



Article

Wand-Like Interaction with a Hand-Held Tablet Device—A Study on Selection and Pose Manipulation Techniques

Ali Samini  and Karljohan Lundin Palmerius * 

Department of Science and Technology, Linköping University, 58183 Norrköping, Sweden; ali.asamini@gmail.com

* Correspondence: karljohan.lundin.palmerius@liu.se

Received: 28 February 2019; Accepted: 18 April 2019; Published: 24 April 2019



Abstract: Current hand-held smart devices are supplied with powerful processors, high resolution screens, and sharp cameras that make them suitable for Augmented Reality (AR) applications. Such applications commonly use interaction techniques adapted for touch, such as touch selection and multi-touch pose manipulation, mapping 2D gestures to 3D action. To enable direct 3D interaction for hand-held AR, an alternative is to use the changes of the device pose for 6 degrees-of-freedom interaction. In this article we explore selection and pose manipulation techniques that aim to minimize the amount of touch. For this, we explore and study the characteristics of both non-touch selection and non-touch pose manipulation techniques. We present two studies that, on the one hand, compare selection techniques with the common touch selection and, on the other, investigate the effect of user gaze control on the non-touch pose manipulation techniques.

Keywords: augmented reality; video see-through; user-perspective; device motion; pose manipulation; selection

1. Introduction

Smart devices such as cellphones and tablets are highly suitable for video see-through Augmented Reality (AR), as they are equipped with high resolution screens, good cameras, and powerful processing units. An enabling factor for the application of such devices is the possibility to provide intuitive and efficient interaction. In this article we consider two aspects of interaction, selection [1,2], and 3D manipulation of virtual objects [3,4].

In head-mounted AR, based on devices such as the HoloLens, wands or mid-air gestures can be used to provide direct 3D pose manipulation [5,6] by directly mapping 3D gesture to pose change (Figure 1, left). This is not common for hand-held AR, as the use of wands or gestures requires the user to hold the device with one hand and at the same time limits interaction to the use of only one hand or wand (Figure 1, center). Touch-based on-screen interaction techniques are commonly used in hand-held video see-through Augmented Reality (AR), where a single touch is used for selection and multi-touch gestures are used for manipulation [7,8], see Figure 1, right. Multi-touch techniques provide robust interaction, but they lack the direct 3D connection between gesture and pose. This issue affects primarily selecting in depth and manipulating the third dimension, which is not available on a 2D screen and requires the introduction of 2D metaphors that are mapped to 3D pose manipulation. Touch-based interaction is also affected by finger occlusion and one-handed operation that can potentially increase fatigue.

To enable direct 3D interaction, including on hand-held AR, an alternative is to use the device as a wand [9,10]. This technique enables a direct 3D mapping between the user actions and subsequent

change in object pose. This also enables continuous control while holding the device with both hands, which eliminates finger occlusion and potentially also decreases fatigue.

In this article we explore, study, and refine techniques for selecting and manipulating objects in hand-held video see-through AR, to approach a more wand-like interaction experience.

For selection, we explore the existing wand like selection techniques, and suggest a small modification to an existing device position-based selection technique. We also introduce an alternate technique based on the device's 3D position. Furthermore, we compare the performance of the modified and introduced selection techniques in a user study. For manipulation we further develop a recent device manipulation technique [11], to keep the object consistently visible throughout transformation changes, and explore its characteristics in a user study, together with the original technique.

The main contributions are

- a description and analysis of touch and non-touch selection and manipulation in hand-held video see-through AR,
- a new approach to selection with minimal touch and hand movements,
- a new use of user perspective rendering together with device pose-based manipulation for fast, precise, and continuous pose control, and
- two user studies comparing both new and existing techniques for selection and pose manipulation in hand-held video see-through AR.



Figure 1. Three examples of Augmented Reality (AR) interaction: head mounted display and wand, hand-held and wand, and hand-held with touch.

2. Related Work

The most important aspects of interaction on hand-held devices can be divided into selection and pose manipulation of objects. This includes both touch-based techniques and alternatives. Recent applications of hand-held AR apply user perspective rendering (UPR) to improve the mapping between what is seen on the screen and around it which also has implications for selection and manipulation.

2.1. User Perspective Rendering

The most straightforward way to setup the view in hand-held video see-through AR is to use the frustum of the device's back-facing camera to render both the camera view and augmentations on the screen [12–15], device perspective rendering (DPR). In contrast, user perspective rendering (UPR) is achieved by taking both user's eye position and screen pose into consideration to create a perspective for the user that makes the graphics appear like a real object would if replacing the screen with an empty frame. This reduces the registration error between what is seen on the screen and around the screen. This has been shown to work based on homography [16–18] and with projective geometry [17,19,20]. The UPR view does not rotate with the rotation of the hand-held device, so rotating the device can instead be used for manipulation, without affecting the view.

2.2. Pose Manipulation

There are three classes of pose manipulation techniques for hand-held AR: on-screen multi-touch gestures, mid-air wand or gesture action, and device-based interaction. For an overview of techniques and their characteristics, we refer to a comprehensive review [21] by Goh et al.

Pose manipulation in hand-held AR, with touch-enabled devices, is commonly done using multi-touch on-screen gestures, see for example [7,8,22,23]. While the multi-touch interaction is a useful technique, it has some issues. Performing 2D gestures for 3D pose manipulation requires metaphors for controlling the third dimension, thereby removing the intuitiveness and direct one-to-one mapping. Also, on-screen finger occlusion decreases the visual feedback, forcing the user to hold the device with one hand while performing the gestures with the other hand can increase fatigue, especially on larger and thus heavier devices such as tablets.

A more direct object manipulation can be achieved by using mid-air gestures. Hürst and van Wezel [24] presented the results from two experiments, that indicate that the finger tracking for gesture-based interaction has “high entertainment value, but low accuracy”. Also removing finger occlusion, it requires the user to hold the device with one hand during operation.

As an alternative to wand or mid-air gestures the hand-held device itself can be used as a wand, where movements of the device are directly mapped to the manipulated object, which we call *device pose-based manipulation* (DM). To our knowledge the first example in the literature was presented by Henrysson et al. where a mobile phone, tracked through its camera with fiducial markers on a table, was used to manipulate objects [9]. Sasakura et al. created a molecular visualization system which is controlled by a mobile device for 6 DoF pose manipulation of the molecules [25]. Tanikawa et al. presented a DM technique that they call ‘Integrated View-input AR’, which they tested in a JENGA-like game [26].

Common to the DM techniques referenced above is that they lock the object to the hand-held device (Figure 2(1a)), and therefore the object is moved with the device as if they are connected via a virtual rod (Figure 2(1b)). The center of rotation is co-located with the device’s center, and thus it is difficult to rotate the object without moving it (Figure 2(1c)). This problem increases with the distance between the object and the device. Mossel et al. presented an alternative DM technique in their HOMER-S application [10]. Their method directly maps the device pose to the object and therefore translates and rotates it with the same amount as the device and in the same direction (Figure 2(2)). This solves the problem that the rotation affects translation, however this approach introduces two other issues. First, since they use DPR, rotating the device rotates the view, and therefore the object will not be visible on the screen for yaw and pitch rotations (Figure 2(2c)). Also, since there is a one-to-one translation mapping, the amount of movement of the object will in perspective look smaller, which results in the object potentially moving out of view because of the large movements during translation (Figure 2(2b)). A hybrid technique was suggested by Marzo et al. [27], which translates the object with the same DM technique as HOMER-S and rotates it using the multi-touch. This solves the rotation issue of HOMER-S DM but inherits the multi-touch issues which we discussed above.

In an effort to solve the rotation issues, Samini and Palmerius presented a DM technique they called “Relative-UPR” [11], which applies UPR to resolve the rotation issues with the DPR-based DM techniques. With their method, the object is rotated around its own center and, therefore, is kept in the view (Figure 3c). A comparison was made in a user study, between the DPR-based DM techniques and the introduced UPR-based method, called Relative-UPR, showing mixed results for the two approaches depending on situation. Relative-UPR was shown to remove the issues with the DPR-based DM and continuously keep the object in the view during rotations, however it is unable to do the same when translation is involved in the movement (Figure 3b). As an effort to try to mitigate this problems we suggest a modification that locks the object to the user’s view. We call this *Fix-UPR*.

