Move'n'Hold: Scalable Device-Based Interaction for Mixed Reality Handheld Displays

Vera M. Memmesheimer
Human Computer Interaction Lab,
RPTU Kaiserslautern-Landau
Germany
v.memmesheimer@rptu.de

Kai J. Klingshirn
Human Computer Interaction Lab,
RPTU Kaiserslautern-Landau
Germany
klingshi@rptu.de

Bahram Ravani
Department of Mechanical and Aerospace Engineering,
University of California, Davis
USA
bravani@ucdavis.edu

Achim Ebert
Human Computer Interaction Lab,
RPTU Kaiserslautern-Landau
Germany
achim.ebert@rptu.de

ABSTRACT

While handheld displays provide a widely available, low-budget access to Mixed Reality, developing appropriate interaction techniques remains challenging. For example, one-handed touch- and gestures-based methods are prone to fatigue and occlusion issues and two-handed device-based techniques are constrained to small-range manipulations or their pre-defined thresholds limit user control. In this paper, we introduce *Move'n'Hold* – a universally applicable interaction paradigm for translating and rotating virtual objects solely through a handheld display's movement and peripheral touch. *Move'n'Hold* combines direct mapping between device and object manipulations when only left-thumb-touch is applied with automated repetitions of these initial manipulations that are started or stopped when right-thumb-touch is added or released. Hence, our technique allows switching individually between natural manipulation for small, precise movements and continuous manipulation for large, coarse movements. Our evaluation revealed *Move'n'Hold* as an intuitive and easy-to-learn input technique for multidimensional object translations and rotations. The technique also provides high scalability in terms of the distance, direction, complexity, and speed of manipulation. At the same time, it supports different user preferences and interaction styles. Our results further show that learning translation prior to rotation enhanced the overall user experience.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**.

ECCE '23, September 19–22, 2023, Swansea, United Kingdom

ACM ISBN 979-8-4007-0875-6/23/09…$15.00
https://doi.org/10.1145/3605655.3605656

KEYWORDS

Mixed Reality, Augmented Reality, Handheld Display, Device-based Interaction

ACM Reference Format:
Vera M. Memmesheimer, Kai J. Klingshirn, Bahram Ravani, and Achim Ebert. 2023. Move'n'Hold: Scalable Device-Based Interaction for Mixed Reality Handheld Displays. In *European Conference in Cognitive Ergonomics (ECCE '23), September 19–22, 2023, Swansea, United Kingdom.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3605655.3605656

1 INTRODUCTION

Handheld displays (HHDs) such as smartphones or tablets are widespread, low-budget technologies for Mixed Reality (MR) applications. As HHDs are meant to serve as a window to the MR scene, the device needs to be held up high to see the virtually augmented view of the device camera. This setting poses new challenges for interaction techniques [5]: For example, touch-based and mid-air gestures-based interaction require the HHD to be held with one hand, are prone to fatigue, and can cause scene occlusion. Device-based interaction which maps the HHD's movements to the object reduces these issues but impedes large-range manipulations.

In this paper we present the design, implementation, and evaluation of *Move'n'Hold* – a highly scalable, unified interaction paradigm for object translation and rotation that seeks to address usability issues of existing MR-HHD interaction techniques. It provides simplicity through a minimalist user interface and a manageable set of input modalities (i.e., device movement, left- and right-thumb-touch). In this technique, the selected object is translated or rotated according to the HHD's movement in real time while the user applies touch only with the left thumb. Automated repetitions of the initial movement, without having to move the HHD further, can be started by adding right-thumb-touch. By releasing right-thumb-touch, continuous object manipulation stops and natural object manipulation can be resumed. As such, *Move'n'Hold* supports small and large, fine and coarse, slow and fast multidimensional manipulations. At the same time, the user retains full control over the activation and speed of continuous movement.

To evaluate the scalability of *Move'n'Hold* we conducted a main and a follow-up experimental study in which participants performed object translations and rotations that differ in complexity, directions, and distances. We assessed the effectiveness by the number of successfully completed tasks, learnability by task completion times and subjective feedback, cognitive load by the NASA TLX [1], and intuitiveness by the QUESI [12]. Further questions were asked regarding the overall experience, preferred interaction styles, and *Move'n'Hold*'s usefulness for different task configurations. To investigate the impact of the order in which translation and rotation techniques are learned, we taught translation prior to rotation in the main study and vice versa in the follow-up study and then compared the results.

