4. Once the virtual object is selected, a bound is created between the piece and the device. This bound is a fixed geometric relation implying that all six degrees of freedom of the object are anchored to the device. As such, this method gives the user total 6 DOF control in the interaction process.

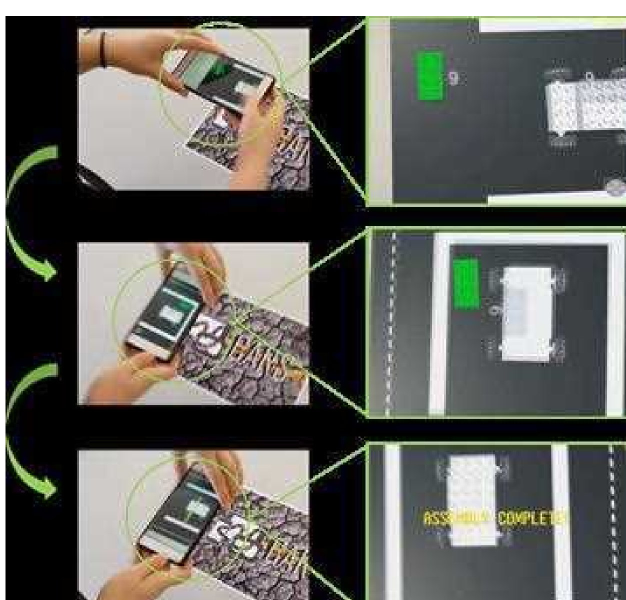

Fig. 4. Method 2 - Using movement of a mobile device to manipulate virtual content.

3.3 Method 3 - interaction through HMD and Controllers

The third method explores the HTC Vive Pro HMD and its 3D controllers capabilities to manipulate virtual objects, as shown in Figure 5. Instead of using markers like in the previous methods, the prototype uses the HMD cameras (90fps at a 480 pixel resolution) to capture the real-world environment and merge it with the virtual scene, thus creating an augmentation of reality. The controllers provide a natural 6DOF interaction mechanism, allowing to select objects by 'press' and 'hold' the controllers trigger respectively and move them in the surrounding environment attached to the controller.

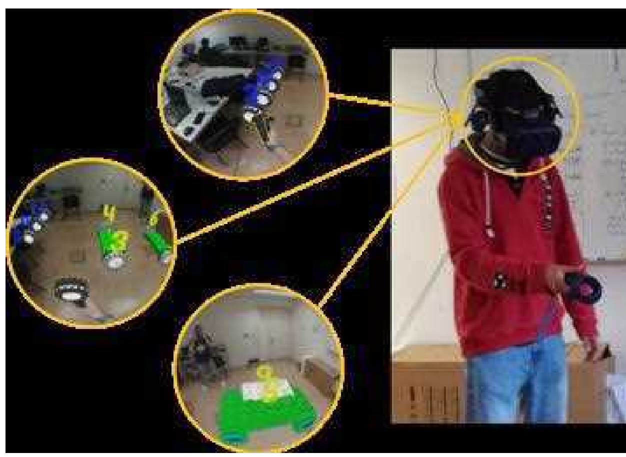

Fig. 5. Method 3 - Using a See-through AR-enable HMD and its controllers to manipulate virtual content.

In all interaction methods, a color change of the selected piece indicates a correct placement near its final desired position and orientation. To end the interaction when close to the desired pose using the different interaction methods, rotation and translation thresholds were empirically estimated. We set the empirically thresholds of a successful match to 8 degrees for rotation and 0.01 meters for translation. The prototype was developed using the Unity 3D game engine, combined with specific SDKs for the different devices used. We also used the Vuforia library to recognize predefined markers, thus placing the virtual content in the real-world environment.

4 SETUP AND METHOD

This study aimed to compare the usability and acceptance of the three AR-based assembly guidance methods.

4.1 Experimental Design

A within-group experimental design was used. The null hypothesis (H0) considered was that the three experimental conditions are equally usable to mount a pre-defined brick block structure. The independent variable was the interaction method provided to the users, with three levels corresponding to the experimental conditions: 1- Touch gestures using a mobile device (C1-Touchscreen): The user performed the assembly using touch and gesture in the mobile device screen to manipulate the pieces to their desired pose (Figure 6 - Left); 2- Mobile Device movement (C2-Movement): The user performed the assembly with movements of the device in 3D space. Thus, the piece's pose changed while the user moved the device (Figure 6 - Middle); 3- Controllers and the See-through

HMD (C3-HMD): The user performed the assembly using the controllers from the AR-enabled See-through HMD, allowing to move the pieces to their correct pose (Figure 6 - Right).