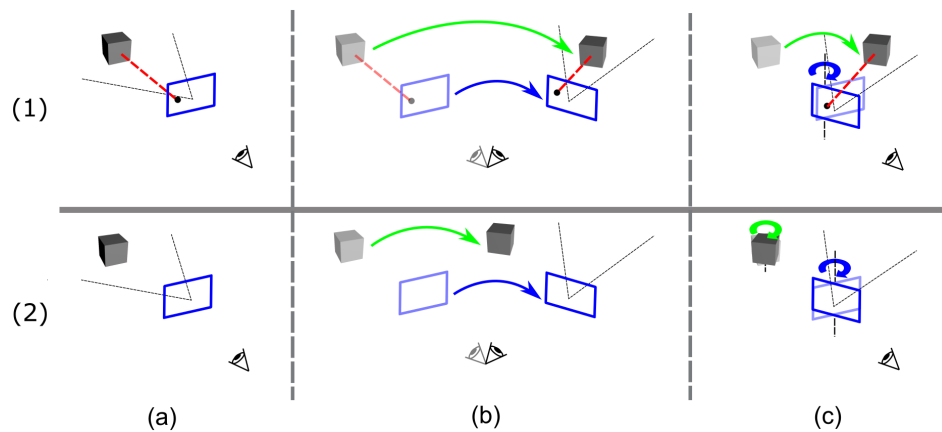


Figure 2. Device pose-based manipulation techniques, based on device perspective rendering, are used to move/rotate a virtual cube. (1) Fixed technique: (a) virtual cube is locked to the device when it is selected, (b) moving the device moves the virtual cube as they are connected with a virtual rod and thus the virtual cube stays in the AR view and visible on the screen, (c) rotating the device rotates and also translates the virtual cube making it difficult to only rotate the virtual cube around its center. (2) Relative technique: (a) virtual cube is selected, (b) the virtual cube is moved with the same amount as the device but in perspective and therefore, it gets out of the AR view and thus not visible on the screen after the movement, (c) rotating the device rotates the virtual cube by same amount around its own center. It is possible to rotate the virtual cube without translating it, however rotating the device rotates the AR view and thus the virtual cube gets out of AR view and will not be visible on the screen.

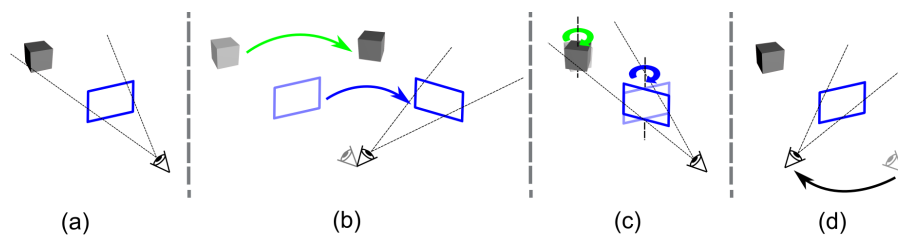


Figure 3. Relative-user perspective rendering (UPR) device pose-based manipulation technique: (a) the virtual object is selected, (b) the virtual object is moved with the same amount as the device but in perspective and therefore gets out of the AR view, (c) virtual object is rotated by rotating the device, which does not rotate the AR view, and therefore stays in the AR view, (d) changes in the pose of user's view changes the AR view which consequently moves the virtual object out of the user's view.

2.3. Selection

When multi-touch pose manipulation is used, it is natural to also use touch for selection, however when using manipulation techniques that do not require on-screen touch, using fingers to select objects reintroduces the occlusion and fatigue issues. There are non-touch alternatives which can be coupled with non-touch manipulation techniques such as DM. Baldauf and Fröhlich presented two alternatives they call “Target” and “Snap Target” [28]. In their application the selection is done on hand-held devices, to activate controls on a larger, remote screen. The Target selection technique uses a centered cross-hair, while the Snap Target applies a dynamic cross-hair that snaps to a control when it is near to the center of the screen. They compared both techniques along with touch selection in a user study, but could not find any significant difference between the selection techniques for task completion time. The number of selection errors, however, were significantly different where touch attained more errors than Target and Snap. Their Target and Snap Target techniques reduce the amount of touch interaction and is suitable to be used with DM manipulation. However, their application is based on 2D selection of controls placed in a plane and therefore does not take into account the issues which are raised from selection in a 3D environment, such as object occlusion.

Tanikawa et al. proposed and evaluated three types of selection they call “Rod”, “Center”, and “Touch”, for their DM technique described above (Integrated View-input AR, [26]). The Rod technique selects an object located at the tip of a virtual rod shown on the screen. The Center technique uses an “aiming point” that is shown at the center of the screen (identical to Target described above), and the Touch is a regular touch selection. They compared these selection techniques in a JENGA-like game. Their results did not show any significant difference of task completion time between the selection techniques. However, overall more users preferred Rod and Center techniques over Touch selection. Both the Rod and Center selection techniques and their subsequent manipulation rely on device movements and thus fit well together.

A potential issue with these techniques occur in applications where multiple objects are selected in sequence: the fair amount of device movement required with the center-based techniques can result in fatigue. Because of this we suggest a minor modification of the center select techniques, and introduce and test an alternative selection approach specifically designed to minimize device movements, thereby also testing the influence of device movements on performance.

3. Modified Interaction Techniques

For the user study, we modified the center-based techniques and introduce a new technique for selection and one for pose manipulation. We call these Center-select, Icons-select, and Fix-UPR, respectively.

3.1. Center Selection

A basic implementation of a center-based selection technique would pick an object when it is at the center of the screen. The snap-target modification increases the reach and reduces the precision necessary when selecting small objects. To further reduce the amount of device movement, an alternative modification is here made that selects the object closest to the screen center anywhere on the screen (Figure 4(left)). In our implementation the actual selection is done by touching at any part of the screen, which is possible to performed with either thumb while holding the device with both hands.

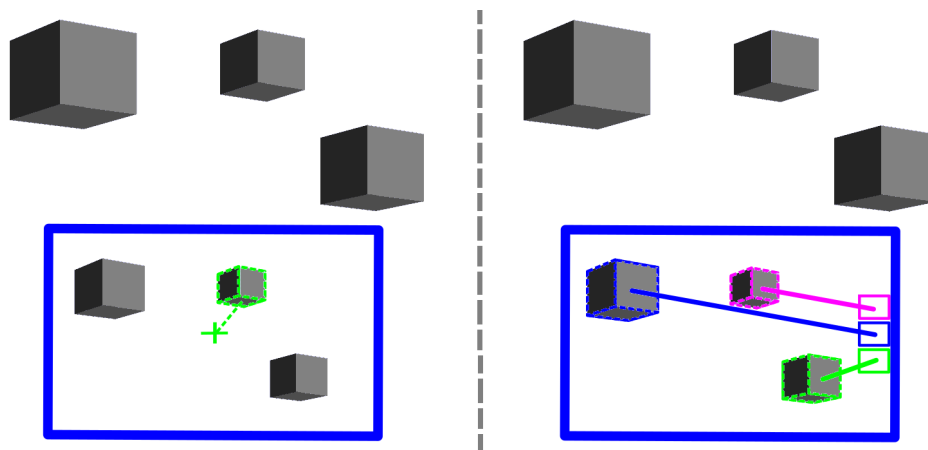


Figure 4. Selecting an object in virtual space. **Left**, Center-select: the nearest object to the center of the screen can be selected. **Right**, Icons-select: objects are assigned to icons and are selected by their corresponding icons.