2 RELATED WORK

While MR was initially introduced as an umbrella term for Augmented Reality (AR; real scenes augmented with virtual contents) and Augmented Virtuality (AV; virtual scenes augmented with real contents) [10], the terms MR and AR are nowadays often used as synonyms. In this paper, we use the term MR to refer to real environments that are visually augmented by virtual contents. Apart from head-mounted displays (HMDs), handheld displays (HHDs) offer a widely available, low-budget access to MR applications. However, the development of appropriate interaction techniques remains challenging. Due to the difficulty of mapping 2D touch-input to 6DOF manipulations, Fuvattanasilp et al. [4] proposed an approach that reduces the number of DOFs that have to be manipulated at once: Their technique in placing an object involves, first defining an initial 2D position using touch and then depth can be set on a ray originating from the defined point while setting the orientation is constrained to rotations around the object's gravity vector. Mossel et al. [11] implemented a touch-based control technique that uses single touch input in combination with the HHD's current pose. In that way 2D translations on the screen are back-projected to translate and rotate the object. Mode switching is enabled through buttons. Translations in the xy-plane were implemented similarly in [6, 9]. Marzo et al. [9] used the distance and angle between two touch points to handle z-axis translations and rotations as well as their middle point to control an arcball rotation. In Grandi et al.'s [6] approach, z-axis translations can be performed via one-touch tapping and sliding, x- and y-axis rotations via two-touch sliding, and z-axis rotations via two-touch rotations. Further approaches implemented mid-air gestures-based manipulation: Kim and Lee [8] proposed a hybrid approach in which touch is reduced to object selection, and object manipulation is performed via hand gestures that are detected by an external sensor attached to the HHD. Botev et al. [3] implemented a gestures-based manipulation technique that provides computer-vision-based hand tracking. While mid-air gestures are usually performed behind the screen, virtual augmentations may still be occluded. Hence, device-based interaction techniques that allow users to hold the device with two hands are deemed a promising alternative [5]. Previous work allowed users to select objects by touching them (e.g., [11]) or centering them in the screen (e.g., [2]). Subsequently, objects can be manipulated as device movements are mapped to the virtual object's pose [2, 6, 9, 11]. Mossel et al. [11] map the HHD's pose to the objects, whereby rotations are performed around the point at which the selection ray hits the object. Blattgerste et al. [2] implemented a grabbing metaphor which automatically places the object in front of a smartphone such that it can be manipulated and dropped by moving the device into the desired position and orientation. Grandi et al. [6] found that touch-based interaction was better suited for rotating objects while device-based interaction was better suited for translating objects. Similar findings were reported by Marzo et al. [9] who compared a multi-touch, device-based, and hybrid approach. While the hybrid mode provided best performance, the device-based mode turned out to be more intuitive and its performance was only worse for large rotations. Samini and Palmerius [13] however state that device-attached objects are hard to rotate without translating them, and objects that are rotated relatively around their own center will move out of view during rotations around the x- and y-axis. Addressing these issues, they proposed an approach that allows rotating an object relative to the device while user perspective rendering is implemented to prevent the object from moving out of view. Since this approach still requires repositioning the device during large-range rotations, Su et al. [14] proposed a rotation technique that continuously rotates objects around a constant speed when the HHD's rotation exceeds pre-defined threshold values. Pre-defined thresholds and speed however limit flexibility and controllability. Furthermore, starting and resuming continuous movement using this approach still requires device repositioning. Moreover, the work of Su et al. [14] implements different interaction paradigms for translation and rotation. Continuous movement is only available for large rotations while large translations still have to be performed by repositioning the device.

Thus, it can be summarized that object manipulation techniques for MR-HHDs should (RQ1) allow holding the HHD with two hands to prevent scene occlusion and fatigue, (RQ2) provide input channels for manipulations in three dimensions, (RQ3) enable large rotations while preventing the object from accidental translations and moving out of view, (RQ4) separate object translation and rotation, and (RQ5) be easy and intuitive to learn. Apart from this, we consider (RQ6) scalability (i.e., a unified interaction paradigm for object translations and rotations of different distances, directions, and complexity) as well as (RQ7) controllability (i.e., users can activate continuous movement, adjust its speed, and switch between small and precise as well as large and coarse movements individually) to be highly relevant. However, an interaction technique meeting all these requirements is still missing in the literature.

3 INTERACTION DESIGN

Seeking to address this research gap, we designed *Move'n'Hold* which extends the previously proposed natural input paradigm of mapping device movement to objects (referred to as *Move*, see e.g., [2, 6, 9, 11]) such that users can individually combine and switch between natural and continuous movement. With *Move'n'Hold* users can hold the HHD with two hands throughout interaction as input is solely based on the HHD's position and orientation in space in combination with peripheral touch.

To manipulate an object, the user has to move the HHD in space such that the object appears in the screen's center. When this is the case, the object is highlighted and becomes manipulable. To

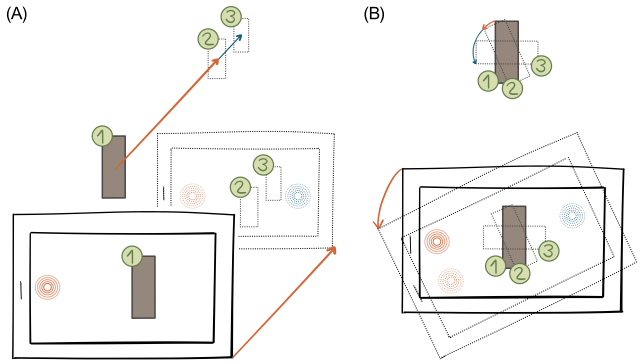

Figure 1: To translate (A) or rotate (B) the virtual rectangle (1), left-thumb-touch is applied first while translating/rotating the HHD. As long as only left-thumb-touch is applied, the HHD's translation/rotation is mapped to the rectangle in real time and the rectangle follows the HHD's movement (2). When right-thumb-touch is added, the rectangle is continuously translated/rotated in the direction of the initial movement without requiring further movement of the HHD. The rectangle's movement stops as right-thumb-touch is released (3).