Fig. 6. Study interaction conditions using: C1 - Touchscreen gestures using a mobile device (Left); C2 - Mobile device Movement (Middle); C3 - Controllers and the See-through HMD (Right).

Performance measures and participants' opinion were the dependent variables. The order in which the conditions were performed, as well as participants' demographic data and previous experience with AR and assembly were registered as secondary variables. To minimize learning effects during the experiment, the participants were split into three groups and each group performed the three conditions in different orders.

4.2 Tasks

The participants were required to assemble a single virtual car using Lego brick with all three experimental conditions. The virtual car is made of 9 pieces, representing a generic assembly process (applicable to general assembly tasks rather than assembly tasks in specific domains) with a pre-defined solution (Figure 7), aiming to minimize bias towards a population with expertise in a specific domain. The assembly process consisted of 9 procedural instructions that were 3 dimensional in nature. During each step, participants were required to place the brick piece in a pre-defined pose (specific position and orientation) shown visually by a transparent Lego brick.

4.3 Measures

The data collection was conducted under the guidelines of the Declaration of Helsinki. Two types of measurements were taken: task performance, and participants' opinion. Task performance is defined based on the time of completion (logged by the device, measured in seconds). We considered time required to place each specific brick piece in the correct pose and the total time required to assemble the full model/car.

Participants' opinion was obtained through a post-task questionnaire (taking into account the works by [13, 15]), including: demographic information (age, gender, previous experience with VR, AR, and AR in assembly tasks) and questions concerning the three conditions in order to assess the performance and ease of use of each condition, as well as preferences. Questions 1 to 4 were generic to the study, while questions 5 to 11 were repeated for each condition (Table 1).

4.4 Procedure

All participants used the three experimental conditions, but the order was varied among participants (and registered as a secondary variable). At the beginning of the experiment, participants were instructed about the experimental setup, the tasks and gave their informed consent. An adaptation

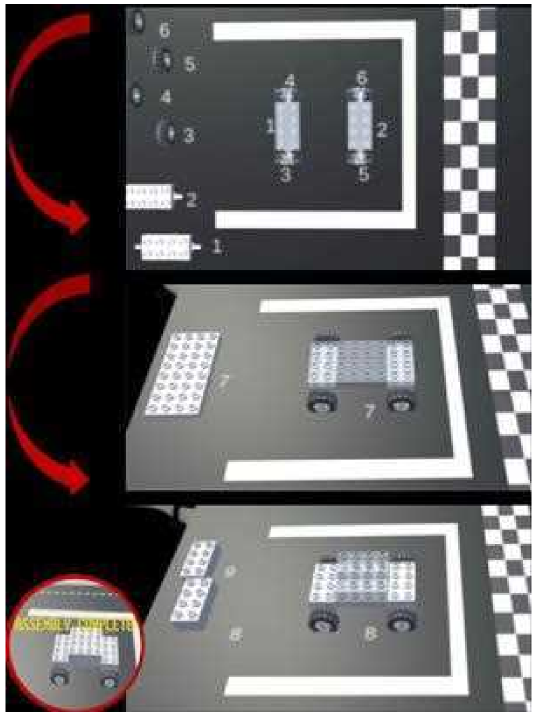

Fig. 7. Brick block - assembly steps, with pieces identified by numbers in the augmented interface .

period was provided to experiment moving some Lego pieces to better understand the different interactions available. Then, they were asked to consider two levels of priorities: perform the task as accurately and as fast as possible. in the next stage, participants completed the assembly task while observed by an experimenter who assisted them if they asked for help. The experimenter used a standard form to make annotations (e.g. main difficulties, etc.). Immediately after completing the task using the three conditions, participants answered the post-study questionnaire.

4.5 Participants

Twenty seven participants (5 female) aged from 18 to 46 years old, performed the assembly task and completed the post-study questionnaire. Participants had various professions within the academic community (e.g. Master and PhD students, Researchers and Teachers). 17 participants had previous experience using AR. From these, 7 had never used AR for assembly guidance purposes before.

5 RESULTS AND DISCUSSION

This section presents and discusses the main results from the performance measures and opinions using Exploratory Data Analysis (EDA) [24], ANOVA [43] and non-parametric tests [19].