3.2. Icons-Based Selection

To reduce the required device movement when selecting a number of objects in sequence, we introduce the Icons-select method. The technique assigns a number of objects to selectable icons. This provides the user with a list of objects to choose from, effectively moving touch selection to the edge of the screen. We place the icons at either side of the screen, within the reach of the thumb of the user’s dominant hand (Figure 4(right)), so that selection can be made while holding the device with both

hands. To indicate the icon–object relationships each icon is connected to its assigned object by a line. To further indicate the icon–object association each object is highlighted with the same color that is used for its icon and connection line.

It can be argued that a user of the touch-select technique is free to move the device so that the selectable objects end up at the edge of the screen, close to their thumbs while holding the device, so that they can select objects without releasing either hand. Our experience from pilot studies, however, is that users tend to work at the center of the screen.

If there are more objects than the number of icons we suggest choosing the objects closest to the screen center, based on their on-screen projections. The order of icon–object assignments is also important: as illustrated in Figure 5a, the crossed connection lines can reduce the perception of the assignments. A straightforward assignment method finds the most suitable object as illustrated in Figure 5b–d: starting from the top icon, the object with the smallest angle between its icon–object line and screen’s up direction is assigned to each icon. Here, moving the device changes the relative position of the screen and objects, so the icon assignments must be dynamically updated. Hysteresis can be used to make these assignments more stable. A suitable threshold value for our studies was determined in a pilot study.

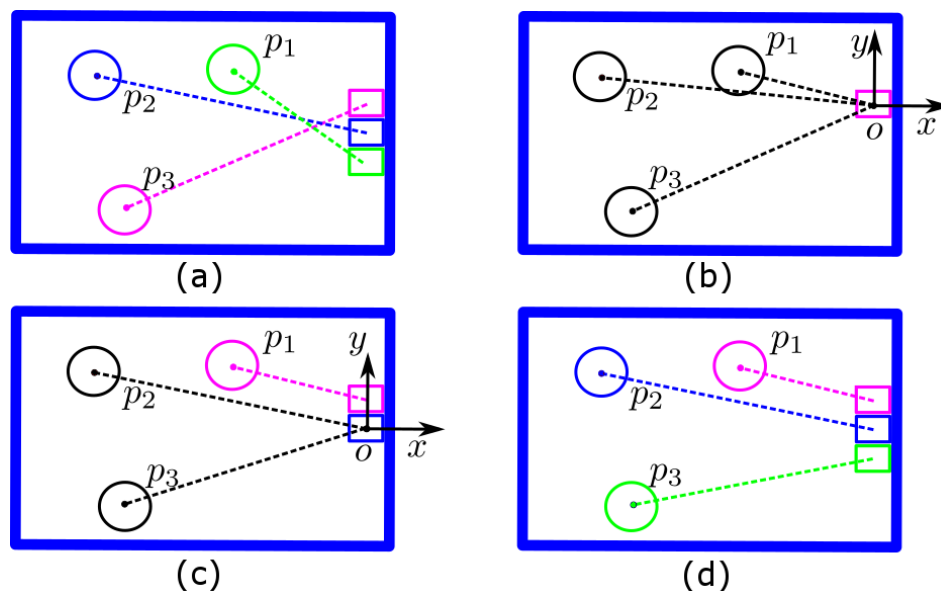


Figure 5. Icon–object assignment in Icons-select technique. p_1 , p_2 , and p_3 show the projection centers of three objects, on the screen. (a) An example of icon–object crossed lines without the assignment technique; (b–d) avoiding cross lines by assigning the icons, one by one from top to bottom, to the unsigned object with the smallest angle $\angle poy$.

Using more icons gives a larger object pool to select from, but at the same time makes it more difficult to see which icon is assigned to a certain object. We performed a pilot study to determine the convenient number of the used icons and the subjects suggested to use three icons as they believed more makes it difficult to recognize the icon–object assignments and fewer reduces the very purpose of the technique, which is to provide multiple options.

3.3. Fixed-View Manipulation

The previously presented UPR based DM technique, Relative-UPR [11], is able to keep the object within the user’s view during rotations, however it is unable to do the same with translations (Figure 3b). To resolve this issue we suggest a new UPR-based DM technique, Fix-UPR. This technique locks the object’s 3D position to its projected position on the screen, as defined by the user’s current UPR-based view, see Figure 6a. Regardless of how the view changes, because of device or head

movements, the object will always be rendered in the same location on the screen (Figure 6b,d). The rotation, on the other hand, is controlled by device orientation: as the device is rotated around the intersection point, the object is correspondingly rotated around its center without translation (Figure 6c). This way Fix-UPR is able to keep the object consistently in the view while manipulating both position and orientation.

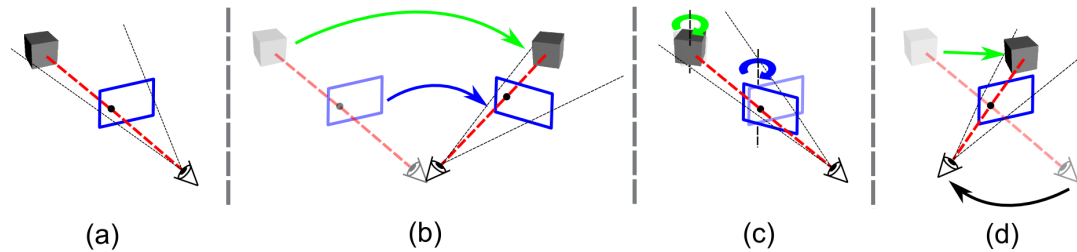


Figure 6. Fix-UPR device pose-based manipulation technique: (a) the virtual object is locked to its projection point on the screen. As defined by the user’s view, (b) moving the device the virtual object will always be visible at the projection point on the screen, (c) rotating the device with the center of rotation at the projection point will rotate the virtual object without around its own center. (d) Moving the user’s view will move the virtual object, and thus keeps it consistently visible on the screen at the projection point.

Figure 7 illustrates how an object is transformed by Fix-UPR (from one pose, o_1 , to another pose, o_2). In this technique the orientation and position are controlled in different ways, and are therefore calculated separately. The orientation calculations can be done in, for example, 3×3 matrix representation of rotation. When the object is first grabbed, its orientation relative to the screen, R_{Δ} , is first calculated as

$$R_{\Delta} = R_{s_1}^{-1}R_{o_1} \tag{1}$$

where $R_{s_1}^{-1}$ is the inverse of the screen orientation and R_{o_1} is the orientation of the object. After this, when the screen is rotated, the new object orientation, R_{o_2} , is calculated as

$$R_{o_2} = R_{s_2}R_{\Delta} \tag{2}$$

where R_{s_2} is the new orientation of the screen after the movement. The object is also moved relative to screen pose changes, which is done through its projection point on the screen. When the object is first grabbed, three things are calculated. First, the objects projection on the screen, p_1 , is calculated as the intersection between the plane of the screen and the line from the user’s eye and the object. Also, the distance between the user’s eye and the screen point, $a_1 = |p_1 - v_1|$, is needed, as well as the user’s eye distance to the object, $b_1 = |x_{o_1} - v_1|$. After this, when the screen is moved, a new projection point, p_2 , can be estimated based on how the screen as moved as

$$p_2 = T_2T_1^{-1}p_1 \tag{3}$$

where T_1 and T_2 are 4×4 column transformation matrices in homogeneous coordinates for the initial and new screen poses, respectively. Since the screen can be moved closer or further away from the user’s eye, this depth motion can also be used to intuitively move the object in depth direction. Thus, the new position for the object, x_{o_2} , is calculated as