initiate manipulation, touch has to be applied with the left thumb while the HHD is moved in space. Touch can be applied anywhere on the left side of the display such that the device can still be held with two hands. The HHD can be either translated (Fig. 1A+2) or rotated (Fig. 1B+3). As long as left-thumb-touch is applied, the HHD's movement is mapped to the object relative to the object's center (*Move*). Thereby, the HHD can also be translated or rotated along multiple axes at once. While rotating objects, users also have the option to lock axes such that objects are only rotated around the selected axis. The user can toggle between rotations around all axes, or rotations limited to the x-, y-, or z-axis by clicking a button on the left side (Fig. 4). To evoke continuous movement, the user has to apply some initial movement with *Move* while applying left-thumb-touch first and can then add right-thumb-touch while the left thumb remains on the screen. Again, right-thumb-touch can be applied anywhere on the right side of the display. As long as touch with both thumbs is applied, this initial movement is applied continuously to the selected object. As such, the speed of the continuous movement can be set individually by adjusting the length of the initial movement.

In contrast to the work of Su et al. [14] who proposed continuous movement based on pre-defined thresholds and constant speed only for rotating objects, our method is applicable to both rotation and translation and allows to start continuous movements at any time and speed for step-wise manipulations without repositioning the device. When both left- and right-thumb-touch are applied, the user has the following three options: (1) release touch with both thumbs to stop manipulation and reset the continuous manipulation direction, (2) release right-thumb-touch and maintain left-thumb-touch to continue with manipulation via direct mapping by moving the HHD in space, or (3) maintain left-thumb-touch and repeatedly release and apply right-thumb-touch while the HHD is held still to pause and resume continuous movement. *Move'n'Hold* provides all of these options for both translations and rotations. For example, the second row in Fig. 3 shows how a user finishes a rotation task with approach (1), while the second row in Fig. 2 shows the completion of a translation task with approach (2).

4 IMPLEMENTATION OF *MOVE'N'HOLD*

We developed the MR-HHD application with Apple's ARKit within the AR Foundation API in Unity and deployed it to an Apple iPad Pro (11", 3rd Gen.) via XCode. While we used a tablet instead of a smartphone due to its larger screen, *Move'n'Hold* can be easily applied to smartphones, too. In each frame, a ray is shot through the camera to check if objects are centered in the HHD's screen. If the ray hits an object's box collider, the object is highlighted, it is saved as the current target object, and can be manipulated. Ray shooting is paused while manipulation is in progress (i.e., left-thumb-touch is registered) and continues when no touch is registered.

If an object is selected, translation is active, and only left-thumb-touch is registered we update this object's position as follows: In each frame, we compute the vector that describes the HHD's translation from the previous to the current frame and add it to the object's position. To this end, we register and save the HHD's position in each frame, whereby the origin of the coordinate system corresponds to the position of the HHD's camera when the app is launched. If touch is registered on both display sides, we compute the vector describing the HHD's translation while only left-thumb-touch was applied. This vector is saved as the continuous translation vector along which the selected object will be translated as long as left- and right-thumb-touch are registered. To prevent very fast movements and ensure that the user remains in control of object manipulation, we perform a linear interpolation by 0.1 as the interpolant, which was determined to provide good control in a pre-study. In each frame, we linearly interpolate between the object's current position and the object's current position to which the continuous translation vector was added and update the object's position to the interpolated value. The continuous translation vector is stored until both left- and right-thumb-touch are released.

Object rotation is based on the same interaction paradigm like object translation. While translation relies on the HHD's position in space, rotation relies on the HHD's orientation in space. If an object is selected and left-thumb-touch is applied, we update the object's orientation in each frame by multiplying the unit quaternion that describes the rotation of the HHD between the current and the previous frame. Again, the HHD's orientation is registered and saved in each frame. As such, we can compute the unit quaternion that describes the HHD's rotation during the initial natural movement while only left-thumb-touch was applied and store it as the unit quaternion for the continuous rotation. As soon as left- and right-thumb-touch is registered, the object will be continuously rotated. To this end, we update the object's orientation in space in each frame with the continuous rotation unit quaternion. Similar to object translations, we interpolate by a 0.1 interpolant between the object's current orientation and the current orientation multiplied with the continuous rotation unit quaternion. We then set the object's orientation to the interpolated value. Since we found

