Design and Assessment of a Collaborative 3D Interaction Technique for Handheld Augmented Reality

Jerônimo G Grandi*[1] Henrique G Debarba†[2] Iago Berndt‡[1] Luciana Nedel§[1] Anderson Maciel¶

[1] Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[2] Artanim Foundation, Geneva, Switzerland

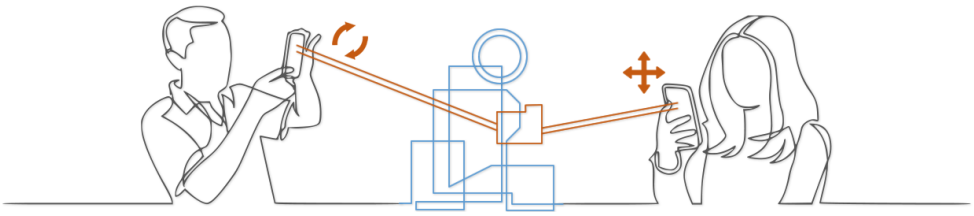


Figure 1: Two users simultaneously manipulating a virtual object in augmented reality. Each user can have a different perspective view of the scene. Rays are drawn from the device to the selected object to inform users the current selection. Virtual icons indicates the current transformation.

Abstract

We present the design of a handheld-based interface for collaborative manipulations of 3D objects in mobile augmented reality. Our approach combines touch gestures and device movements for fast and precise control of 7-DOF transformations. Moreover, the interface creates a shared medium where several users can interact through their point-of-view and simultaneously manipulate 3D virtual augmentations. We evaluated our collaborative solution in two parts. First, we assessed our interface in single user mode, comparing the user task performance in three conditions: touch gestures, device movements and hybrid. Then, we conducted a study with 30 participants to understand and classify the strategies that arise while working in pairs, when partners are free to make their task organization. Furthermore, we investigated the effectiveness of simultaneous manipulations compared with the individual approach.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction techniques; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed/augmented reality Human-centered computing—Collaborative and social computing

1 Introduction

The premise of augmented reality (AR) is to enhance sensory perception through computer-generated information, mainly visual information, from virtual objects to meta-data about the environment. Different AR displays exist, such as head-mounted see-through displays, surface mapped projections and video-mediated rendering. The latter includes handheld devices containing a rear camera and a screen, such as mobile phones and tablets.

Current handheld devices seem to be the ideal device to fulfill AR requirements for everyday applications. They have sensors to capture touch, movement, and image. This allows the use of computer vision and integration of inertial sensors to define the pose of the device in the physical space. Virtual elements then overlay the real world captured by the device's embedded camera directly on the mobile device screen. Touchscreen gestures are widespread among users and are being smoothly adopted to trigger interactive content. Besides, interest for mobile AR increased with the appearance of games such as Pokemon GO, reaching a broad audience. Very recently, Apple and Android released their APIs (ARKit [1] and ARCore [2]) for native AR support in their operational systems.

While some virtual augmentations are informative, others are interactive. Interactive content may be in various formats, but the most common are three-dimensional objects. The manipulation of 3D objects in AR environments is a complex task that requires the control of multiple degrees-of-freedom (DoF) for selecting, translating, rotating and scaling objects. Moreover, virtual objects are intangible, and interaction with them can only be achieved through full-body tracking or mediated by a handheld device. Touch gestures on a 2D screen provide a straightforward method to interact with virtual objects. Alternatively, 3D interactions with the device movements [28] and around-the-device [17], provide a direct mapping between the input and the respective manipulation. Furthermore, combined with touch gestures these interactions provide intuitive, high precision and fast control in multiple DOF.

The widespread availability of mobile phones allows users to access AR through a personal perspective, and to share and collaborate with other users when interacting with virtual content [7, 24]. Since it is interesting that multiple people cooperate in virtual spaces, an interface capable of handling and synchronize inputs of many users for cooperative work is desirable. However, as of today, research in the 3D object manipulation field mainly focus on single-user interaction.

In this work, we present a novel collaborative 3D user interface for virtual objects' manipulation in handheld augmented reality. Our solution creates a shared medium where several users can interact through their points-of-view and simultaneously manipulate 3D virtual augmentations. We integrate touch gestures and device movements into a hybrid manipulation for fast and precise interaction. Our technique handles inputs from several participants with their devices. Participants can either perform manipulations alone

*e-mail: jggrandi@inf.ufrgs.br
†e-mail: henrique.debarba@artanim.ch
‡e-mail: iago.berndt@inf.ufrgs.br
§e-mail: nedel@inf.ufrgs.br
¶e-mail: amaciel@inf.ufrgs.br

[1] https://developer.apple.com/arkit
[2] https://developers.google.com/ar

or manipulate objects together, simultaneously. We implement UI elements to keep users aware of the other's actions. The design to combine all these features makes our approach unique and is the main contribution of the paper.

Besides, we present two experiments. In the first, we evaluated the interaction interface where we compare the single user performance in three interaction conditions: touch gestures, movements and hybrid. Then, we conducted a study to understand and classify the strategies that arise while working in pairs, when partners are free to make their task organization. Furthermore, we investigated the effectiveness of simultaneous manipulations compared with single user manipulations.

2 Related Work

2.1 3D Interaction in Mobile Augmented Reality

Touch gestures are well-established input for object manipulations. Touchscreens are available with various surface sizes and hardware apparatus, such as tabletops and handheld devices. The touch gestures are mapped from the device's 2D screen to 3D transformations and are used as manipulation metaphors [19, 21, 27]. Nonetheless, there is also literature on the mapping of mobile device's sensors and gestures for the selection and manipulation of virtual objects for single users [15], collaborative [12] contexts and in AR environments [31].

The use of the mobile device touchscreen for 3D interaction in mobile AR has been explored to some extent. Notably, Boring et al. [7] used the built-in camera of a mobile phone to directly interact with a distributed computing environment, performing tasks such as selection, positioning and transferring of photos across different displays. Also in mobile AR interfaces for the exploration [11] of medical images in handheld devices and control of a docking object orientation [8]. Tiefenbacher et al. [32] compared the performance of manipulating 3D objects in the camera-, object- and world- coordinate systems. They evaluated the approaches in an augmented reality environment during a docking task with custom transformation gestures. Faster translation time was achieved when using a camera coordinate system, while for rotation the object-centric approach performed better.