$$x_{o_2} = v_2 + \frac{b_1}{a_1}(p_2 - v_2) \tag{4}$$

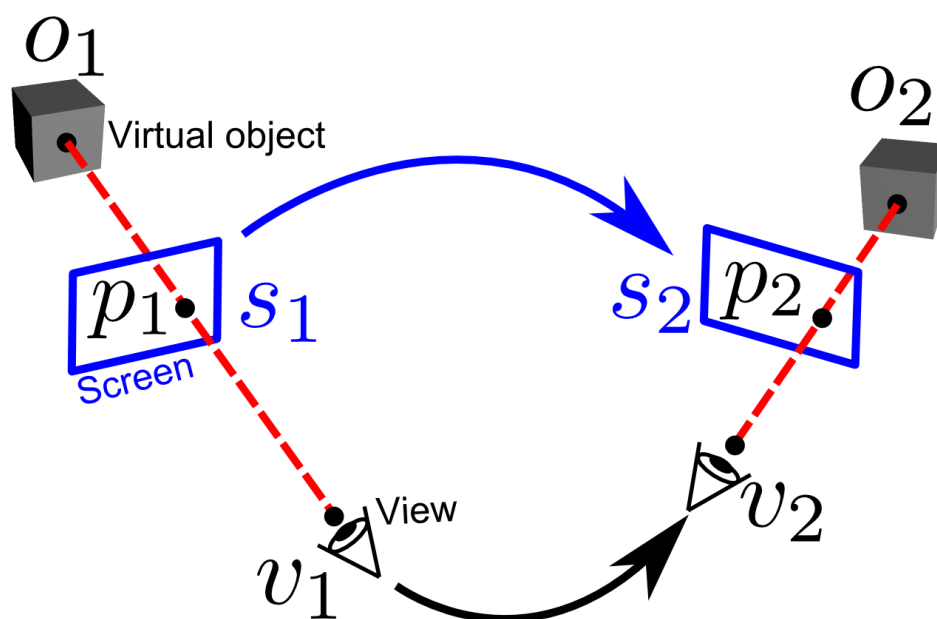


Figure 7. An object is transformed by Fix-UPR technique from o_1 to o_2 . The s_1 and v_1 are the the screen and the user's view before the movement, respectively, and the s_2 and v_2 are the same variables after the movement.

4. User Studies

Two user studies have been executed to determine which DM-based interaction technique has better overall performance. The performance is assessed in terms of speed, fatigue, and accuracy in both studies. The first study focuses on object selection and compare three selection techniques, of which two are DM-based and one is based on touch selection (Center-select, Icons-select, and Touch-select). The focus of the second study is on object manipulation and compare two UPR-based DM techniques (Fix-UPR and Relative-UPR).

4.1. Test Environment and Prototype

Both experiments were performed indoors in an area of 2×2 m. For tracking of the tablet and user's eyes we used external optical tracking. This decision is made due to lack of stability and accuracy of the existing internal tracking solutions such as using the device's camera or Global Positioning System (GPS). We can assume that in the future we can get the same amount of accuracy from the AR internal tracking. The test area was covered by a PhaseSpace optical motion tracking system that uses illuminative LED markers for real-time 6-DoF tracking. The video see-through AR application ran on an off-the-shelf Microsoft Surface-Pro 3 tablet PC. The back-side camera of the device was used for realtime video AR. The LED markers and their driver unit were mounted on a 3D printed cover that was attached to the tablet. The head tracking was provided via a pair of safety goggles, also equipped with LED markers, see Figure 8.

The communication between the tracking system and the tablet PC was provided by VRPN [29] over WiFi. The software was developed using C++, OpenGL, and SDL2 [30]. The camera calibration, image capture, and management were done with OpenCV [31].



Figure 8. The apparatus: a tablet PC covered with a 3D printed frame. The frame is mounted with the LED illuminative markers and their unit driver with batteries. A pair of safety goggles equipped with LED markers is connected to the same driver unit.

4.2. Participants and General Procedure

Both studies were performed in sequence by the same group of 18 test subjects, who were recruited from the university (5 female and 13 male) with age ranging from 22 to 36 (median 30). The order of the studied techniques were completely balanced within each study between the 18 subjects, to minimize the impact of learning effects and fatigue. However, all subjects performed the selection study first and then the manipulation study, to avoid experience of selection in the manipulation study to affect the selection study. The subjects were given time to rest between two studies to minimize the effect of fatigue in the second study.

Upon arrival each subject was asked to fill in a form with their personal information and possible disabilities that were expected to affect the test. We did not find any physical disability or uncorrected vision problem. The test subjects were then given a pre-test questionnaire to input their level of experience with hand-held devices and AR. All subjects were mostly experienced with hand-held devices, some of which reported to be familiar with AR and a few with the UPR.

After completing the questionnaire, a tutor thoroughly described UPR, DM, and the other techniques used in the study. The subjects were trained separately for each experiment before starting the test. The practice finished upon the subjects' request, when they felt well-prepared for the actual experiment. Each subject first performed the selection study followed by a questionnaire, and then the manipulation study also followed by a questionnaire.

Both studies were started by a touch on the screen, with the subjects standing in the middle of the test area, holding the tablet and wearing the tracking goggles. The test was automated by the application and data were recorded from the start. The study took about 45 min for each subject, including pauses.

4.3. Selection Study

The aim of the selection study was to compare the user's performance with the different DM and touch selection methods, Center-select, Icons-select, and Touch-select, in terms of speed, fatigue, and accuracy. Therefore, no manipulation is included in this study. The Center-select is DM-based, Icons-select is list-based, and the Touch-select technique was added for comparison, since it is a common technique in the literature.

4.3.1. Expected Outcome and Hypotheses

Our expectations from the analysis of the different techniques during pilot tests include that Touch-select is more difficult to use, since the subjects need to hold the device with one hand. Also, Center-select and Icons-select require less movements of the hand and fingers over the screen than Touch-select, and is therefore faster. Icons-select requires less movement of the device between multiple selections, since the objects can be chosen from a list. The accuracy, on the other hand, may be lessened by the confusion produced from the object-icon connections that may switch as the device is moved. The level of accuracy is expected to be higher with Touch-select due to the direct connection between touch action and selected object. These observations are formulated into the following hypotheses:

H1: Subjects will perform the selection task faster with Icons-select and Center-select compared to Touch-select.

H2: Subjects will move the device less with Icons-select compared to Touch-select and Center-select.

H3: Subjects will perform less wrong selections with Touch-select and Center-select compared to Icons-select.

4.3.2. Design and Procedure

A three dimensional array of virtual cubes are rendered through the AR display. The task for the subjects was to select, one-by-one, the randomly highlighted one (Figure 9). The performance metric was recorded as the time of completing the task (representing speed), the length of the trajectory that the device moved (representing fatigue), and the number of erroneous selections (representing accuracy). These were considered the dependent variables and the selection techniques were the independent variables.

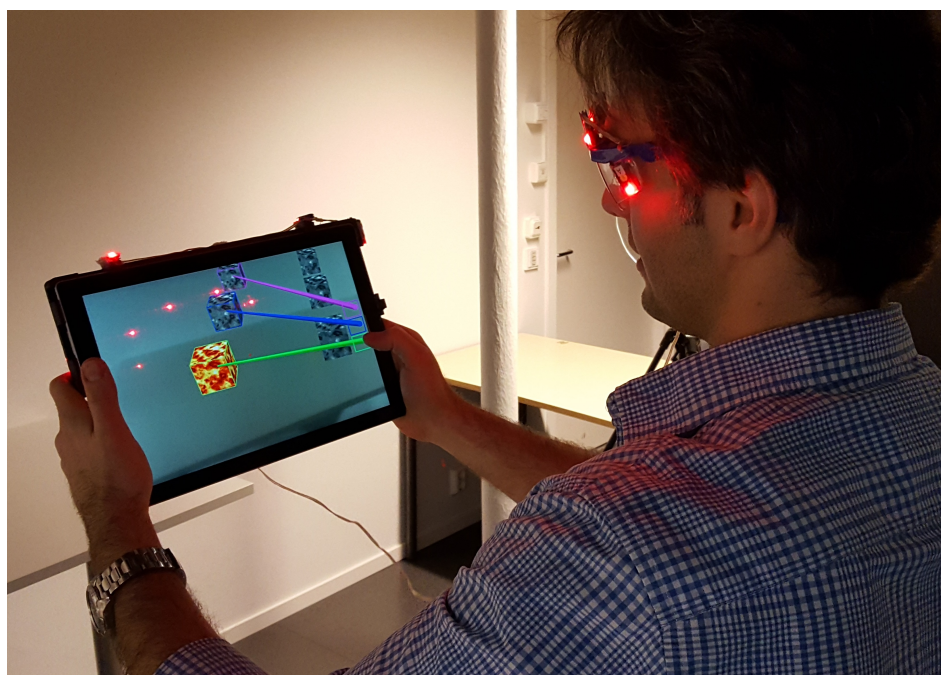


Figure 9. The selection test, here performed with Icons-select technique.

