Article

AR Object Manipulation on Depth-Sensing Handheld Devices

Koukeng Yang, Thomas Brown and **Kelvin Sung** *

Computing and Software Systems, University of Washington Bothell, Bothell, WA 98011, USA
* Correspondence: ksung@uw.edu

check for updates

Abstract: Recently released, depth-sensing-capable, and moderately priced handheld devices support the implementation of augmented reality (AR) applications without the requirement of tracking visually distinct markers. This relaxed constraint allows for applications with significantly increased augmentation space dimension, virtual object size, and user movement freedom. Being relatively new, there is currently a lack of study on issues concerning direct virtual object manipulation for AR applications on these devices. This paper presents the results from a survey of the existing object manipulation methods designed for traditional handheld devices and identifies potentially viable ones for newer, depth-sensing-capable devices. The paper then describes the following: a test suite that implements the identified methods, test cases designed specifically for the characteristics offered by the new devices, the user testing process, and the corresponding results. Based on the study, this paper concludes that AR applications on newer, depth-sensing-capable handheld devices should manipulate small-scale virtual objects by mapping directly to device movements and large-scale virtual objects by supporting separate translation and rotation modes. Our work and results are the first step in better understanding the requirements to support direct virtual object manipulation for AR applications running on a new generation of depth-sensing-capable handheld devices.

Keywords: augmented reality; direct object manipulation; handheld mobile devices; marker AR

1. Introduction

The Merriam Webster dictionary defines augmented reality (AR) as "an enhanced version of reality created by the use of technology to overlay digital (this paper uses "digital," and "virtual" interchangeably) information on an image being viewed through a device" (https://www.merriam-webster.com/dictionary/augmented+reality). Delivering AR on modern handheld platforms (handheld or mobile devices do not include general wearable devices like wearable backpack computers [1] or smart glasses [2]. This paper uses "handheld" and "mobile" interchangeably), such as ubiquitous smartphones or tablet devices, is an effective way of connecting the general public to this technology to promote the creation of next-generation applications [3,4].

Until recently, explicitly positioned, predefined visual markers were relied on to establish a correspondence between real and virtual worlds on handheld platforms, e.g., [5–7]. These markers must be maintained within the view of an application at all times, are processed dynamically, and tracked in real time [8]. Virtual information is integrated into the physical environment based on the location and orientation of these visual markers.

The recent introduction of depth-sensing capabilities on moderately priced commercial handheld products through either dedicated hardware [9] or software libraries [10,11] eliminates the visual marker requirement. Newer generations of AR applications on these devices (e.g., [3,4]) can detect, track, and integrate virtual information into the augmented environment based on geometric information of actual physical objects and do not require any visually distinct markers.

This paper categorizes visual-marker-based AR applications as "marker AR" and the associated virtual object manipulation methods as "marker methods." "Markerless AR" will be used to refer to newer depth-sensing AR applications that are capable of tracking elements in the physical environment without visual markers, and the corresponding virtual object manipulation methods will be referred to as "markerless methods."

While marker methods have been well examined (e.g., [6,7,12–16]), there is currently a lack of study on markerless methods due to the relative novelty of the associated capable devices. Although marker and markerless AR differ only by a seemingly simple marker-tracking requirement, the implications on virtual object manipulation methods and the resulting AR applications can be substantial.

The understanding of issues concerning markerless methods is an important first step in realizing the potential of markerless AR. This paper presents our approach to this research opportunity and the corresponding results. Our study examines the ramifications of eliminating the marker-tracking requirement, articulates the corresponding impacts, designs new test cases in response to these impacts, and re-evaluates existing marker methods in a markerless AR setting. Our results indicate that many of the results from marker AR studies are valid for markerless AR, although specific applicability depends on the size of the virtual object being manipulated.

This paper begins with a survey of existing marker methods and explores the implications of removing the visual marker requirement for handheld AR systems. The sections that follow describe the choice of the marker methods for testing, the test suite implementation, and the test cases explored. The details of user testing and results are then presented, with a discussion of implications concluding the paper.

2. Literature Review and Background

The field of AR is vast [8]. Our work focuses on user interactions with virtual objects for handheld AR applications.

2.1. Object Manipulation in Marker Augmented Reality (AR)

Manipulating virtual objects on handheld platforms presents unique challenges due to a smaller screen size, a lack of dedicated input devices, and the fact that one of the user's hands is typically holding the device [6]. Early approaches focused on available hardware sensors and considered the manipulation of virtual object attributes independently. Examples of such approaches include separating translation and rotation of an object by mapping the position to physical device movement and the orientation to a virtual keyboard [6] or mimicking physical world interaction by tracking one of the user's fingers [7].

More recently, researchers have demonstrated that properly designed gestures on multi-touch screens can significantly outperform finger tracking [12]. Additionally, it has been recognized that interaction designs should build around manipulation requirements rather than device hardware functionality. These observations have led to mapping the translation and rotation of virtual objects to the natural movements of handheld devices [13,14] and intuitively combining multi-touch with device movements [15].

2.2. Marker Methods

The demonstrated effectiveness and ever-present touch screens on handheld devices identify the following as prime candidates for markerless AR virtual object interactions: (1) 3DTouch [13], (2) HOMER-S [13], (3) integrated view input (IVI) [14], and (4) hybrid of touch screen and movement [15].