5.1 Overall completion time in each condition

Figure 8 shows the boxplots of the time required to complete the car assembly in each experimental condition. Participants were faster when they used condition C3-HMD. The average times for the three methods are 219.5s, 365.1s and 72.9s, respectively. As the preconditions for ANOVA were not met, the alternative non-parametric Friedman test was used. The median times for

Table 1. Post-task questionnaire for evaluating the performance and ease of use of the three conditions. Questions 1 - 4 were generic to the study. Questions 5 - 11 were repeated for each condition. Questions 1 and 5 to 9 used a standard Likert-type Scale with 7 levels: 1- Totally disagree; 7- Totally agree.

#	Question	Response Type
Q1	Were the assembly instructions easy to understand?	Likert-type Scale with 7 levels
Q2	Order the methods according to the preference.	Open Answer
Q3	After the experiment I had.	dizziness, sickness, disorientation, none
Q4	Add additional comments you may consider relevant.	Open Answer
Q5	The manipulation and interaction are intuitive.	Likert-type Scale with 7 levels
Q6	It is easy to place the object in the correct position.	Likert-type Scale with 7 levels
Q7	Manipulation and interaction have some irritating characteristics.	Likert-type Scale with 7 levels
Q8	Manipulation and interaction could improve with training.	Likert-type Scale with 7 levels
Q9	Please select your degree of satisfaction towards this condition.	Likert-type Scale with 7 levels
Q10	What were the main difficulties?	Open Answer
Q11	Please add additional comments regarding this specific condition, you may consider relevant.	Open Answer

the three methods were significantly different (p-value=0.00) and multiple pairwise comparisons (considering the Bonferroni correction) showed significant differences among all the methods (HMD-Touch, p-value=0.001; HMD-Movement, p-value=0.000; Touch-Movement, p-value=0.013). When categorizing the time of completion by the secondary variables participants' gender and previous experience with VR, AR and assembly, no significant differences were found. However, when considering the time of completion regarding the secondary variable order of using each experimental condition, the equality of means was rejected by an one-way ANOVA (p-value=0.037) for C3-HMD. A pairwise comparison using the post-hoc test Least Significant Distance of Fisher showed that with this condition participants took significantly longer when they used it in the first place than when they used in second or third place. For the other conditions (C1-Touch and C2-Movement) this effect was not significant.

5.2 Completion time by piece in each condition

Figure 9 presents the bar chart corresponding to the participants' average completion time by piece, while using each experimental condition. . Participants were faster when they used condition C3-HMD. For condition C2-Movement the first piece took longer to place in the correct pose than the other pieces, which might be due to a less intuitive interaction, as mentioned later by some participants. It is also noticeable a time decrease along the sequence of pieces for this condition, suggesting that a training period is necessary to master this interaction method.

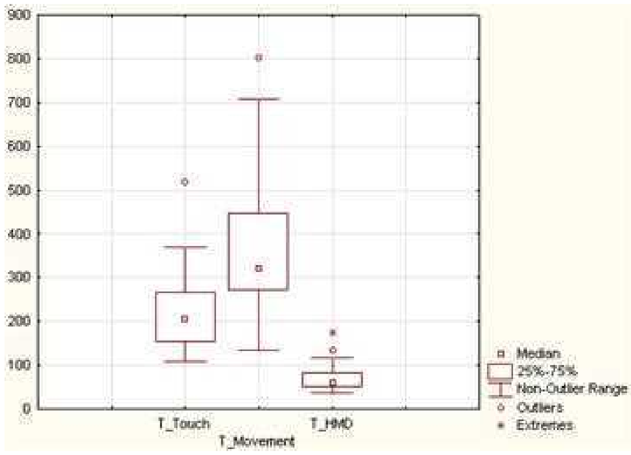

Fig. 8. Boxplots of completion times by experimental conditions C1-Touch, C2-Movement and C3-HMD.

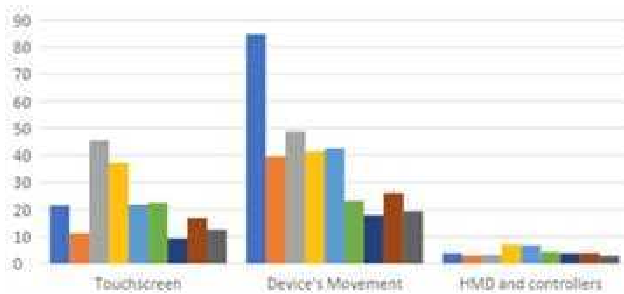

Fig. 9. Average time by experimental condition for each piece: C1-Touch; C2-Movement; C3-HMD. A different color is assign to every Lego Piece.