The advantage of manipulating 3D objects in augmented reality scenarios using mobile devices is the ability to use the physical movements for interaction. This approach is a natural way to place objects in the scene, as it mimics the real actions. It relies on the quality of the tracking for the amplitude of the movements, but on the other hand, it does not require external apparatus attached to the device. Henrysson et al. [13] were the first to propose the use of movements to manipulate virtual objects. It was possible to alter the object's transformations by changing the device's pose. Samini and Palmerius [28] designed a device movement technique that uses the user perspective rendering approach. The method was compared with a fixed and relative device perspective approaches for near and far objects. Hybrid techniques take the advantages of the device movement and touch gestures. Mossel et al. [23] proposed two techniques, *3DTouch* and *HOMER-S* for one-handed interactions with handheld devices. The first implements touch gestures along with interface widgets to choose one transformation at a time. The *HOMER-S* works with user's movements. The user first selects the action and then moves the device in the physical environment to transform the 3D object. During a manipulation task, the two techniques are integrated and can be combined. Similarly, Marzo et al. [22] combine multi-touch and device movements. However, their interface is designed for two hands interaction.

2.2 Collaborative Manipulation of 3D Objects

The main objective of Collaborative Virtual Environments (CVEs) is to allow multiple users to interact with objects and to share a virtual space. In collaborative spaces important aspects have to be taken into account for the design of interaction methods, such as, awareness of the task and others, co-presence, co-location/remote (whether the users are in the same or different locations), active/passive (whether the users control the interaction), synchronous/asynchronous (whether users control the same object at the same time), symmetrical/asymmetrical interaction (whether users have the same interaction capabilities) [1, 20, 25]. Another important aspect in collaborative virtual environments is the network latency. While this is crucial for reliable interactions in VEs, it is not the focus of this work. An in-depth survey about network latency in VEs is presented by Khalid et al. [16].

Early works that apply collaborative aspects in augmented reality environments were compiled by Billinghurst and Kato [5]. We are more interested in works that explore the synchronous approach, where two or more users manipulate the same object at the same time. A study conducted by Aguerreche et al. [3] compares three main synchronous manipulations for collaborative 3D objects manipulation in virtual environments: collaborative tangible device, proposed by themselves [2], DOF separation [26] and mean average of the actions [9]. Grandi et al. [12] designed an handheld-based interface for collaborative object manipulation for shared displays. They conducted an experiment to compare the performance of different group sizes during synchronous manipulation tasks.

While prior work focuses on single user manipulations in AR scenarios, we allow users to cooperate by sharing the transformation tasks. Moreover, we propose a solution for fluid simultaneous manipulations of a same 3D object, without the need to block or divide the actions. Thus, groups can adopt their own task organization without system limitations. In Section 5, we investigate the effect of simultaneous manipulations during pair work.

3 Design of a Collaborative Handheld AR Technique

We propose a flexible set of actions for manipulation using handheld-based devices. It engenders the availability of *touch gesture* and *device movement* inputs that aid both precise and fast spatial transformations in augmented reality scenarios. The user decides the most appropriate input depending on the task needs. A simple touch+hold action on the screen surface switches between touch gesture manipulation (Sec. 3.1) and device movement manipulation (Sec. 3.2). Moreover, our technique was designed to support an unlimited number of simultaneous participants. All users have access to all available functions and can apply transformations to different objects or to the same object simultaneously while observing the scene from a different point of views.

3.1 Touch Gestures Manipulations

We convert finger gestures on the touchscreen into 3D transformations to manipulate a total of 7 DOF of a selected 3D object. More specifically, there are 3 DOF for translation, 3 DOF for rotation and 1 DOF for uniform scale. The transformations are applied relative to the touchscreen plane orientation (i.e., the device orientation), similarly to the proposed by Grandi et al. [12] and Katzakis et al. [15] (Figure 2). The transformations are performed using one and two fingers. We based our touch gestures implementation on the *DS3* technique [21] with variations. The touch and slide of one finger in the device *xy* orientation plane move the object in the same direction as the finger slide. The gesture sequence of one tap followed by a touch and slide of one finger horizontally moves the object towards the *z* device axis, unlike the *DS3* that uses another finger for *z* translation (Figure 2a-b). Two fingers touch and slide enables rotation. The slide of two fingers in a horizontal direction affects yaw, vertical sliding changes pitch and pivoting affects roll rotations (Figure 2c-d). Finally, we add the pinch and spread of two fingers to uniformly modify the object's scale regardless of the screen plane orientation (Figure 2e). The rotation and scale ges-

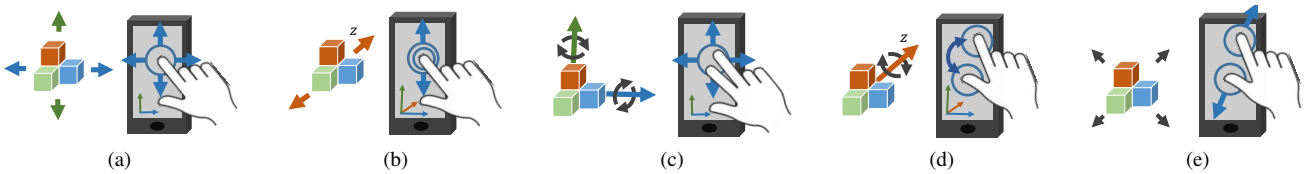


Figure 2: Manipulations performed over the selected object: (a) A touch and slide translates in xy axis, (b) one tap followed by a touch and slide translates in z axis, (c) touch and slide with two fingers rotates in xy axis, (d) touch and rotate two fingers rotates in z axis and (e) pinch and spread of two fingers to uniformly modify the object's scale.