The test was done with 27, $5 \times 5 \times 5$ cm virtual cubes that were placed in $3 \times 3 \times 3$ cube formation with 40 cm distance. This distribution of objects, also in depth, introduces occlusion between objects which makes the selection task more difficult, requires the user to sometimes move sideways, increases risk of erroneous selection, increases movement, and also makes the scenario more realistic. The cubes were placed in front of the subjects at the beginning of the test, however during the test the subjects were allowed to freely move around in the test area, even circulating the array of objects. The cubes were gray and randomly turned yellow, one-by-one, to be selected. When correctly selected the cube turned gray and the next was randomly highlighted.

The subjects performed the three tests, Touch-select, Center-select, and Icons-select, in different a order that was perfectly balanced within the group. Thirty selections were performed using each technique, thus $30 \times 3 = 90$ selections in total. Three randomized sets, one for each technique, determined the order of highlighting. These sets were also balanced with respect to the overall positions and distances between the cubes to select.

The selection test took about 15 min. The subjects were asked to express their preferences in a post-experiment questionnaire after completing the task.

4.3.3. Results

Figure 10 and Table 1 present the results from the selection test. The recorded data were analyzed with repeated measures ANOVA, with statistical significance at $p < 0.05$, separately for the task completion time, the amount of device movement and the number of the wrong selections. The Kolmogorov–Smirnov test verified that the data were normally distributed, and the Mauchly’s test indicated that the sphericity assumption was not violated.

The task completion time differed significantly between the selection techniques, Wilk’s Lambda = 0.481, $F(2,16) = 8.615$, $p = 0.003$. Post hoc test with Bonferroni correction revealed that the mean task completion times for Touch-select and Center-select were significantly faster than for Icons-select, $p = 0.031$ and $p = 0.002$, respectively. We did not find any significant difference between Touch-select and Center-select ($p \approx 1.0$).

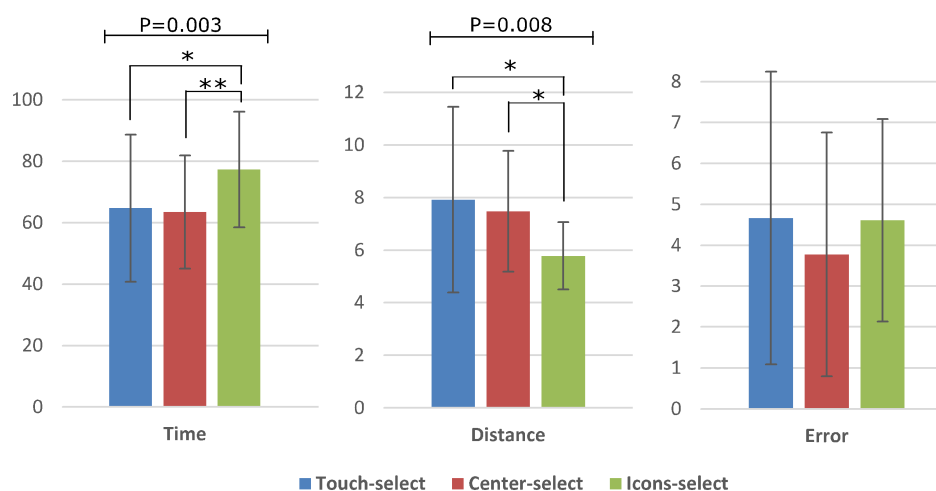


Figure 10. Results of the selection test. The candles show mean task completion time in seconds, mean distance that the device is moved in meters, and the number of wrong selections for three compared selection techniques. The standard deviation and significance are also indicated.

Table 1. Descriptive statistics of the selection test. The mean time is the task completion time for all (N) users in seconds. The mean distance is the amount of device movement to complete the task, in meters. The mean error is the number of erroneous selection during the test.

Measure	Technique	Mean	Std. Dev.	N
Time	Touch-select	64.762	23.951	18
	Center-select	63.461	18.396	18
	Icons-select	77.331	18.787	18
Distance	Touch-select	7.916	3.534	18
	Center-select	7.486	2.294	18
	Icons-select	5.783	1.288	18
Error	Touch-select	4.67	3.581	18
	Center-select	3.78	2.981	18
	Icons-select	4.61	2.477	18

There was a significant effect, of the selection technique, on the amount of device movement to complete the task, Wilk's Lambda = 0.548, $F(2,16) = 6.6$, $p = 0.008$. Post hoc test with Bonferroni correction revealed that the distance was significantly shorter with Icons-select compared to both Touch-select ($p = 0.032$) and Center-select ($p = 0.011$). The Touch-select and Center-select did not differ significantly ($p \approx 1.0$).

We could not find any significant effect of the selection techniques on the number of wrong selections, Wilk's Lambda = 0.916, $F(2,16) = 0.0729$, $p = 0.498$.

Thus, the results do not support H1, which stated Icons-select will be faster than the other techniques. They do, however, support H2, that predicted that the required device movement for Icons-select would be less than with the other techniques. Also, the results do not support H3, stating that the number of wrong selections would be less using Touch-select and Center-select compared to Icons-select.

4.3.4. Post-Test Questionnaire

The post-test questionnaire was designed to capture the preferences of the subjects. The questions and answers are listed in Figure 11. The subjects were first asked how easy each technique was to learn, where 61% of them expressed Icons-select, 88% Center-select, and 95% Touch-select to be somewhat or very easy to learn. The subjects then were asked how easy each technique was easy to use (physically). A proportion of 50% chose Icons-select, 94% Center-select, and 61% Touch-select to be somewhat or very easy to use. When asked about method fatigue, 45% of the subjects chose Icons-select, 23% Center-select, and 67% Touch-select to be somewhat or very tiring. Then the subjects were asked how natural and intuitive each technique felt. A proportion of 50% found Icons-select, 89% Center-select, and 72% Touch-select to be somewhat or very natural and intuitive. Finally, the subjects revealed their overall preferred selection technique, where 61% of the subjects chose Center-select, 28% Touch-select, and the remaining Icons-select (11%).

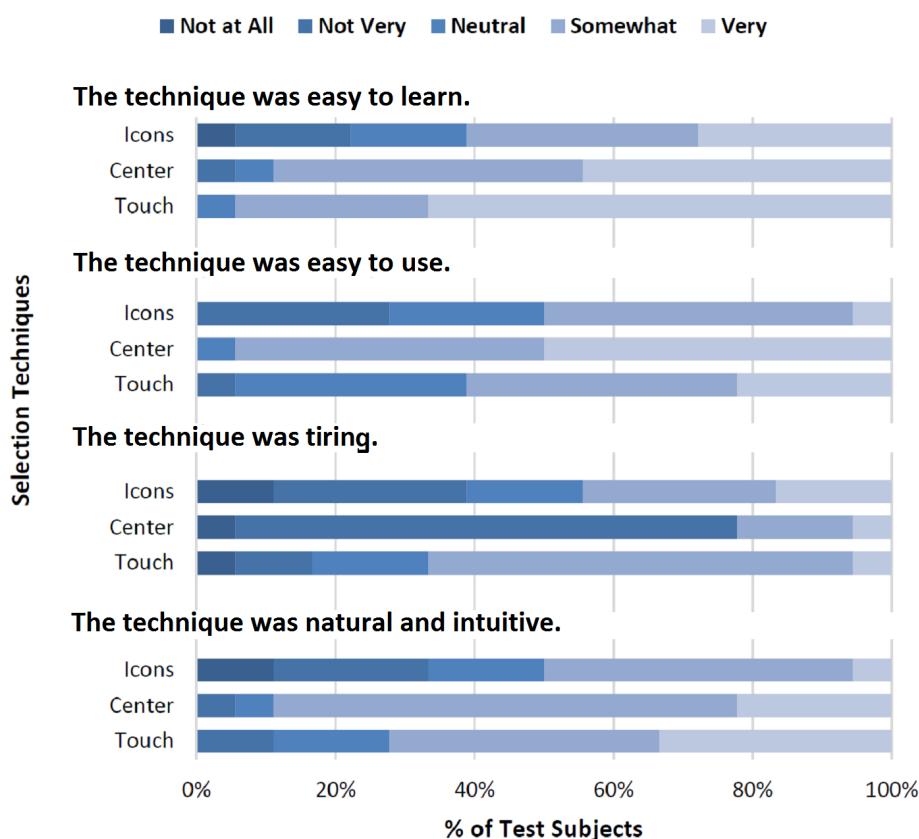


Figure 11. The questions and answers from the post selection test questionnaire.