5.3 Post-task Questionnaire

Answers to the post-task questionnaire show that the majority of the participants found the assembly instructions easy to understand (Q1: median=7 - totally agree). Analysing the answers to question Q2, the preference order of conditions seems related to the completion time obtained with each condition. For instance, twenty six participants indicated C3-HMD as their preferred condition and all were faster with this condition. Moreover, the participants who considered C1-Touch as their second preference achieved their second best time with this condition, and the participants who considered C2-Movement as the least preferred achieved their worst time with this condition.

Although participants preferred condition C3-HMD it must be noticed (Q3) that 5 out of 27 participants reported symptoms of nausea, sickness and/or disorientation, after using this condition, which may cause significant impact in the participants performance for long periods of usage. While the device used is a robust VR HMD, the camera resolution remains very limited, displaying a blurred representation of the real world when in video see-through AR mode, which in turns impacts the experience. In addition, previous work also suggest that video see-through systems often suffer from increased latency, which may also contribute to the cybersickness symptoms [14, 17].

Another limitation, is the fact that it is limited to a specific physical space, due to set up constrains as it requires two tracking cameras and a computer with solid characteristics, which also translate into an expensive investment. As such, conditions C1 and C2 may prove a good compromise between an efficient and accurate interaction mechanism, portability and budget.

As questions Q5 to Q9 were answered using an ordinal scale and each user performed the three conditions (matched sample) the equality of medians was tested with the Friedman test (non parametric ANOVA). In all cases there were significant differences between condition C3-HMD and each of the other conditions (C1-Touch and C2-Movement); however, no differences were found between these latter conditions, suggesting they were considered as generally similar by the participants. Details concerning the analysis of the answers to each question are presented next.

Regarding Q5, all interaction methods were classified as intuitive by all participants. Yet, condition C3-HMD stands out as the most intuitive (median = 7), as expected, and condition C2-Touch as the less intuitive (median = 5). The Friedman test rejected the null hypothesis -equality of medians (p-value=0.000), indicating differences among methods; in pairwise comparisons significant differences were found between C2-Movement and C3-HMD (p-value=0.000), as well as between C1-Touch and C3-HMD (p-value=0.006).

Condition C3-HMD also presents the better results (median = 7) regarding the ease to place virtual pieces in the correct pose (Q6). The Friedman test rejected the null hypothesis - equality of medians (p-value=0.000), indicating differences among methods; in pairwise comparisons significant differences were found between C2-Movement and C3-HMD (p-value=0.000), as well as between C1-Touch and C3-HMD (p-value=0.000). Despite, all participants believe the accuracy associated with all interaction methods could improve with training (Q8) and no significant differences among methods were found.

Regarding Q7, the participants considered that in general the methods do not have irritating characteristics; however condition C2-Movement was the least favorite and gathered some specific comments. For instance, one participant stated that *"Although the interaction was easy to understand, it was hard to execute. The metaphor used seemed easier than condition 1, but after testing it, it was not as easy as expected"*. Another participant shared this idea, reporting that *"This method was intuitive enough in the sense that it came closer to direct interaction. Nevertheless, it took me some time to adjust to do the fine movements required to mount the model"*.

Moreover, participants highlighted this condition was difficult to master and should provide some additional mechanisms in the user interface in order to display the distance between the virtual piece and its desired pose, or even changing the piece color between a specific set of colors, thus easing the placement process. Some users also complained about the accuracy required to place the pieces in the desired pose and the lack of feedback. These complaints happened when the pieces were misaligned. For example, users reported that the perspective effect led them to think that the pieces were sometimes above or below the desired pose and they did not know what to do. As in previous cases, the Friedman test rejected the null hypothesis -equality of medians (p-value=0.000), indicating differences among methods; in pairwise comparisons significant differences were found once more between C2-Movement and C3-HMD (p-value=0.000), as well as between C1-Touch and C3-HMD (p-value=0.033).