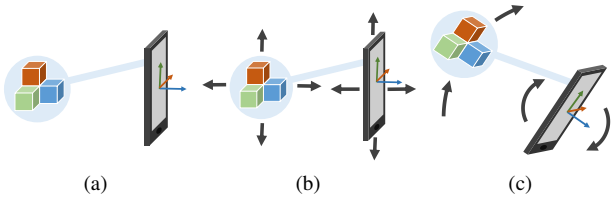


Figure 3: Manipulations performed on the attached object: (a) object-device attachment, the object and the device keep an invariant rigid transformation. In this way, the object translates (b) and rotates (c) with the device movements.

tures are not combined. Thus, it is necessary to release the fingers and start another gesture to change the transformation. All transformation modes are correctly applied depending on the gesture. Therefore, there is no necessity for UI buttons to switch modes.

3.2 Device Movements Manipulations

This approach attaches the object to the physical pose of the device [28]. The transformation with movements is activated by pressing and holding a circular button in the lower-right corner of the interface and is halted once the button is released. While a finger is pressing the button, the object translates and rotates with the device while keeping an invariant rigid transformation relative to it (Figure 3). We do not allow changes in the transform rate and transformations decouple to preserve the absolute mapping. Clutch can be used to reach a total rotation beyond arms limits (i.e., perform object rotation, reposition the device, then perform a new object rotation). Depending on the object-device bound distance, the rotation affects the object position, as shown in Figure 3c. While moving the object, it is possible to slide another finger on the screen closer or far away to scale the object.

The equation $T = V(V^{-1}V_{prev})V^{-1}$ defines the matrix calculation to transform the object during the manipulation. The transformation first converts the model to the camera coordinate system, symbolized by the matrix V. After, we apply to V the same transformation that the camera suffered during the last iteration, for that, the view matrix of the previous frame is used, represented by V_{prev}. Finally, the transformation is multiplied by the inverse of the view matrix to return the model to the world coordinate system.

3.3 Simultaneous Manipulations

Our technique supports simultaneous manipulation of virtual objects by multiple users. While working in groups, users can either independently manipulate objects or interact simultaneously with the same object. Thus, cooperative strategies can be established depending on the task needs and the ability of the team members.

When two or more users are manipulating the same object, the actions performed by each individual counts as a transformation step. We multiply each user transform matrix by the virtual object transform matrix. Thus, every contribution from each user is summed up in the final object's transformation without restrictions or weights (Figure 4a). Therefore, if two users move the object in opposite directions, the position of the object will not change (Figure 4b). On the other hand, if they manipulate the different transformations in parallel, the transformations are combined (Figure 4c). The simultaneous manipulations can occour with any combination of *touch gestures* and *device movements*.

We added two virtual elements to make users aware of the other participants' actions in the virtual scene. Regarding selection, we draw rays from the device location to the currently selected objects. The ray informs about the focus of interest of other users interacting in the same AR environment, as shown in Figure 1. We distinguish between colors the user selection ray and the other participants' selection rays. We also render icons on the virtual rays indicating the transformation being performed by each user, so that users can be aware of each other's actions without the need for verbal communication. The icons can indicate that a mobile phone is either performing a movement or translation, rotation or scale with touch on the selected object.

4 Experiment 1: Single User Assessment

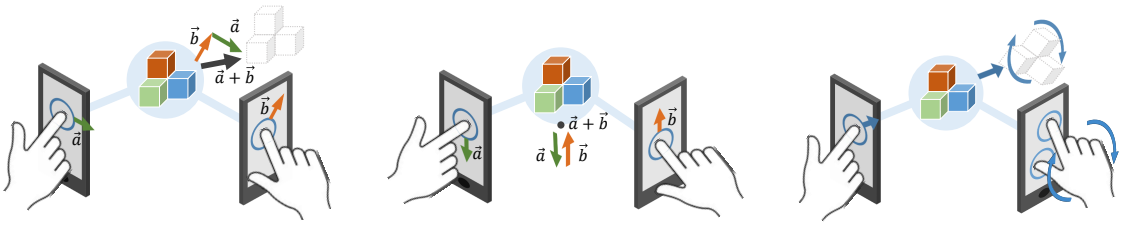
Our goal is to compare task completion time and error rate between the three 3D manipulation methods: *Touch gestures*, *Device movements* and *Hybrid*. Thus, we carried this experiment with a single user. We hypothesize the *Hybrid* technique to be the fastest and will have the lowest error rate. Theoretically, the *Device movements* technique has a time advantage over the *Touch gestures* since the manipulation is analogous to a person carrying an object. However, the physical effort required by the *Device movements* technique may introduce more error during precise positioning, which is less expected with the *Touch gestures* technique.

4.1 Task and Stimuli

We designed a 3D docking task that comprises translation in 3-DOF and orientation in 3-DOF. A docking task consists of transforming a virtual object to a target position and orientation, and it is a widely adopted task to evaluate interfaces and techniques for spatial manipulations [10, 33]. In our experiment, we asked participants to dock a virtual *moving piece*, controlled by the user with a similar virtual *static piece*. Both *moving piece* and *static piece* had the color matched. The *static piece* was 50% semi-transparent while the *moving piece* was opaque (Figure 7).

The pieces configuration stimuli are composed of cube blocks similar to the Shepard and Metzler [30] construction. The blocks are assigned with different colors to avoid ambiguity when matching the target piece (Figure 7). The blocks have 6cm long edges.

For each docking task, only one *moving piece* and their respective *static piece* appears in the scene. We placed the *static piece* always in the center of the scene while the *moving piece* has four possible spawn positions. The distance between the two pieces at the start of a trial is fixed at 35cm.



(a) Forces addition (b) Forces cancellation (c) Transformations merge

Figure 4: Concurrent access to transformations. Every action performed by each user counts as a transformation step. In (a) and (b) the final translation is the sum of the transform vectors. If users apply a translation in opposite directions, the object will not move. If they manipulate different transformations in parallel, the transformations are merged (c).

We render shadows and apply collision forces between the piece and the table to provide additional depth cues and interaction with the physical environment.