4.4. Manipulation Study

The manipulation experiment was designed to evaluate and compare the performance when using the two UPR DM techniques. Relative-UPR, which resolves the issues of DPR based DM techniques as is discussed in [11], is compared with the suggested Fix-UPR technique, which aims to resolve the translation issue of Relative-UPR and, thus, keep the object continuously in the view.

4.4.1. Expected Outcome and Hypotheses

Based on our pilot trials and technical analysis on both UPR based DM techniques, we believe that Fix-UPR resolves the transform issue with the relate-UPR. Fixating the object to the view consistently keeps the object visible on the screen, unlike Relative-UPR. Therefore, we believe, users of Relative-UPR will have to release the object, perform manual perspective corrections, to get it back into view, and then pick it up again to continue the manipulation. Thus, Fix-UPR will lead to faster manipulation and require less movement of the device. These observations are formulated into hypotheses as follows:

H4: Subjects will perform the manipulation task faster with Fix-UPR compared to Relative-UPR.

H5: Subjects will move the device less with Fix-UPR compared Relative-UPR.

H6: Subjects will release and grab objects more with Fix-UPR compared Relative-UPR.

4.4.2. Design and Procedure

The two UPR-based DM techniques (Relative-UPR and Fix-UPR) were compared using a virtual puzzle-solving task. The task was to complete a nine piece puzzle, by placing the pieces at their correct position, with their correct orientation. The independent variable was the interaction techniques and the dependent variables were performance metrics in the form of task completion time (representing speed), the length of the trajectory that the device was moved (representing fatigue), and the

perspective correction count, represented by the number of re-grabs of each puzzle piece before it is docked.

A 60 × 30 cm puzzle board was placed 50 cm away from the user's starting point, facing the subject. The nine puzzle pieces (20 × 10 cm, in color) had to be moved from a random position onto a completed version of the puzzle (in gray scale), see Figure 12. The back of the pieces had no indication of identity or orientation. To prevent puzzle solving skills of the subjects effect the results, the puzzle was chosen to be very simple. Finding the correct orientation was also facilitated by the rectangular design of the pieces.

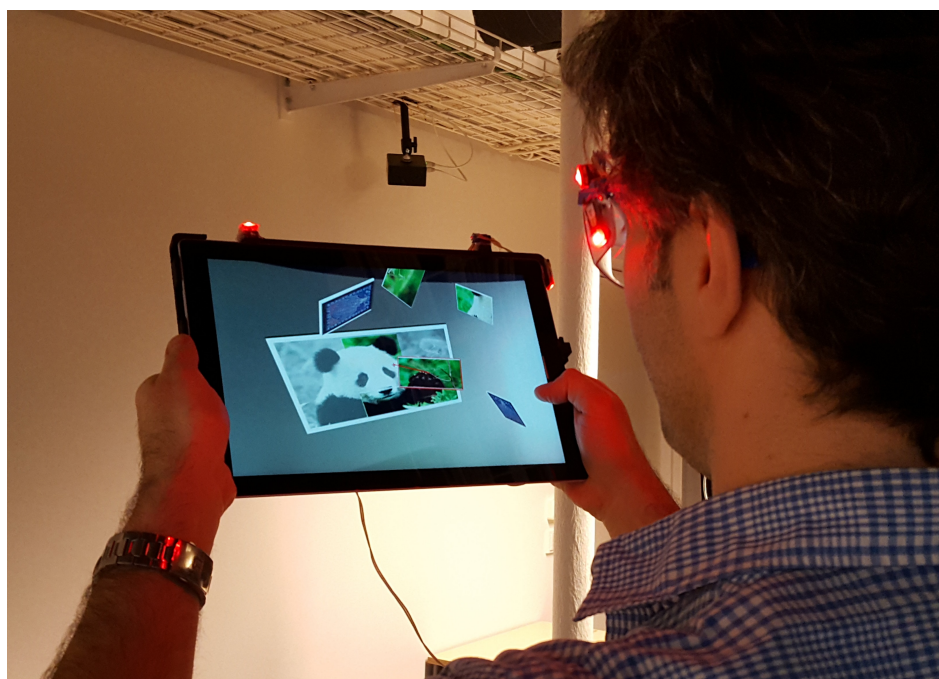


Figure 12. The manipulation experiment, where the subject is solving a virtual puzzle.

The subjects solved the puzzle one time with Relative-UPR and one time with Fix-UPR, with the order balanced between the subjects. Thus, each subject had to dock nine puzzle pieces to solve the puzzle. The randomized sets of puzzle piece poses were balanced to have overall equal transformations needed to solve the puzzle. Based on a pilot study, Center-select was the preferred selection technique, and thus used for picking the pieces.

The subjects started by standing in front of the puzzle board. During the test they were allowed to move around the pieces and the board. When a piece was positioned at the correct position on the board, with the correct orientation parallel to the board, it snapped to its place and became un-selectable for the remaining duration of the test. Suitable amounts of acceptable errors for rotation and translation were determined through the pilot study.

The test took about 20 min for each subject, followed by filling in a post-test questionnaire where the subjects were asked to input their overall preferences for the techniques.

4.4.3. Results

Figure 13 and Table 2 present the results from the manipulation test. Separate repeated measures ANOVAs, with statistical significance at $p < 0.05$, were done on the task completion time in seconds, the device movement in meters, and the number of object re-grabs. The normal distribution of the data has been verified by the Kolmogorov–Smirnov test and the sphericity assumption has been confirmed for all ANOVAs.

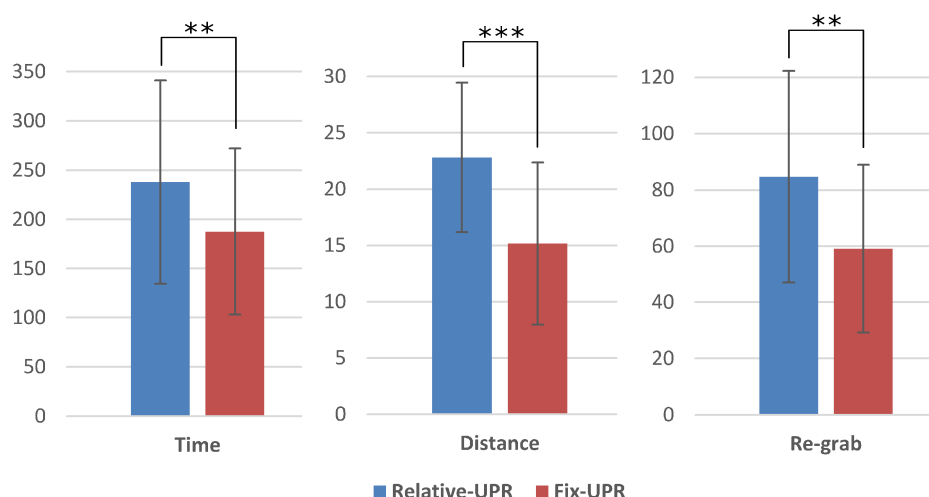


Figure 13. Results of the pose manipulation test. The candles show mean task completion time in seconds, the mean distance that the device is moved in meters, and the number of object reselects for both compared techniques. The standard deviation and significance are also indicated.