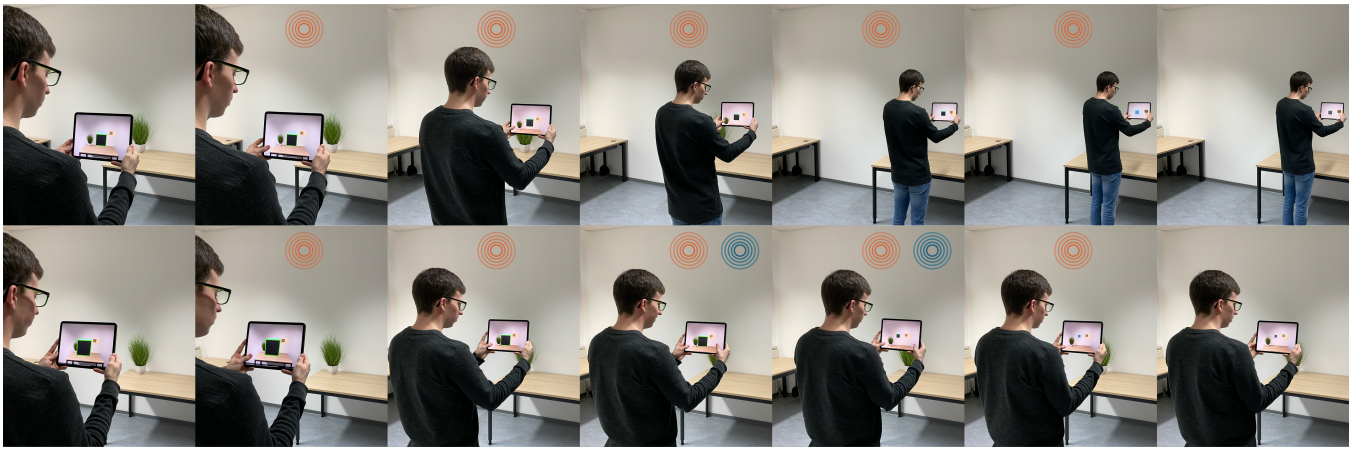

Figure 2: The two photo series show the same task completed with *Move* (first row) and *Move'n'Hold* (second row): A virtual box is translated along the z-axis. Comparing the photo series shows that object translation with *Move'n'Hold* requires less movement than with *Move*. First row (*Move*): The HHD is moved while applying left-thumb-touch (i.e., the HHD's translation is directly mapped to the object). Left-thumb-touch is released after task completion. Second row (*Move'n'Hold*): The HHD is first moved while applying left-thumb-touch , then right-thumb-touch is added to start continuous movement (i.e., the object keeps moving in the same direction while the HHD is held still). Next, right-thumb-touch is released to stop continuous translation and resume translation via direct mapping. Left-thumb-touch is released after task completion.

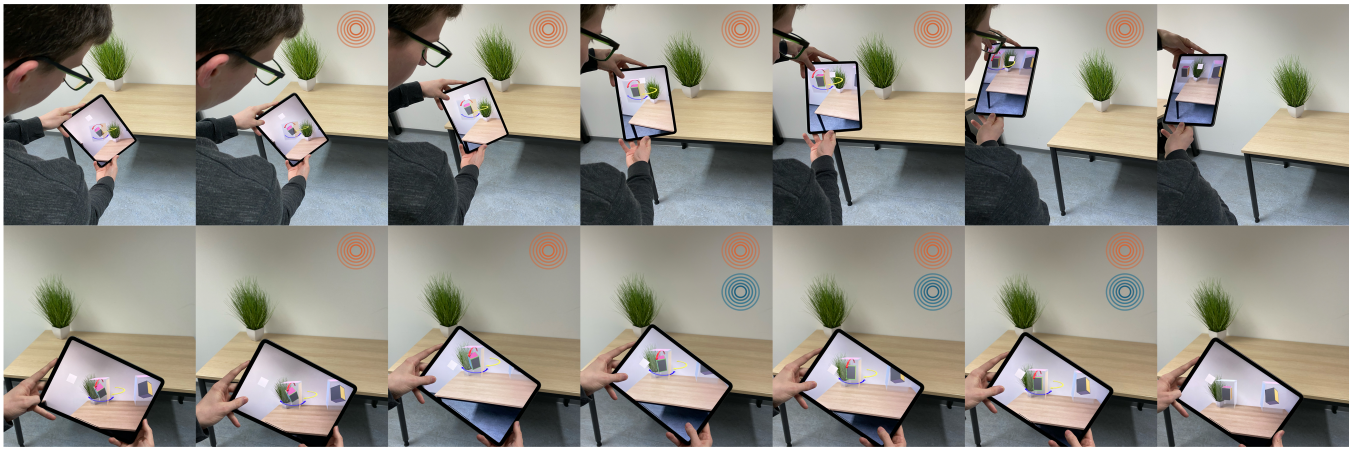

Figure 3: The two photo series show the same task completed with *Move* (first row) and *Move'n'Hold* (second row): A virtual box is rotated around the x- and the y-axis. Comparing the photo series shows that object rotation with *Move'n'Hold* requires less movement than with *Move*. First row (*Move*): The HHD is moved while applying left-thumb-touch (i.e., the HHD's rotation is directly mapped to the object). Left-thumb-touch is released after task completion. Second row (*Move'n'Hold*): HHD is moved while applying left-thumb-touch , then right-thumb-touch is added to start continuous movement (i.e., the object keeps rotating in the same direction while the HHD is held still). Both left- and right-thumb-touch are released as the virtual box reaches its target orientation and the task is completed.

that rotating the HHD only around a single axis is very difficult for humans in contrast to object translations, we implemented an additional feature that allows locking axes during rotation. By clicking a peripheral button the user can toggle between rotation around all axes, the x-axis, the y-axis, or the z-axis. If any axis is selected, rotations around the other axes are ignored and the visibility of the auxiliary arrows is adapted accordingly (Fig. 4).