4.2 Apparatus

The experimental setup is composed of an Apple iPad mini 4 with 7.9 inches screen (approx. 324 ppi pixel density) and weights 299g. The 8MP rear camera and the Vuforia SDK [3] were used to track the mobile device physical pose relative to the fiducial markers. We used multiple markers and the *extended tracking* feature to extend the interaction range. These markers were placed on top of a table. We used a server application to manage and send to the mobile device the experiment parameters (modality order, trials randomization, and piece spawn position) and to record all user interactions during the experiment, a dedicated WiFi connection and a table with patterns for tracking. Both server and client's application were developed using the Unity3D game engine [4] and the communication is made with the Unity UNET network API.

4.3 Subjects

Twenty subjects participated voluntarily in this experiment (six female), aged 25.05 years in average (SD=3.27). All subjects read and agreed to an Informed Consent Form before the experiment. They were all Computer Science students. Two of them reported minimal movement restrictions on the wrist and finger that did not affect their performance during the experiment. They all had either normal or corrected to normal vision.

4.4 Experimental Setup

The experiment follows a repeated measures within-subject design with *Technique (Touch gestures, Device movements and Hybrid)* and *rotational angles (45° and 90°)* as the independent variables. The dependent variables were *time* and *accuracy* (position and orientation) to complete each docking. We have also collected the user *transformation actions (translation, rotation)* and the *user's physical positions* in the environment for post hoc association with the aforementioned dependent and independent variables.

The participants answered a characterization form and performed a mental rotation test at least one day before the experiment. The mental rotation experiment used is similar to that used by Shepard and Metzler [30]. The test was used to assess the participant's ability to understand 3D rotations. In the experiment session, before the trials,the participants were guided to experiment the input commands. In this phase, only one object was displayed and the participant had no target objective.

The presentation order of the three conditions was counterbalanced that resulted in 6 different group orders. For each input technique, participants had one practice session composed of two docking trials and a recorded session composed of eight docking trials.

In the first practice trial, we displayed, in real-time on the device's screen, the actual values of position and orientation errors. The text values changed colors (from white to green) for each parameter to inform when a threshold (1.5cm and 15° difference [33]) was achieved. For the second practice trial, the reference position and orientation errors were displayed after the participant confirm and finish the docking. Then, participants were asked to perform the eight valid trials. We asked participants to balance accuracy and speed. No reference errors were displayed during the recorded trials to avoid a bias toward accuracy [14]. Participants were orally informed of their progress when four and two trials were missing to complete the block of the trials. Each trial started when the user selected the virtual object and finishes when the user confirms the docking by pressing a button in the lower left corner of the device's screen. Each virtual piece appears in sequence after the previous docking confirmation. It was possible to select only the *moving piece* and once selected it could not be unselected. After the recorded session, participants were allowed to rest and were asked to assess their workload level with the NASA's Task Load Index [5].

In summary, the experiment consisted of: **20** participants × **3** techniques × **8** trials (4 - 45° and 4 - 90° of rotational difference) = **480** unique docking.

At the end of the experiment participants answered the Single Easy Question (SEQ) [29] for each manipulation condition ("Overall, How difficult was the tasks with *condition*?"). The SEQ was rated on a 7-point Likert scale, ranging from 1 ("Very Hard") to 7 ("Very Easy"). Then, we applied a System Usability Score (SUS) questionnaire to assess the overall usability of the experimental setup. The experiment sessions lasted approximately 50 minutes.

4.5 Results

4.5.1 Accuracy and Time

We removed 3 trials where subjects failed to attain the minimal precision of 5 cm and 15° relative to the reference docking object. For the statistical analysis, we take the median of the time, translation error and rotation error for each combination of method and rotation angle per subject. Repeated measures ANOVA was used to test the statistical significance of the manipulated factors. For the precision analysis, we consider the precision attained by the subjects when they indicated that they were satisfied with the docking.

Figure 5 shows the time by error reduction rate for position and rotation. The smaller amount of time needed to reduce the error suggests that *Movement* manipulation is more efficient for positioning the object, while *Touch* manipulation is more efficient for rotating the object. It also suggests that subjects could take advantage of

[3] www.vuforia.com
[4] www.unity3d.com
[5] https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLXScale.pdf

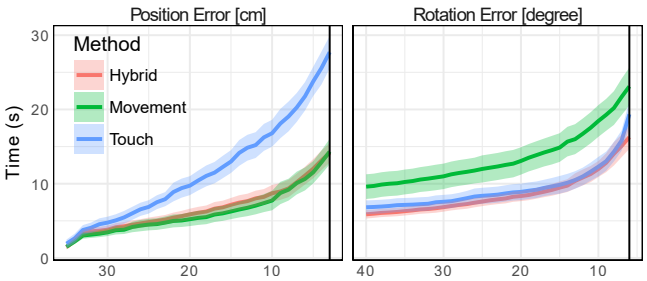


Figure 5: Error reduction by time for position and rotation. The faster the error is reduced, the better. The shaded areas represent the standard error of the mean.

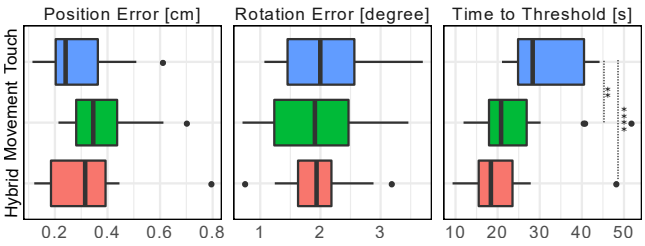


Figure 6: We found a statistically significant advantage of Hybrid over Movement and Touch, and of Movement over Touch for the time using a 1.15cm and 8° threshold. There is no significant difference in position error and rotation error between interaction methods.

the *Hybrid* approach, presenting an error reduction rate very similar to *Movements* regarding positioning, and *Touch* in term of rotation. We test the statistical significance for the point in time when subjects achieved a minimal precision of 1.15cm and 8°, these values were chosen as they represent the performance attained in every trial of the experiment. Figure 6 shows the time to reach the threshold for each method and the position and rotation errors achieved when users were satisfied with the docking. We found a statistically significant difference for input method ($F_{(2,38)} = 24.9$, $p < .001$) but not for the initial rotation factor ($F_{(1,19)} = .5$, $p > .48$), nor their interaction ($F_{(2,38)} = .03$, $p > .96$). Post-hoc T-test of the method indicates that *Hybrid* was more efficient than *Movement* ($t_{(19)} = 3.5$, $p < .005$) and *Touch* ($t_{(19)} = 6.7$, $p < .001$) manipulation, and that *Movement* had advantage over *Touch* ($t_{(19)} = 4.2$, $p < .001$). Our results point that subjects effectively took less time when interacting with the hybrid approach.