Table 2. Descriptive statistics of the manipulation test. The mean time is the task completion time for all (N) users in seconds. The mean distance is the amount of the device movement to complete the task, in meters. The re-grab measure is the number of object re-selects to complete the task.

Measure	Technique	Mean	Std. Dev.	N
Time	Relative-UPR	237.803	103.288	18
	Fix-UPR	187.578	84.557	18
Distance	Relative-UPR	22.809	6.638	18
	Fix-UPR	15.17	7.203	18
Re-grab	Relative-UPR	84.67	37.674	18
	Fix-UPR	59.11	29.846	18

The analysis indicates that the manipulation technique significantly affect the task completion time, Wilk's Lambda = 0.645, $F(1,17) = 9.369$, $p = 0.007$, in favour of Fix-UPR. There was a significant effect of the manipulation technique on the length of the device movement, Wilk's Lambda = 0.243, $F(1,17) = 53.078$, $p = 10^{-6}$, in favour of Fix-UPR. The number of re-grabs for each puzzle piece was significantly different between the techniques, Wilk's Lambda = 0.597, $F(1,17) = 11.480$, $p = 0.003$, in favour of Fix-UPR.

The results support all three hypotheses: the subjects performed the experiment faster with Fix-UPR than the Relative-UPR, with Fix-UPR the subjects had to move the device less compared to Relative-UPR, and the number of re-grabs performed by the subjects to complete the task was less with Fix-UPR compared to Relative-UPR.

4.4.4. Post-Test Questionnaire

The questions and answers from the post manipulation experiment questionnaire are presented in Figure 14. A total of 94% of the subjects found Fix-UPR and 83% Relative-UPR to be somewhat or very easy to learn. When asked about ease of use 94% found Fix-UPR and 83% Relative-UPR to be somewhat or very easy to use (physically). A total of 33% of the subjects expressed Fix-UPR and 61% Relative-UPR to be somewhat or very tiring to use. Furthermore, 89% of the subjects found Fix-UPR and 50% Relative-UPR to be somewhat or very natural and intuitive. Finally, the subjects were asked

which technique they preferred as manipulation technique, where 83% chose Fix-UPR and the rest Relative-UPR (17%).

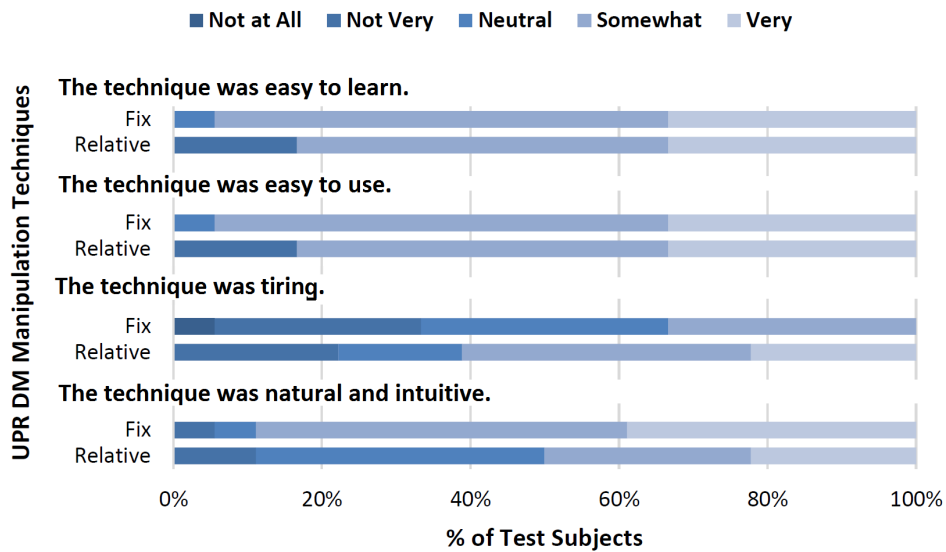


Figure 14. The results from the post manipulation test questionnaire.

5. Discussion

In Sections 2 and 3 above, selection and manipulation techniques are analyzed and discussed based on their principles and inherent behavior. These principles were put to the test in the study presented in the previous section. Some of the results from these studies can be considered to follow the intrinsic behavior of the compared techniques. There are, however, a few aspects in the outcome should be highlighted and some that need closer examination.

5.1. Selection Study

The results from the selection study revealed that Icons-select took significantly more time to complete the task compared to other techniques. We believe that having multiple selection options cost additional time for the test subjects to find the association between icon and the desired object, which is even further increased when the icon–object assignments changes rapidly. This is more of a problem with many objects at close, but decreased by adding hysteresis, which on the other hand leads to a less intuitive order of icon–object assignments. This also may explain why more subjects described Icons-select to be harder to learn and perform compared to Touch-select and Center-select.

The Touch-select was chosen by the subjects to be the most tiring technique. We believe that this is likely because the users need to hold the device with one hand during the test. The statistics revealed that the amount of device movement with Icons-select was significantly less than the compared techniques, which is in contrast with the subjects’ indication of the method being more tiring than Center-select. This, we believe, can be explained by the longer mean task time with Icons-select that overall led to more fatigue.

We predicted that the Icons-select would cause more errors, by providing multiple selection choices. However, the results did not yield any statistically significant difference between the techniques. In our opinion this can be explained by the longer task time with Icons-select, which can mean that the subjects took more time to carefully choose the correct icon, and thus made fewer mistakes. This also agrees with some observations during the tests.

Overall we believe the Icons-select technique is more useful for crowded scenes to select multiple objects in a longer period of time as it is less tiring compared to the other techniques. However in

the applications where speed or accuracy is a priority, the Center-select or Touch-select would be a better choice.

5.2. Manipulation Study

The statistical results from the manipulation study indicate that Fix-UPR is faster than Relative-UPR. It also required significantly less device movement to complete the puzzle test. We believe the reason is that Fix-UPR consistently keeps the object in the view, while with Relative-UPR the subjects are required to realign the object and the view. This also explains why with Fix-UPR we saw significantly less re-grabbing of the objects. The subjects' preferences also agreed with the statistical results, indicating that Fix-UPR was easier to learn and use, and also less tiring and more natural and intuitive.

Overall, we believe that adding the gaze control to the device pose-based interaction makes it more robust and useful for most of the hand-held AR applications where the pose of a virtual object needs to be manipulated.

6. Conclusions

For this article we have explored and studied interaction techniques, specifically for selecting and manipulating objects in hand-held video see-through Augmented Reality. For selection, we have explored two techniques that reduce the amount of touch and conducted a user study that compare them and also compare them with a touch selection technique. For manipulation we have further developed an existing technique and explore through another user study its characteristics, together with the original technique.

In the first study we compare the commonly used Touch-select technique with two other techniques that allow the user to hold the device with two hands during the operation and removes the problem with finger occlusion. The compared techniques were Center-select, a technique selecting the object nearest to the center of the screen based on device movements (DM), and Icons-select, a list-based technique that allows the users to select one of multiple objects using icons, and the commonly used touch selection. The study contained 18 participants \times 3 techniques \times 1 task \times 30 selections = 1620 selections in total. The results revealed that more subjects preferred Center-select and chose it as more natural and intuitive technique. Center-select was also statistically significantly faster than Icons-select, however the mean time difference with Touch-select was not significant. The Icons-select technique required significantly less movements of the device compared to both Touch-select and Center-select, and was chosen by the subjects to be less tiring than Touch-select, however the subjects believed Center-select was less tiring than the Icons-select.