5 USER STUDY

5.1 Experimental Design

Revisiting the requirements for object manipulation techniques mentioned at the end of section 2, it can be summarized that the design and implementation of *Move'n'Hold* allows users to hold the HHD with two hands (RQ1), supports 3D manipulation (RQ2) as

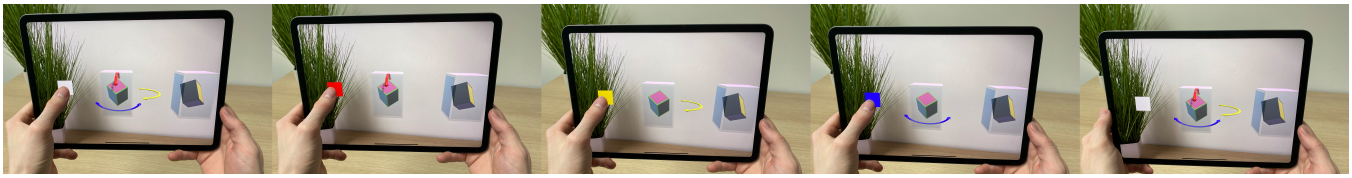

Figure 4: For object rotation, users can lock axes such that only the HHD's rotation around the selected axis is mapped to the object and rotations around all other axes are ignored. Users can press the button on the left to toggle between all axes (white), x- (red), y- (yellow), and z-axis (blue). Depending on the selected option, the button's color and the auxiliary arrows are adapted accordingly.

well as large rotations (RQ3), and separates translation and rotation (RQ4). To investigate how easy and intuitive it is for users to learn and apply *Move'n'Hold* (RQ5) as well as to evaluate its scalability (RQ6) and controllability (RQ7) we conducted a main and a follow-up experimental study with a total of 31 participants (16 male, 15 female; 21 – 36 years old). While all participants had experience with HHDs such as smartphones or tablets only 7 had prior experience with MR and only 6 had experience with MR-HHDs. Different task configurations were posed to evaluate *Move'n'Hold's* effectiveness and flexibility for different levels of complexity, directions, and distances. In total, each participant completed 50 translation tasks and 50 rotation tasks. We assessed intuitiveness with the QUESI [12] and cognitive load with the NASA TLX [1]. Learnability was assessed based on task completion times and subjective feedback. The participants answered additional questions regarding the usefulness of *Move'n'Hold*, user experience, and preferred interaction styles. Since participants of a pre-study reported that they preferred learning the translation before the rotation technique, we let participants in the main study (n = 20) perform the tasks in this order. To investigate how this order affects learnability and user experience, we conducted a follow-up study with a new set of participants (n = 11) that performed rotation tasks prior to translation tasks.

5.2 Task Configurations and Procedure

For the translation tasks, the MR scene included a blue and an orange pair of boxes: An opaque manipulable box (15cm x 15cm x 15cm) and a transparent target box (25cm x 25 cm x 25 cm). To solve the task, the manipulable boxes had to be moved into the corresponding target boxes. As soon as this was the case, the next task block was started automatically (i.e., two new manipulable boxes appeared at new positions). Similarly, the MR scene for the rotation tasks included two pairs of manipulable boxes (10cm x 15cm x 10cm) and target boxes (20cm x 30cm x 20cm). The manipulable boxes with differently colored sides were placed inside their transparent target boxes and had a different starting orientation. To solve the tasks, the manipulable boxes had to be rotated such that they aligned with their target box. After both pairs of boxes were solved, two new manipulable boxes with different orientations appeared automatically.

The experiments were based on four task configuration parts (A, B, C, D) which differed in terms of distance, direction, and complexity (Table 1 and Fig. 5). Within each part, the position and orientation of the manipulable boxes were updated for each task block. The starting position and orientation of the manipulable boxes were computed relative to the target boxes' position and orientation by subtracting (left box) or adding (right box) 30cm (short), 50cm (medium), or 70cm (long) for translation or 20 deg (short), 40 deg (medium), 60 deg (long) for rotation according to the specified complexities and directions. For instance, in Part A and D, the position and orientation was only adapted in one dimension (i.e., manipulations along and around the x- (right–left), y- (up–down), or z-axis (forward–back)). In Part B and C the manipulable boxes were translated and rotated in two and three dimensions respectively.

First, participants were introduced to the basic manipulation technique and completed Part A with *Move* after a short training session. Next, *Move'n'Hold* was introduced and again some training time was provided before Part A, B, and C were completed with *Move'n'Hold*. This was followed by a repetition of Part A with *Move* in order to explore how the learnability of the device-based object manipulation is affected by the order in which translation and rotation techniques are taught. For Parts A-C, participants were allowed to walk around while completing a set of tasks but were required to come back to the starting position when a new set of tasks was displayed. To investigate how different directions affect basic manipulations, the participants were asked to stay in the starting position while completing Part D with *Move* in the end. The participants could skip tasks if they found them too difficult. This procedure was followed for translation and rotation tasks. After task completion, subjective feedback was collected via the NASA TLX [1] and the QUESI [12]. Furthermore, participants were asked about their preferred interaction styles and answered questions (see caption of Fig. 6) regarding the usefulness of *Move'n'Hold* and user experience.

6 RESULTS

6.1 Main Study Conducted

6.1.1 Effectiveness and Learnability. All tasks were completed successfully and no participant took up the offer to skip a task because it was perceived too difficult. Hence, we rate *Move'n'Hold* to be an effective tool for performing multidimensional small- and large-range translations and rotations. Furthermore, the participants perceived the interaction technique as easy to learn (S4) and thought it will be easy to relearn after a lengthy interruption (S5) (Fig. 6A). A comparison of the completion times for each task block in Part D revealed that basic manipulations can be performed equally fast in all directions: A repeated measures anova showed no significant differences between the task completion times for translations ($p>0.3$)

Table 1: Task configurations.