On the other hand, we found no significant effect of *Technique* on position and rotation error. That is, subjects could dock the objects with similar precision regardless of the *Technique* in use.

4.5.2 Single Easy Question, Workload and Usability

Non-parametric Friedman tests were conducted to compare the effect of the three manipulation techniques (*Hybrid, Movements and Touch Gestures*) to the NASA TLX workload questionnaire and the SEQ. The post-hoc analysis was conducted with Wilcoxon signed-ranks test with a Holm-Bonferroni correction for multiple comparisons.

For the SEQ score, we found a significant effect of manipulation condition ($X^2(2) = 27.757$, $p < .001$). Post-hoc indicates a significant difficulty increases from *Hybrid* to *Movements* ($p < .001$), from *Hybrid* to *Touch Gestures* ($p < .001$) and from *Movements* to *Touch Gestures* ($p < .001$). The *Hybrid* was ranked as the easiest ($Med = 6$ $IQR = .25$) and the *Movements* as the hardest ($Med = 3$ $IQR = 1$), while *Touch Gestures* ($Med = 5$ $IQR = 1$) was ranked between two other conditions.

The result of the NASA TLX user workload for each tested condition revealed a significant effect between conditions ($X^2(2) = 14.354$, $p < .001$) with *Hybrid* $Med = 61$,$IQR = 17.5$, *Movements* $Med = 70.7$, $IQR = 12.8$ and *Touch Gestures* $Med = 62.7$, $IQR = 16.5$. The post-hoc test indicates that significant workload increase occurs from *Hybrid* to *Movements* ($p < .006$) and from *Touch Gestures* to *Movements* ($p < .007$). No significant workload differences occurs between *Touch Gestures* and *Hybrid* ($p > .7$).

An analysis of each individual NASA TLX factor revealed a significant effect on Performance ($X^2(2) = 6.51$, $p < .04$), Effort ($X^2(2) = 7.42$, $p < .03$) and Frustration ($X^2(2) = 14$, $p < .001$). Significant Effort factor increase occurs from *Hybrid* to *Movement* ($p < .025$) and from *Touch Gestures* to *Movement* ($p < .02$). Significant Frustration factor increase occurs from *Hybrid* to *Movement* ($p < .003$) and from *Touch Gestures* to *Movement* ($p < .046$). No significant difference was found for Performance on the Post-hoc analysis.

The SUS of our experimental setup ranged from 65 to 100 (M=77.12, SD=9.5). According to surveys that compare SUS scores for different systems, the system is ranked as a "Good" [4].

5 Experiment 2: Pair Work Assessment

The pair work assessment described here focuses on the evaluation of collaborative aspects when two users are manipulating virtual objects in the same scene. Different from works that impose and compare different collaborative strategies [18], we aim at observing and classifying the strategies that emerge when users are free to make their task organization. During public demonstrations of our collaborative AR interface, we observed that groups often adopted different strategies to accomplish a constructive task. Thus, in this experiment, we propose to control the level of occlusion in the augmented scene to observe how interaction strategies change, and how these strategies compare regarding performance.

We expect that users will tend to organize themselves depending on the level of occlusion in the augmented scene. We hypothesize that two strategies will appear: *Independent Interaction*, where users divide the problem and each solves part of it as in a single user approach, and a *Shared Interaction*, where the task is performed sequentially with both users focusing attention on the same sub-task. Moreover, we would like to analyze if trials where users apply the *shared interaction* approach will have any effect on time when compared with trials performed separately as in single user mode.

5.1 Task and Stimuli

The virtual scene setup of the previous experiment was reused for this experiment. We added collision detection between pieces and gravity attraction besides the already included shadows to make the pieces act similarly as physical blocks. Differently from the previous experiment, where only one pre-selected *movable piece* and one *static piece* were shown at a time, here, all *movable pieces* are shown and the users need to select and dock the correct piece. Each *static piece* property – spawn position, rotation and scale – is randomly chosen from a list of valid transformation. Since scale is introduced in this experiment, we defined four possible scales: 60%, 80%, 120%, 140% of the *movable piece* size. When one *movable piece* is docked, both the *movable piece* and the respective *static piece* disappear and a new *static piece* appear in the next docking space. The task is completed after all pieces are docked.

To stimulate cooperative work, we created three conditions where we vary the level of occlusion in the scene. One with no occlusions, the second with moderate occlusion, and the last condition with high occlusion. For that, we added virtual walls on the augmented scene to force users to look for new vantage points from which the objects are not occluded (Fig. 7). The working space had 57x120cm and was divided into a 57x40cm docking space, where the *demands* appear, and the supplies space (57x80cm), where the

18 pieces are initially placed. The walls are 24cm high (the exact height of two stacked blocks) and 2cm thick. The longer walls have 80cm and the shorter walls have 57cm. The moderate occlusion space is divided into two 27.5x78cm partitions, while the highest occlusion condition has eight 23x17.5cm partitions.

5.2 Apparatus

This experiment was conducted with a couple of Apple iPad Air 2 that only differ from the iPad mini 4 of the previous experiment on a screen size that is 1.8 inches larger (9.7 inches) and is 138g heavier (437g). The device replacement was necessary due to the availability of two identical devices.

5.3 Subjects

Thirty subjects voluntarily took part in this experiment (thirteen female), aged 22.2 years in average (SD=2.93). They were all students with no movement restrictions on wrists and arms. We arranged the participants in pairs. They were allowed to choose their partners. Two pairs had never met before. Five subjects participated in the previous experiment (single user), two of them together.