The Relative-UPR DM-based manipulation technique, which according to literature performs better than other DM-based techniques [11], has been further improved to resolve its issues with translation, by locking the object to the view. The second study was designed to compare this modified technique, Fix-UPR, with its originally suggested form. The second study contained 18 participants \times 2 techniques \times 1 task \times 9 dockings = 324 dockings in total. The results from the second study show that the subjects overall preferred Fix-UPR, which was also selected as the more natural and intuitive technique. This result is supported by the statistical analysis, where Fix-UPR was significantly faster than Relative-UPR. Furthermore, the subjects selected Fix-UPR to be less tiring, matching the significantly reduced amount of device movement with the Fix-UPR compared to Relative-UPR. The number of re-grabs for each object, which indicates the manual perspective corrections, was significantly less with Fix-UPR. This is explained by its ability to keep the object consistently in the view.

Author Contributions: Conceptualization, A.S. and K.L.P.; Resources, A.S.; Software, A.S.; Supervision, K.L. Palmerius; Writing—original draft, A.S.; Writing—review and editing, A.S. and K.L.P.

Funding: This work has been financed by the Swedish Foundation for Strategic Research (SSF) grant RIT15-0097.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR	Augmented Reality
UPR	user perspective rendering
DPR	device perspective rendering
DoF	degrees of freedom
DM	device pose-based manipulation

References

1. Mossel, A.; Venditti, B.; Kaufmann, H. DrillSample: Precise selection in dense handheld augmented reality environments. In Proceedings of the Virtual Reality International Conference: Laval Virtual, Laval, France, 20–22 March 2013; p. 10.
2. Wagner, D.; Pintaric, T.; Ledermann, F.; Schmalstieg, D. *Towards Massively Multi-User Augmented Reality on Handheld Devices*; Springer: Berlin, Germany, 2005.
3. Rekimoto, J. Transvision: A Hand-Held Augmented Reality System For Collaborative Design. 1996. Available online: https://www.researchgate.net/profile/Jun_Rekimoto/publication/228929153_Transvision_A_hand-held_augmented_reality_system_for_collaborative_design/links/0fcfd50fde14404200000000.pdf (accessed on 24 April 2019).
4. Jung, J.; Hong, J.; Park, S.; Yang, H.S. Smartphone As an Augmented Reality Authoring Tool via Multi-touch Based 3D Interaction Method. In Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, Singapore, 2–4 December 2012; pp. 17–20.
5. Hilliges, O.; Kim, D.; Izadi, S.; Weiss, M.; Wilson, A. HoloDesk: Direct 3D Interactions with a Situated See-through Display. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA, 5–10 May 2012; pp. 2421–2430.
6. Hincapie-Ramos, J.D.; Ozacar, K.; Irani, P.; Kitamura, Y. GyroWand: An Approach to IMU-Based Raycasting for Augmented Reality. *IEEE Comput. Graph. Appl.* **2016**, *36*, 90–96. [[CrossRef](#)] [[PubMed](#)]
7. Hancock, M.; ten Cate, T.; Carpendale, S. Sticky Tools: Full 6DOF Force-based Interaction for Multi-touch Tables. In Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, Banff, AB, Canada, 23–25 November 2009; pp. 133–140.
8. Martinet, A.; Casiez, G.; Grisoni, L. Integrality and Separability of Multitouch Interaction Techniques in 3D Manipulation Tasks. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 369–380. [[CrossRef](#)] [[PubMed](#)]
9. Henrysson, A.; Billinghurst, M.; Ollila, M. Virtual Object Manipulation Using a Mobile Phone. In Proceedings of the 2005 International Conference on Augmented Tele-Existence, Christchurch, New Zealand, 5–8 December 2005; pp. 164–171.
10. Mossel, A.; Venditti, B.; Kaufmann, H. 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, 20–22 March 2013; pp. 12:1–12:10.
11. Samini, A.; Lundin Palmerius, K. 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, 2–4 November 2016; pp. 121–128.
12. Squire, K.D.; Jan, M. Mad City Mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *J. Sci. Educ. Technol.* **2007**, *16*, 5–29. [[CrossRef](#)]
13. Huynh, D.N.T.; Raveendran, K.; Xu, Y.; Spreen, K.; MacIntyre, B. Art of Defense: A Collaborative Handheld Augmented Reality Board Game. In Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games, New Orleans, Louisiana, 4–6 August 2009; pp. 135–142.
14. Mulloni, A.; Seichter, H.; Schmalstieg, D. Handheld Augmented Reality Indoor Navigation with Activity-based Instructions. In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, Stockholm, Sweden, 30 August–2 September 2011; pp. 211–220.

15. Dunser, A.; Billingham, M.; Wen, J.; Lehtinen, V.; Nurminen, A. Technical Section: Exploring the Use of Handheld AR for Outdoor Navigation. *Comput. Graph.* **2012**, *36*, 1084–1095. [[CrossRef](#)]
16. Yoshida, T.; Kuroki, S.; Nii, H.; Kawakami, N.; Tachi, S. ARScope. In Proceedings of the ACM SIGGRAPH 2008 New Tech Demos, Los Angeles, CA, USA, 11–15 August 2008; p. 4:1.
17. Hill, A.; Schiefer, J.; Wilson, J.; Davidson, B.; Gandy, M.; MacIntyre, B. Virtual transparency: Introducing parallax view into video see-through AR. In Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, Switzerland, 26–29 October 2011.
18. Tomioka, M.; Ikeda, S.; Sato, K. Approximated user-perspective rendering in tablet-based augmented reality. In Proceedings of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, SA, Australia, 1–4 October 2013; pp. 21–28.
19. Samini, A.; Lundin Palmerius, K. A Perspective Geometry Approach to User-perspective Rendering in Hand-held Video See-through Augmented Reality. In Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology, Edinburgh, Scotland, 11–13 November 2014; pp. 207–208.
20. Yang, J.; Wang, S.; Sörös, G. User-Perspective Rendering for Handheld Applications. In Proceedings of 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Los Angeles, CA, USA, 18–19 March 2017.
21. Goh, E.S.; Sunar, M.S.; Ismail, A.W. 3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review. *IEEE Access* **2019**, *7*, 40581–40601. [[CrossRef](#)]
22. Reisman, J.L.; Davidson, P.L.; Han, J.Y. A Screen-space Formulation for 2D and 3D Direct Manipulation. In Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, Victoria, BC, Canada, 4–7 October 2009; pp. 69–78.
23. Liu, J.; Au, O.K.C.; Fu, H.; Tai, C.L. Two-Finger Gestures for 6DOF Manipulation of 3D Objects. *Comput. Graph. Forum* **2012**, *31*, 2047–2055. [[CrossRef](#)]
24. Hürst, W.; van Wezel, C. Gesture-based interaction via finger tracking for mobile augmented reality. *Multimed. Tools Appl.* **2013**, *62*, 233–258. [[CrossRef](#)]
25. Sasakura, M.; Kotaki, A.; Inada, J. A 3D Molecular Visualization System with Mobile Devices. In Proceedings of the 2011 15th International Conference on Information Visualisation, London, UK, 13–15 July 2011; pp. 429–433.
26. Tanikawa, T.; Uzuka, H.; Narumi, T.; Hirose, M. Integrated view-input interaction method for mobile AR. In Proceedings of the 2015 IEEE Symposium on 3D User Interfaces (3DUI), Arles, France, 23–24 March 2015; pp. 187–188.
27. Marzo, A.; Bossavit, B.; Hachet, M. 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, HI, USA, 4–5 October 2014; pp. 13–16.
28. Baldauf, M.; Fröhlich, P. Snap Target: Investigating an Assistance Technique for Mobile Magic Lens Interaction With Large Displays. *Int. J. Hum. Comput. Interact.* **2014**, *30*, 446–458, [[CrossRef](#)]
29. Taylor, R.M., II; Hudson, T.C.; Seeger, A.; Weber, H.; Juliano, J.; Helser, A.T. VRPN: A Device-independent, Network-transparent VR Peripheral System. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology, Baniff, AB, Canada, 15–17 November 2001; pp. 55–61.
30. SDL: Simple DirectMedia Layer. Available online: <http://www.libsdl.org/> (accessed on 1 November 2018).
31. Bradski, G. opencv. *Dr. Dobb's J. Softw. Tools* **2000**. Available online: <http://www.drdobbs.com/open-source/the-opencv-library/184404319> (accessed on 24 April 2019).