part	#task blocks	complexity	direction	distance
A	6	1D	x-, y-, or z-axis	short (task blocks 1-3), long (task blocks 4-6)
B	3	2D	xy-, xz-, or yz-plane	medium
C	1	3D	x-, y-, and z-axis	medium
D	3	1D	x-, y-, or z-axis	short

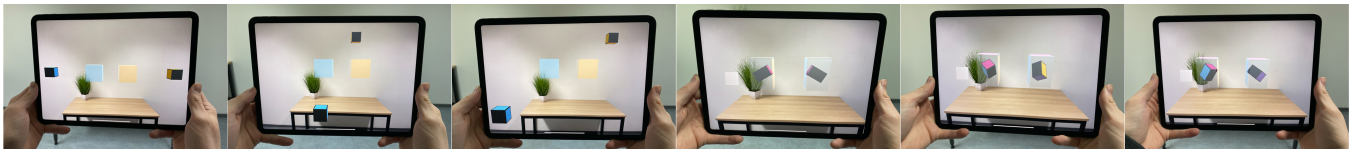

Figure 5: Example of 1D, 2D, and 3D translation (left) and rotation (right) tasks.

Table 2: QUESI [1 (worst) - 5 (best)] and NASA TLX [0 (best) - 100 (worst)] results.

	QUESI [12]		**NASA TLX** [1]	
	translation	rotation	translation	rotation
main	4.3	4.5	31.6	32.2
follow-up	4.4	3.8	29.1	41.3

along or rotations (p>0.1) around the x-, y-, and z-axis. Furthermore, we found that the mean completion time from the first to the last repetition of Part A significantly decreased by 42% for translation ($p \leq 0.0001$) and 43% for rotation ($0.001 < p \leq 0.01$) in paired samples t-tests with Bonferroni correction. On top of that, the standard deviation among the participants' task completion times decreased by 25% for translation and by 64% for rotation tasks, showing that performance discrepancies between users are reduced quickly.

6.1.2 User Experience. To investigate the intuitiveness and the workload of *Move'n'Hold*, we computed QUESI scores [12] and the NASA TLX [1] (Table 2). The QUESI scores [12] indicate that object translation (4.3) and rotation (4.5) were perceived as highly intuitive by the main study participants. Furthermore, the computed NASA TLX [1] for translation and rotation in the main study is lower than in at least 80% of the studies reviewed by Grier [7].

The participants were also asked to rate their agreement with additional statements regarding their experience (Fig. 6A). When manipulating objects the participants had the feeling that their behavior is predictable (S3). They rated the interaction techniques to be suitable for the tasks (S1) and they considered the number of steps to complete the task to be adequate (S2). Overall, the participants were satisfied with the interaction techniques (S6) and would like to use them if they had to translate or rotate virtual objects in the future (S7). For both translation and rotation, the participants rated *Move'n'Hold* to be most useful for long-distance tasks. As shown in Fig. 6C, *Move'n'Hold* was rated slightly more useful for 1D tasks than for 2D and 3D tasks and most useful for x-axis translations, x- and y-axis rotations.

6.1.3 User Preferences and Interaction Styles. Throughout the study, we observed different interaction approaches. While some participants walked around a lot, others mostly stayed at the starting position. This observation is also mirrored in participants' answers regarding their walking preference. When translating objects, 60% of the participants preferred walking, 30% preferred to stay in one place, and 10% did not have a preference. When rotating objects, 50% of the participants preferred walking, 45% preferred to stay in one place, and 5% did not have a preference. Similarly, some users completed the tasks step by step (i.e., translating along and rotating around single axes) while others performed diagonal translations and rotated objects around multiple axes at once. A particularly smart approach for object rotation that we observed was to align the HHD with the manipulable object's front side and then rotate the HHD so that it is aligned with the front side of the target object. In that way, the manipulable object is automatically rotated towards the correct target orientation without having to think a lot about the axes around which the object needs to be rotated. 85% of the participants preferred the option to not lock axes and rotate objects around multiple axes at once. Furthermore, the majority of participants stated that if an axis locking feature was available for translating objects, they would not (50%) use it or use it less often (20%) than for rotating objects. During the rotation of objects, we observed that some users performed rotations around the y-axis by walking around the object. Moreover, we observed that *Move'n'Hold* was often used to bring distant objects closer to the HHD and to correct unintended actions while completing translation tasks.

6.2 Follow-up Study Conducted

To investigate if learning translation prior to rotation is indeed easier, we conducted a follow-up study in which participants performed the rotation prior to the translation tasks. While participants in the main and follow-up study rated the second manipulation technique to be more intuitive, the discrepancies between the QUESI scores [12] for translation and rotation were higher when rotation was performed first (Table 2). This effect is also reflected in the agreement with S4 and S5 (Fig. 6A+B). While in the follow-up study learning the translation technique (S4) was perceived only slightly easier than in the main study, learning the rotation technique (S4)