5.4 Experimental Setup

The experiment followed a repeated measures within-subject design with the *Occlusion (No Occlusion, Moderate Occlusion and High Occlusion)* as the independent variable. The dependent variables collected were total piece manipulation *time* needed to complete each docking, and *manipulation time* of each user in each docking. Based on the results achieved in the *Experiment 1* (Sec. 4), we adopted the *Hybrid* manipulation during this experiment.

The participants answered a characterization form before arriving at the experiment. We explained the interface operation and allowed the users to practice the transformations and train in two docking trials. Then, the participants performed 3 blocks of 8 recorded trials. Latin squares determined the presentation order of the trial blocks with 6 different group orders. We asked the pairs to complete each docking as fast as possible. No reference errors were displayed during the recorded trials. The threshold error for a successful docking was 1.15cm in position, 8° in the rotation and 1cm in size. After each block of trials, the users answered a SEQ to assess the task difficulty and questions about behavioral interaction, mutual assistance and dependent action which encompasses the behavioral engagement factor of the *Networked Minds Measure of Social Presence* [6]. In the end, users filled a final form with custom questions and with questions about psychological involvement factor ([6]). The experiment took on average 50 minutes.

In summary, the experiment design had: **30** participants $\times$ **3** levels of occlusion $\times$ **8** trials = **720** unique docking.

5.5 Results

5.5.1 Strategies

We asked participants two questions in the post-test questionnaire about the group strategy: "What was the strategy adopted by your team?" and "Has the strategy changed along the experiment?". Ten groups of fifteen (66.6%) answered the first question by saying that they choose to work together to solve the same piece, while five groups – IDs 1, 6, 7, 10 and 11 – (33.3%) said they adopted division strategy, where each piece was solved by one user alone.

We analyzed how pieces were manipulated during the trials to verify if the strategy pointed in the questionnaire was consistent with the behavior of the pair during the trials. We calculated participation score for each trial that represents the balanced participation of both subjects in the manipulation of a single piece: $participation = 1 - abs(\frac{Time_{User1}}{TotalTime} - \frac{Time_{User2}}{TotalTime})$. A score of 1 represents an equal time of manipulation, while a score of 0 means that a single user carried the task for a given piece. Figure 8 shows the distribution of participation scores of the pieces for each group. The Figure shows that our participation score effectively captured the work strategy reported by users.

Moreover, we evaluated the behavioral engagement dimension of the *Networked Minds Measure of Social Presence* [6] against the two strategies adopted to observe if participants feel more engaged when docking simultaneously the same piece. Figure 9 shows the level of behavioral engagement of each strategy. The Wilcoxson signed-ranks test indicates a significant effect of the behavioral engagement on the strategies ($Z = 225$, $p < .001$). The ANOVA test of the behavioral engagement and participation scores shows that they are closely related ($F_{(1,13)} = 19.81$, $p < .001$), and validate our participation score as a proxy to the pair engagement while performing the experiment.

Finally, we have investigated whether the participation score explains the variation of the docking time ($TotalTime$) and the total manipulation time ($Time_{User1} + Time_{User2}$) of the pieces using one-way ANOVA.The test failed to reject the equality of mean docking time and piece manipulation time across the range of computed participation scores ($F_{(1,13)} = .18$, $p > .67$ and $F_{(1,13)} = .65$, $p > .43$ respectively). That is, there is no strong evidence that the manipulation of the same piece by both users interferes with performance.

5.5.2 Occlusion vs. Task Time and Manipulation Time

The ANOVA test revealed a statistically significant effect of occlusion on task completion time ($F_{(2,28)} = 9.53$, $p < .001$). The post-hoc t-test indicates significant time increase occurs for *High Occlusion* as compared to *No Occlusion* ($t_{14} = 2.9$ $p < .03$) and *Moderate Occlusion* ($t_{14} = 4$ $p < .004$). Equivalence could not be rejected for between *No Occlusion* and *Moderate Occlusion* ($t_{14} = .6$ $p > .5$).

The statistical test failed to reject equivalence in the time required to manipulate and dock the object across the levels of occlusion ($F_{(2,28)} = 1.0$, $p > .37$). This indicates that the difference in task time is due to the added search time caused by high level of occlusion.

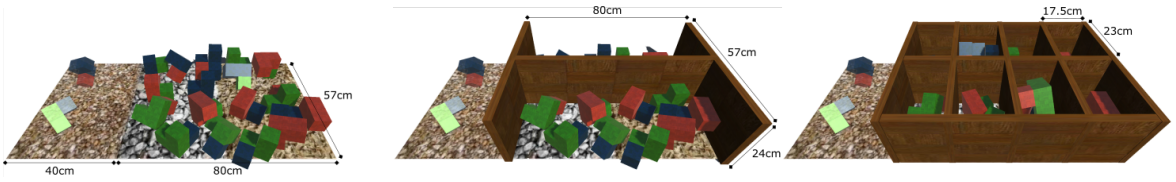
6 Discussion

6.1 Interaction Technique

The finding in our first experiment indicates that the *Hybrid* method is the most suitable for 6-DOF manipulations. We went further and observed that the time performance of the *Hybrid* is as good as the *Movements* for positioning and as good as *Touch Gestures* for rotations. This suggests that users can effectively coordinate the use of the most suitable method for each transformation and that the change between modes is made intuitively and seamlessly without the need of context-aware techniques [23]. We also find that users reach similar precision with all methods when they have to indicate when they were satisfied with the docking. The precision similarity was possible because we let the users change their point-of-view. In situations where users have limited movements and need to manipulate distant objects, the *Movements* method alone is unusable [28]. The lower workload reported for *Hybrid* and the SEQ corroborate with the performance analysis. The highest workload was achieved with *Movements*. The effort and frustration factors significantly affected the *Movements* workload as users have to move the device to transform the objects position and orientation. Users found it easier to manipulate with *Hybrid* and harder with *Movements*, while *Touch Gestures* received an intermediate score. *Touch Gestures* was the slowest to reach the threshold.