Regarding Q9, degree of satisfaction, the three conditions, C1-Touch, C2-Movement and C3-HMD, obtained the medians of 5, 4 and 7, respectively. The Friedman test rejected the null hypothesis - equality of medians (p-value=0.000), indicating differences among methods; in pairwise comparisons significant differences were found once more between C2-Movement and C3-HMD (p-value=0.000), as well as between C1-Touch and C3-HMD (p-value=0.000).

Concerning future improvements, participants identified the following AR enhancements: presentation of hints (lines, circular arrows and/or others) to improve the guiding process required to manipulate the pieces into the correct poses, thus allowing to recover from localization and/or orientation errors more easily. When ask about other interaction methods which could be a good alternative, participants suggested: haptic gloves, hand recognition and a mix between the characteristics of conditions 1 and 2.

Participants also recognized that the AR methods could be more beneficial when applied to more complex tasks. Hence, a new study including more complex tasks and a larger group of participants must be conducted to obtain more significant results, providing more insight concerning the proposed methods.

6 CONCLUSIONS AND FUTURE WORK

Several industry applications are starting to use Augmented Reality (AR) for assisting in assembly tasks due to its great potential. This paper presented a prototype using three different AR-based interaction methods for assembly procedures (Touch gestures using a mobile device; Mobile Device movement; Controllers and See-through HMD). A controlled user study was performed with 27 participants to compare the performance, ease of use and acceptance of the methods using a case study based on brick blocks. Participants were faster using the AR condition based on the use of controllers and a See-through HMD. It also appears that the mobile device movement condition seems to require training to be correctly used as the time to place the first piece was significantly higher with this condition. Although participants also preferred HMD-controllers condition, the device used displays a blurred representation of the real world, which some participants report to have caused symptoms of nausea, sickness and/or disorientation. Therefore, for long periods of usage, this condition can cause an impact in the users experience.

The study presents some limitations, namely because the See-through HMD condition is significantly different from the handheld conditions This had a significant impact on the results as the direct manipulation is a more effective interaction but also causes some discomfort to users. Further studies with all the interactions implemented in a HMD set-up (inclusive handheld ones) might be a possibility to minimize differences between the set-ups and realize a better comparison. Also, the task used (Lego car assembly) is fairly simple. The use of a more complex and realistic task might show more clearly the benefits and limitations of the different methods.

Work will continue through the integration of these methods into a collaborative AR platform for co-located cooperation. Collaborative scenarios provide an interesting challenge and literature shows few works exist that aim to explore interaction methods in such scenarios. Therefore, we aim to conduct a large scale collaborative user-testing to obtain more insight concerning: which method stands out in a collaborative context and why, user awareness, interest and social interaction.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their thoughtful comments and efforts towards improving our manuscript. Likewise, to everyone involved in case studies, and discussion groups, thanks for collaborating in providing essential data and relevant feedback. The present study was developed in the scope of Produtech – SIF – Solutions for the Industry of the Future [POCI-01- 0247-FEDER-024541]. It was also supported by IEETA - Institute of Electronics and Informatics Engineering of Aveiro, funded by National Funds through FCT, in the context of the project [UID/CEC/00127/2019].

REFERENCES

[1] João Alves, Bernardo Marques, Miguel Neves, Rafael Maio, André Santos, Inês Justo, Raquel Rainho, Dany Costa, Beatriz Sousa Santos, and Paulo Dias. 2019. Augmented Reality Platform to foster Collaborative Manipulation of 3D virtual objects: A Prototype. In *International Conference on Graphics and Interaction, ICGI 2019*.

[2] J. Alves, B. Marques, M. Oliveira, T. Araújo, P. Dias, and B. S. Santos. 2019. Comparing Spatial and Mobile Augmented Reality for Guiding Assembling Procedures with Task Validation. In *2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. 1–6.

[3] Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (1997).

[4] R. T. Azuma. 2016. The Most Important Challenge Facing Augmented Reality. *Presence* 25, 3 (Dec 2016), 234–238.

[5] Mafkereseb Kassahun Bekele, Roberto Pierdicca, Emanuele Frontoni, Eva Savina Malinverni, and James Gain. 2018. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *J. Comput. Cult. Herit.* 11, 2 (March 2018), 7:1–7:36.

[6] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends in Human–Computer Interaction* 8, 2-3 (2015), 73–272.

[7] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative Augmented reality. In *Communications of the ACM*.

[8] Mark Billinghurst, Hiroyuki Kato, and Ivan Poupyrev. 2008. Tangible augmented reality. In *SIGGRAPH 2008*.