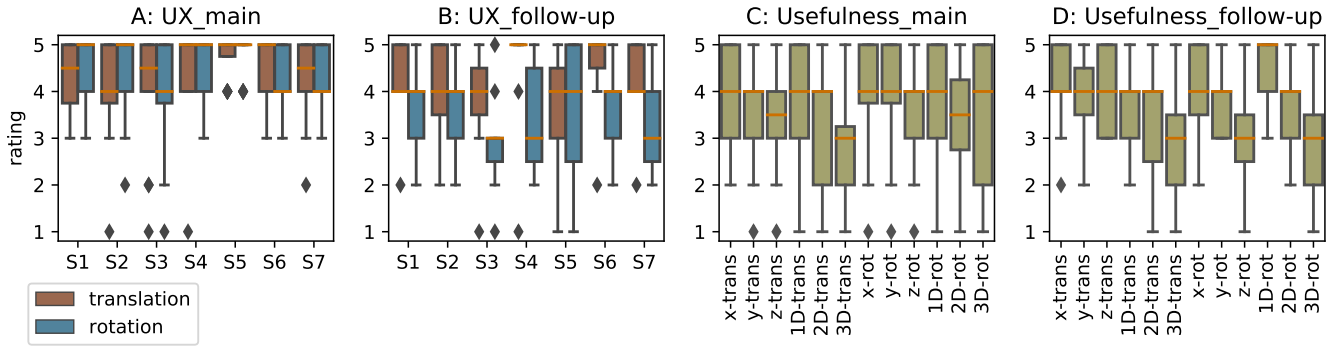

Figure 6: (A+B) Agreement with statements 1-7 regarding *Move'n'Hold* from 1 (fully disagree) to 5 (fully agree): *S1* The translation/rotation technique is well suited to the requirements of the task; *S2* The number of steps to translate/rotate objects is adequate; *S3* When translating/rotating objects, I have the feeling that their behavior is predictable; *S4* Learning the translation/rotation technique was very easy; *S5* Relearning the translation/rotation technique after a lengthy interruption will be easy; *S6* Overall, I am satisfied with this translation/rotation technique; *S7* If I had to translate/rotate virtual objects in the future, I would like to use this technique. (C+D) Rated usefulness of *Move'n'Hold* from 1 (not useful at all) to 5 (very useful).

was perceived much easier if the translation technique was already known than if rotation had to be learned before knowing the translation technique. Furthermore, both translation and rotation were perceived to be much easier to relearn (S5) in the main study as compared to the follow-up study. In general, Fig. 6A+B shows higher discrepancies between the agreement with S1-7 for translation and rotation in the follow-up study than in the main study. While the NASA TLX [1] for translation was slightly lower in the follow-up study (29.1) than in the main study (31.6), the rotation technique's task load was clearly higher when performed prior to translation (41.3) than if the translation technique was already known (32.2). Furthermore, we found decreasing learning effects in the second condition for both groups (i.e., learning effects for translation were less strong in the follow-up (24%, $0.0001 < p \leq 0.001$) than in the main study (42%, $p \leq 0.0001$), and learning effects for rotation were less strong in the main (43%, $0.001 < p \leq 0.01$) than in the follow-up study (53%, $0.001 < p \leq 0.01$)). Based on these findings, we rate *Move'n'Hold* to provide learnability across different manipulation techniques and recommend teaching translation prior to rotation.

7 DISCUSSION

With the design of *Move'n'Hold* we do not only intend to enhance object translation and rotation methods for MR-HHDs independently of each other, but also to introduce a joint interaction paradigm that applies for both translation as well as rotation and consequently reduces the overall cognitive and temporal effort required for task completion. The results from the main and the follow-up study conducted did not only reveal *Move'n'Hold* to be easy and intuitive to learn (RQ5) but also showed that users indeed benefit from the knowledge gained while translating objects with *Move'n'Hold* when they move on to rotating objects and vice versa (see section 6.2). In this context, we also found that learning translation prior to rotation enhanced the overall user experience and thus recommend this teaching strategy for applications that implement *Move'n'Hold*.

Move'n'Hold was found to provide high scalability (RQ6) regarding object translation and rotation tasks differing in distance, direction, and complexity. While all tasks could be completed successfully, *Move'n'Hold* was rated most useful for 1D tasks (see Fig. 6C+D) which could indicate that it is easier for the users to predict an object's position and orientation when it is continuously translated or rotated along or around only one axis. Regarding object rotation, *Move'n'Hold* was rated to be particularly useful for rotations around the x- (right–left) and y-axis (up–down) (see Fig. 6C+D). In contrast to rotations around the z-axis (forward–back), the HHD's display becomes invisible to the user if it is rotated too far around the x- or y-axis. *Move'n'Hold* allows the user to only perform a small rotation around the x- or y-axis and then start continuous movement such that the object remains visible to the user while it is rotating. Thus, we conclude that *Move'n'Hold* effectively solves this usability issue and enhances scalability regarding object rotations in different directions.

Furthermore, we rate *Move'n'Hold* to provide high controllability (RQ7) as the participants of our studies could effectively combine natural and continuous movement to complete the tasks according to their individual preferences and interaction styles as described in section 6.1.3.

In order to ensure the interpretability of the results and keep the number of tasks to be completed by each participant acceptable (50 translation tasks and 50 rotation tasks), we only focused on task configurations that differ in terms of direction, distance, and complexity. While we did not consider different object shapes and sizes in our studies, we expect the translation and rotation with *Move'n'Hold* to be easily applicable to objects in different sizes and shapes. In contrast, the selection of objects that differ in size and shape should be further investigated in the future. Further research should also take into account a more heterogeneous group of users and evaluate the accessibility of *Move'n'Hold*, for instance by expanding the age range of the participants and considering visual or motor impairments.