6.2 Pair Work Strategies

As hypothesized, we have observed that pairs adopted two main strategies. The 66.6% of the pairs manipulated the objects simultaneously, which we called *Shared Interaction*. The remaining pairs manipulated the objects individually, which we called *Independent Interaction*. The time performance between the two groups was



(a) (b) (c)

Figure 7: The three occlusion conditions presented in the experiments. (a) no occlusion, (b) moderate occlusion and (c) high occlusion. On the left of each condition is the *docking space* (40cm) and on the right is the *supplies space* (80cm).

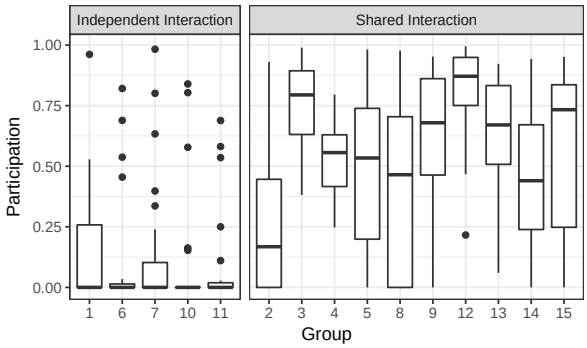


Figure 8: Participation score to complete the trials for each group. Higher scores represent more simultaneous manipulations. Groups are classified by their strategy reported in the post-test questionnaire.

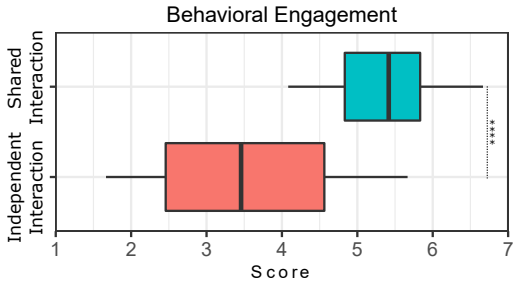
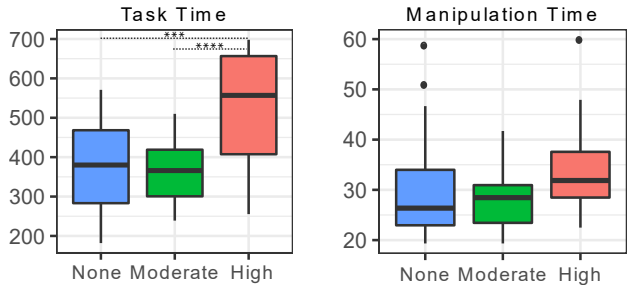


Figure 9: The users behavioral engagement in the two strategies. The results are reported in the Likert scale where 1 is low engagement and 7 is high engagement. Shared Interaction (M = 5.32, SD = 0.58), Independet Interaction (M = 3.57, SD = 1.12).

similar regardless of the strategy. The experience of the participants with non-conventional devices was also similar between groups: M=2.35 SD=1.1 for the *Shared Interaction* group and M=2.35, SD=0.8 for the *Independent Interaction* group. It suggests that the user experience with non-conventional devices did not affect the strategic decision. Moreover, groups did not change their behavior with the increase of occlusion. These results indicate that the strategies are less related to the environmental factors and more related to the users and pairs profile.

We investigated the participants' behavioral engagement in the different strategies. The behavioral engagement is the degree users believe that their actions are interdependent, connected or in response to the other's actions [6]. The results showed that pairs that adopted the shared interaction felt more involved in the task than pairs that adopted the independent strategy, even though, in both cases, they were working together.

A collaborative interface provides each user with an individual



(a) (b)

Figure 10: Time vs Occlusion conditions. Task time is the time taken to complete the task. Manipulation time is the time spent by the users with manipulation commands to dock the pieces.

action perspective of the same scene. The individual viewpoints allow users to place themselves on key locations, avoiding the need for constant movement to check occluded parts. Groups may adopt different strategies to complete the tasks depending on their profile and task requirements, while maintaining similar precision and time performances.

7 Conclusion

In this paper, we presented the design of a novel user interface for collaborative manipulation of 3D augmentations superimposed on the physical environment. The technique was designed for handheld devices as they are versatile and ubiquitous in the everyday tasks. We designed two modes for 3D manipulations, touch gestures and device movements, which combined allow for intuitive 7-DOF transformations. The technique creates a shared medium where multiple users can simultaneously interact with their own devices. The technique solves the problem of concurrent manipulations with an action coordination approach, where every contribution from each user counts as a transformation step and is applied directly to the object.

More than fifty participants have tested the interface in public demonstrations. They could download the app in their smartphones and join on-going demos. Up to five team-members using devices of different models were registered during the demos. All of them easily and quickly understood the purpose and mechanics of the technique, indicating a high affordance. The observation of groups during the demos inspired the design of the two experiments.

We demonstrated the effectiveness of the *Hybrid* approach when compared with solely *Touch Gestures* or *Movements* methods in the interface assessment. Moreover, we observed that users could seamlessly switch between methods and use the most efficient action to correctly transform the object while keeping high time performance. In the second experiment with pairs, we observed that two strategies were adopted when we do not impose any restriction

to collaboration. Interestingly, pairs with different strategies obtained similar task performance while teams that privileged *Shared Interaction* strategy felt more engaged in the task.

In this study, both single user and pair work has been assessed while interacting with 3D objects in an AR environment. In future work, it would be interesting to investigate how users are affected by the use of diverse AR/VR hardware such as HMDs and how designers can incorporate interaction constraints for larger groups to mitigate user errors. Especially during simultaneous manipulations where concurrently *device movements* manipulations can cause the object attached to the device to disappear during the transformation. Future investigations could also explore the support of multiple users, from same or different locales, and assess network latency and its influence in group performance.

ACKNOWLEDGMENTS

We thank all the users that volunteered for the experiment. Thanks are also due to Victor A. J. Oliveira for his help with the text and analysis revision. We finally acknowledge the funding from CAPES, CNPq (311353/2017-7), and FAPERGS (17/2551-0001192-9).

REFERENCES

[1] L. Aguerreche, T. Duval, and A. Lécuyer. 3-Hand Manipulation of Virtual Objects. In *Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments*, JVRC'09, pages 153–156, Aire-la-Ville, Switzerland, 2009. Eurographics Association.