[9] Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy Cockburn. 2004. FingARtips: Gesture Based Direct Manipulation in Augmented Reality. In *Proceedings of the 2Nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia* (Singapore) *(GRAPHITE '04)*. ACM, New York, NY, USA, 212–221.

[10] Sebastian Büttner, Markus Funk, Oliver Sand, and Carsten Röcker. 2016. Using Head-Mounted Displays and In-Situ Projection for Assistive Systems: A Comparison. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Island, Greece) *(PETRA '16)*. ACM, New York, NY, USA, Article 44, 8 pages.

[11] Fabio M. Caputo, Marco Emporio, and Andrea Giachetti. 2018. The Smart Pin: An effective tool for object manipulation in immersive virtual reality environments. *Computers Graphics* 74 (2018), 225 – 233.

[12] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. 2017. Mobile Augmented Reality Survey: From Where We Are to Where We Go. *IEEE Access* 5 (2017), 6917–6950.

[13] Arindam Dey, Mark Billinghurst, Robert W Lindeman, and J Edward Swan Ii. 2018. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI* 5 (2018).

[14] Paul Dizio and James Lackner. 2000. Motion Sickness Side Effects and Aftereffects of Immersive Virtual Environments Created with Helmet-Mounted Visual Displays. (11 2000), 4.

[15] Andreas Dünser and Mark Billinghurst. 2011. Evaluating Augmented Reality Systems. In *Handbook of Augmented Reality*. 289–307.

[16] Iñigo Fernández Del Amo, John Ahmet Erkoyuncu, Rajkumar Roy, Riccardo Palmarini, and Demetrius Onoufriou. 2018. A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications. *Computers in Industry* 103 (2018), 47–71.

[17] Lawrence H. Frank, John G. Casali, and Walter W. Wierwille. 1988. Effects of Visual Display and Motion System Delays on Operator Performance and Uneasiness in a Driving Simulator. *Human Factors* 30, 2 (1988), 201–217.

[18] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with Augmented Reality? A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)*. Association for Computing Machinery, New York, NY, USA, 222–229.

[19] Jean Dickinson Gibbons and Subhabrata Chakraborti. 2011. *Nonparametric Statistical Inference*. Springer Berlin Heidelberg, Berlin, Heidelberg, 977–979.

[20] E. S. Goh, M. S. Sunar, and A. W. Ismail. 2019. 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review. *IEEE Access* 7 (2019), 40581–40601.

[21] Jens Grubert and Stefanie Zollmann. 2017. Towards Pervasive Augmented Reality: Context- Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (2017), 1706–1724.

[22] Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. Plane, Ray, and Point: Enabling Precise Spatial Manipulations with Shape Constraints. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1185–1195.

[23] Anders Henrysson, Mark Billinghurst, and Mark Ollila. 2005. Face to face collaborative AR on mobile phones. *IEEE a International Symposium on Symposium on Mixed and Augmented Reality, ISMAR 2005* (2005), 80–89.

[24] David C. Hoaglin, Frederick Mosteller, and John W. Tukey (Editor). 2000. *Understanding Robust and Exploratory Data Analysis* (1 ed.). Wiley-Interscience.

[25] Jérme Jetter, Jörgen Eimecke, and Alexandra Rese. 2018. Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits? *Computers in Human Behavior* 87 (2018), 18–33.

[26] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura. 2014. The effectiveness of an AR-based context-aware assembly support system in object assembly. In *2014 IEEE Virtual Reality (VR)*. 57–62.

[27] K. Kim, M. Billinghurst, G. Bruder, H. B. Duh, and G. F. Welch. 2018. Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008-2017). *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2947–2962.

[28] D W F Van Krevelen and R Poelman. 2010. A Survey of Augmented Reality Technologies, Applications and Limitations. *The International Journal of Virtual Reality* 9 (2010).

[29] M. Krichenbauer, G. Yamamoto, T. Taketom, C. Sandor, and H. Kato. 2018. Augmented Reality versus Virtual Reality for 3D Object Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 24, 2 (Feb 2018), 1038–1048.

[30] Ze-Hao Lai, Wenjin Tao, Ming C. Leu, and Zhaozheng Yin. 2020. Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. *Journal of Manufacturing Systems* 55 (2020), 69 – 81.