8 CONCLUSIONS AND FUTURE WORK

In this paper, we presented *Move'n'Hold* – a novel, unified interaction paradigm for object translations and rotations with MR-HHDs. *Move'n'Hold* was designed to improve scalability and controllability which are limited in existing state of the art methods. Addressing these issues, *Move'n'Hold* allows users to individually combine natural and continuous object translations and rotations through a minimalist user interface that relies solely on peripheral touch and device movement. The experimental evaluation revealed *Move'n'Hold* to be an easy-to-learn and intuitive input paradigm, that scales between different user preferences and interaction styles and effectively supports multidimensional small- and large-range object manipulations in different directions. Our results further show that *Move'n'Hold* scales between translation and rotation, i.e., users benefit from their experiences while translating objects and can apply their knowledge when performing rotations and vice versa. Moreover, we found that teaching translation prior to rotation enhanced the overall user experience. In future work, we also plan to use motion sensors, pupillometry, and heart rate monitoring as part of an extended evaluation comparing *Move'n'Hold* with *Move* and other modern interaction techniques in terms of cognitive and physical effort. The design of *Move'n'Hold* is not limited to specific scenarios and suitable for general use: The virtual boxes can be easily replaced with any number and type of virtual items such as furniture or machines for design review and planning tasks, text fields and hints for assistance tasks, or learning contents for educational purposes. Furthermore, *Move'n'Hold* can be used to enhance human robot interaction and collaborations. Hence, we are planning to extend *Move'n'Hold* for multi-selection and apply it to real use cases that combine object translation and rotation such as factory layout planning and robot control.

ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 252408385 – IRTG 2057 at RPTU Kaiserslautern-Landau and UC-Davis. Furthermore, it was supported in part by NASA through the HOME (Habitats Optimized for Missions of Explorations) Space Technology Research Institute at UC-Davis. We thank Sofie Schwenkreis for her support in conducting the experiments and data analysis.

REFERENCES

[1] National Aeronautics and Space Administration. 2022. *NASA TLX Paper and Pencil Version Instruction Manual.* Retrieved January 11, 2023 from https://humansystems.arc.nasa.gov/groups/tlx/tlxpaperpencil.php

[2] Jonas Blattgerste, Kristina Luksch, Carmen Lewa, and Thies Pfeiffer. 2021. TrainAR: A Scalable Interaction Concept and Didactic Framework for Procedural Trainings Using Handheld Augmented Reality. *Multimodal Technologies and Interaction* 5, 7 (2021). https://doi.org/10.3390/mti5070030

[3] Jean Botev, Joe Mayer, and Steffen Rothkugel. 2019. Immersive Mixed Reality Object Interaction for Collaborative Context-Aware Mobile Training and Exploration. In *Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems* (Amherst, Massachusetts) *(MMVE '19)*. 4–9. https://doi.org/10.1145/3304113.3326117

[4] Varunyu Fuvattanasilp, Yuichiro Fujimoto, Alexander Plopski, Takafumi Taketomi, Christian Sandor, Masayuki Kanbara, and Hirokazu Kato. 2021. SlidAR+: Gravity-aware 3D object manipulation for handheld augmented reality. *Computers & Graphics* 95 (2021), 23–35. https://doi.org/10.1016/j.cag.2021.01.005

[5] Eg Su Goh, Mohd Shahrizal Sunar, and Ajune Wanis Ismail. 2019. 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review. *IEEE Access* 7 (2019), 40581–40601. https://doi.org/10.1109/ACCESS.2019.2906394

[6] Jerônimo G Grandi, Henrique G Debarba, Iago Bemdt, Luciana Nedel, and Anderson Maciel. 2018. Design and Assessment of a Collaborative 3D Interaction Technique for Handheld Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Tuebingen/Reutlingen, Germany). 49–56. https://doi.org/10.1109/VR.2018.8446295

[7] Rebecca A. Grier. 2015. How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59, 1 (2015), 1727–1731. https://doi.org/10.1177/1541931215591373

[8] Minseok Kim and Jae Yeol Lee. 2016. Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality. *Multimedia Tools and Applications* 75, 23 (2016), 16529–16550. https://doi.org/10.1007/s11042-016-3355-9

[9] Asier Marzo, Benoît Bossavit, and Martin Hachet. 2014. Combining Multi-Touch Input and Device Movement for 3D Manipulations in Mobile Augmented Reality Environments. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (Honolulu, Hawaii, USA) *(SUI '14)*. 13–16. https://doi.org/10.1145/2659766.2659775

[10] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: a class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*, Hari Das (Ed.), Vol. 2351. SPIE, 282 – 292. https://doi.org/10.1117/12.197321

[11] 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* (Laval, France) *(VRIC '13)*. Article 12, 10 pages. https://doi.org/10.1145/2466816.2466829

[12] Anja Naumann and Jörn Hurtienne. 2010. Benchmarks for Intuitive Interaction with Mobile Devices. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. 401–402. https://doi.org/10.1145/1851600.1851685

[13] Ali Samini and Karljohan Lundin Palmerius. 2016. A Study on Improving Close and Distant Device Movement Pose Manipulation for Hand-Held Augmented Reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. 121–128. https://doi.org/10.1145/2993369.2993380

[14] Goh Eg Su, Mohd Shahrizal Sunar, and Ajune Wanis Ismail. 2020. Device-based manipulation technique with separated control structures for 3D object translation and rotation in handheld mobile AR. *International Journal of Human-Computer Studies* 141 (2020), 102433. https://doi.org/10.1016/j.ijhcs.2020.102433