[2] L. Aguerreche, T. Duval, and A. Lécuyer. Reconfigurable tangible devices for 3D virtual object manipulation by single or multiple users. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, page 227, New York, USA, 2010. ACM Press.

[3] L. Aguerreche, T. Duval, and A. Lécuyer. Evaluation of a Reconfigurable Tangible Device for Collaborative Manipulation of Objects in Virtual Reality. In U. E. Chapter, editor, *Theory and Practice of Computer Graphics*, pages 81–88, Warwick, UK, Sept. 2011.

[4] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123, May 2009.

[5] M. Billinghurst and H. Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, 2002.

[6] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th Annual Int. Workshop on Presence*, pages 1–9, 2001.

[7] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch. Touch projector: mobile interaction through video. *SIGCHI Conference on Human Factors in Computing Systems*, pages 2287–2296, 2010.

[8] H. Debarba, J. Franz, V. Reus, A. MacIel, and L. Nedel. The cube of doom: A bimanual perceptual user experience. *IEEE Symposium on 3D User Interfaces 2011, Proceedings*, pages 131–132, 2011.

[9] T. Duval, A. Lécuyer, and S. Thomas. SkeweR: A 3D interaction technique for 2-user collaborative manipulation of objects in virtual environments. In *3DUI 2006: IEEE Symposium on 3D User Interfaces 2006 - Proceedings*, volume 2006, pages 69–72, 2006.

[10] B. Froehlich, J. Hochstrate, V. Skuk, and A. Huckauf. The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, page 199, 2006.

[11] J. Grandi, A. Maciel, H. Debarba, and D. Zanchet. Spatially aware mobile interface for 3D visualization and interactive surgery planning. In *2014 IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–8. IEEE, may 2014.

[12] J. G. Grandi, H. G. Debarba, L. Nedel, and A. Maciel. Design and Evaluation of a Handheld-based 3D User Interface for Collaborative Object Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pages 5881–5891, New York, New York, USA, 2017. ACM Press.

[13] A. Henrysson, M. Billinghurst, and M. Ollila. Virtual object manipulation using a mobile phone. In *Proceedings of the 2005 international conference on Augmented tele-existence - ICAT '05*, page 164, New York, New York, USA, 2005. ACM Press.

[14] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell. Usability Analysis of 3D Rotation Techniques. *Proceedings of the Symposium on User Interface Software and Technology*, pages 1–10, 1997.

[15] N. Katzakis, R. J. Teather, K. Kiyokawa, and H. Takemura. INSPECT: extending plane-casting for 6-DOF control. *Human-centric Computing and Information Sciences*, 5(1):22, dec 2015.

[16] S. Khalid, S. Ullah, and A. Alam. Optimal Latency in Collaborative Virtual Environment to Increase User Performance: A Survey. *International Journal of Computer Applications*, 142(3):35–47, may 2016.

[17] M. Kim and J. Y. Lee. Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality. *Multimedia Tools and Applications*, pages 1–22, 2016.

[18] C. Liu, O. Chapuis, M. Beaudouin-Lafon, and E. Lecolinet. Shared Interaction on a Wall-Sized Display in a Data Manipulation Task. *Proceedings of the 34th international conference on Human factors in computing systems*, pages 1–12, 2016.

[19] J. Liu, O. K. C. Au, H. Fu, and C. L. Tai. Two-finger gestures for 6DOF manipulation of 3D objects. *Eurographics Symposium on Geometry Processing*, 31(7):2047–2055, 2012.

[20] D. Margery, B. Arnaldi, and N. Plouzeau. A general framework for cooperative manipulation in virtual environments. *Virtual Environments*, 99:169–178, 1999.

[21] A. Martinet, G. Casiez, and L. Grisoni. Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. *IEEE Transactions on Visualization and Computer Graphics*, 18(3):369–380, 2012.

[22] A. Marzo, B. Bossavit, and M. Hachet. Combining multi-touch input and device movement for 3D manipulations in mobile augmented reality environments. In *Proceedings of the Symposium on Spatial User Interaction*, pages 13–16, New York, USA, 2014. ACM Press.

[23] A. Mossel, B. Venditti, and H. Kaufmann. 3DTouch and HOMER-S: Intuitive Manipulation Techniques for One-Handed Handheld Augmented Reality. *Proceedings of the Virtual Reality International Conference on Laval Virtual - VRIC '13*, page 1, 2013.

[24] J. Müller, R. Rädle, and H. Reiterer. Virtual Objects As Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load. *Proceedings of the Conference on Human Factors in Computing Systems*, pages 1245–1249, 2016.

[25] T. T. H. Nguyen and T. Duval. A survey of communication and awareness in collaborative virtual environments. In *International Workshop on Collaborative Virtual Environments*, pages 1–8, 2014.

[26] M. S. Pinho, D. A. Bowman, and C. M. D. S. Freitas. Cooperative object manipulation in collaborative virtual environments. *Journal of the Brazilian Computer Society*, 14:54–67, 2008.

[27] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2D and 3D direct manipulation. *Symposium on User Interface Software and Technology*, page 9, 2009.

[28] A. Samini and K. L. Palmerius. A study on improving close and distant device movement pose manipulation for hand-held augmented reality. *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology - VRST '16*, pages 121–128, 2016.

[29] J. Sauro and J. S. Dumas. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 1599–1608. ACM, 2009.

[30] R. N. Shepard and J. Metzler. Mental Rotation of Three-Dimensional Objects. *Science*, 171(3972):701–703, feb 1971.

[31] T. Tanikawa, H. Uzuka, T. Narumi, and M. Hirose. Integrated view-input interaction method for mobile AR. *2015 IEEE Symposium on 3D User Interfaces, 3DUI 2015 - Proceedings*, pages 187–188, 2015.

[32] P. Tiefenbacher, S. Wichert, D. Merget, and G. Rigoll. Impact of Coordinate Systems on 3D Manipulations in Mobile Augmented Reality. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 435–438, New York, USA, 2014. ACM Press.

[33] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock. Evaluation of Docking Task Performance Using Mid-air Interaction Techniques. *Proceedings of the Symposium on Spatial User Interaction*, pages 44–52, 2015.