[31] Joseph J. LaViola Jr., Ernst Kruijff, Doug Bowman, Ivan P. Poupyrev, and Ryan P. McMahan. 2017. *3D User Interfaces: Theory and Practice (second edition).*

[32] Wenkai Li, A. Nee, and S. Ong. 2017. A State-of-the-Art Review of Augmented Reality in Engineering Analysis and Simulation. *Multimodal Technologies and Interaction* 1, 3 (2017).

[33] Jingbo Liu, Oscar Kin-Chung Au, Hongbo Fu, and Chiew-Lan Tai. 2012. Two-Finger Gestures for 6DOF Manipulation of 3D Objects. *Comput. Graph. Forum* 31, 7pt1 (Sept. 2012), 2047–2055.

[34] F. Loch, F. Quint, and I. Brishtel. 2016. Comparing Video and Augmented Reality Assistance in Manual Assembly. In *2016 12th International Conference on Intelligent Environments (IE).* 147–150.

[35] Annette Mossel, Benjamin Venditti, and Hannes Kaufmann. 2013. 3DTouch and HOMER-S: Intuitive Manipulation Techniques for One-Handed Handheld Augmented Reality. In *Proceedings of the Virtual Reality International Conference: Laval Virtual.* Association for Computing Machinery, New York, NY, USA, 10.

[36] K. Murakami, R. Kiyama, T. Narumi, T. Tanikawa, and M. Hirose. 2013. Poster: A wearable augmented reality system with haptic feedback and its performance in virtual assembly tasks. In *2013 IEEE Symposium on 3D User Interfaces (3DUI).* 161–162.

[37] A. Nishihara and J. Okamoto. 2015. Object recognition in assembly assisted by augmented reality system. In *2015 SAI Intelligent Systems Conference (IntelliSys).* 400–407.

[38] Michael E. Porter and James E. Heppelmann. 2017. *A Manager's Guide to Augmented Reality. Harvard Business Review.* Technical Report.

[39] Moritz Quandt, Benjamin Knoke, Christian Gorldt, Michael Freitag, and Klaus Dieter Thoben. 2018. General Requirements for Industrial Augmented Reality Applications. *Procedia CIRP* 72 (2018), 1130–1135.

[40] Rafael Radkowski and Christian Stritzke. 2012. Interactive Hand Gesture-based Assembly for Augmented Reality Applications. , 5 pages.

[41] Christian Sandor, Martin Fuchs, Alvaro Cassinelli, Hao Li, Richard Newcombe, Goshiro Yamamoto, and Steven Feiner. 2015. Breaking the Barriers to True Augmented Reality. *ArXiv e-prints* 1512.05471 (2015), 1–13.

[42] C.Y. Siew, S.K. Ong, and A.Y.C. Nee. 2019. A practical augmented reality-assisted maintenance system framework for adaptive user support. *Robotics and Computer-Integrated Manufacturing* 59 (2019), 115 – 129.

[43] B.G. Tabachnick and L.S. Fidell. 2007. *Experimental Designs Using ANOVA.* Thomson/Brooks/Cole.

[44] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03).* ACM, New York, NY, USA, 73–80.

[45] Matthew Turk. 2014. Multimodal interaction: A review. *Pattern Recognition Letters* 36 (2014), 189 – 195.

[46] Peng Wang, Shusheng Zhang, Mark Billinghurst, Xiaoliang Bai, Weiping He, Shuxia Wang, Mengmeng Sun, and Xu Zhang. 2019. A comprehensive survey of AR/MR-based co-design in manufacturing. *Engineering with Computers* (2019), 1–24.

[47] Xiangyu Wang, Soh K Ong, and Andrew YC Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (2016), 1–22.

[48] Sabine Webel, Uli Bockholt, Timo Engelke, Nirit Gavish, Manuel Olbrich, and Carsten Preusche. 2013. An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems* 61, 4 (2013), 398 – 403. Models and Technologies for Multi-modal Skill Training.

[49] M. L. Yuan, S. K. Ong, and A. Y. C. Nee. 2008. Augmented reality for assembly guidance using a virtual interactive tool. *International Journal of Production Research* 46, 7 (2008), 1745–1767.

[50] Feng Zhou, Henry Duh, and Mark Billinghurst. 2008. Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR. *IEEE International Symposium on Symposium on Mixed and Augmented Reality, ISMAR 2008* (2008), 193–202.

Received June 2020; revised August 2020; accepted September 2020