3DTouch [13] combines device orientation with touch screen finger tracking. Users touch to select an object; choose to translate, rotate, or scale an object; and then swipe on the device screen to carry out the intended transformation. HOMER-S [13], similar to 3DTouch, supports distinct object selection through the touch screen interface. However, instead of swiping on the touch screen, HOMER-S treats

the device as a one-to-one proxy of the selected virtual object and manipulates the object by moving and rotating the device.

The IVI [14] method is similar to HOMER-S in treating the device as a one-to-one proxy for manipulation of the selected virtual object. However, IVI requires users to touch and hold the device screen to maintain object selection during manipulation.

Lastly, the hybrid [15] method is similar to 3DTouch and HOMER-S in maintaining a selected object and requiring users to specify a manipulation mode. The hybrid method is a hybrid because it translates virtual objects by treating the device as a one-to-one proxy of the selected virtual object and rotates them by interpreting swiping gestures on the touch screen. Please refer to the original papers and an online video demonstration, please follow the link in the Supplementary Materials, https://youtu.be/9dztL0pUllM for more details of each of these methods.

2.3. Marker Method Interaction Characteristics

The discussed marker methods can be characterized according to three important aspects.

1. Object selection and manipulation states: Three of the methods (3DTouch, HOMER-S, and hybrid) define discrete modes. Input actions only operate on the currently selected object, and the interpretation depends on the mode of manipulation selected.
2. Touch screen utilization: All the methods perform object selection based on the touch screen, whereas only 3DTouch and hybrid support swipe manipulation.
3. Proxy manipulation: The methods treat the device as a one-to-one proxy of the selected object in varying degrees. 3DTouch does not support proxy manipulation, hybrid supports proxy translation manipulation, and both HOMER-S and IVI support proxy translation and orientation manipulation.

The effects of these characteristics must be examined in the context of markerless AR.

3. The Context of Markerless AR

Without the requirement of maintaining a marker within the application view, many significant limitations of marker AR are overcome by markerless AR. This section examines the limitations of marker AR and articulates the comparative benefits of markerless AR.

It should be noted that the following discussion assumes single-marker AR applications. Marker AR applications that support multiple distributed markers should have increased capabilities. However, they also involve complexities in other dimensions, including preplanning for marker placements, strategies for supporting transitions between markers, etc. These issues are outside the scope of this paper.

3.1. Augmentable Space

The dimension of augmentable space in marker AR is governed by the size of the visual marker because virtual objects are typically projected onto these markers. An increase in augmentable space requires a larger marker and appropriate accommodations for their impact on the interaction environment. In contrast, the entire visible volume of markerless AR constitutes its augmentable space [16].

3.2. Visibility to Specific Physical Objects

The visualization of virtual objects in marker AR is dependent on the continuous detection and tracking of associated visual markers. An occlusion or partial occlusion of the marker can cause erroneous detection and result in the disappearance or incorrect projection of the virtual objects [17]. In this way, users of marker AR applications must maintain appropriate visibility to the physical objects with the markers. Without visual markers, markerless AR does not need to detect or track a predefined set of physical objects in the environment.

3.3. User Movement

User movement in marker AR is restricted by the need to maintain visual markers in the application view. An acute viewing angle results in distorted visual marker registration and can be challenging for maintaining continuous marker tracking. For this reason, users of marker AR can typically only examine and interact with virtual objects based on a restricted set of viewing angles [18]. In contrast, markerless AR users are free to move around and examine virtual objects from any viewing angle.

3.4. Impact on Virtual Object Manipulation Assessment

The above discussion identifies that markerless methods should be assessed based on manipulations that involve the following:

1. a large physical environment,
2. partial occlusion by both physical and other virtual objects, and
3. a large variation of viewing angles.

These requirements suggest that the discussed marker methods should be re-examined for validity in the context of markerless AR. Additionally, improvements to handheld interaction methods designed specifically to address the above requirements could provide interesting interaction alternatives for a markerless AR setting.

4. Study and Implementation

We were interested in examining and understanding the characteristics of marker AR virtual object manipulation methods in the context of markerless AR. Similar to previous studies, our focus was on the most basic virtual object interactions: selection, positioning, and rotation [19]. These are fundamental building blocks that can be combined to accomplish other, more significant tasks.

Our study followed a pattern similar to previous works: identify appropriate object manipulation methods, implement these methods on a modern markerless-AR-capable device, design test cases, collect relevant statistics on end users completing the test cases, summarize observations, and derive conclusions.

4.1. The Marker Methods to be Tested

As discussed above, the marker methods to be re-evaluated were 3DTouch [13], IVI [14], HOMER-S [13], and a hybrid of touch screen and movement [15].

The Modified IVI (MOD IVI) Method

When considering the markerless AR setting, the IVI's stateless, device proxy manipulation aligns well with the increased augmented space and freedom of the user movement observed. We propose the MOD IVI interaction method to enhance the IVI for a larger augmentable space. MOD IVI has the following control interface:

1. Double tap to select and automatically translate an object to a distance of one meter in front of the user.
2. Vertically swipe to move a selected object away/toward the user.
3. Vary the transparency of a selected object based on its distance to the user, with closer objects exhibiting greater transparency.

The last enhancement of varying transparency is designed to address the potential challenges of interacting with a large virtual object that may result in partial occlusion of the interaction environment.

4.2. The Test Suite Implementation

The chosen device for test suite implementation was a Lenovo Phab 2 Pro. This device offers hardware depth sensors and supports the Google Tango AR platform [9]. Google Tango is a precursor to the current Google ARCore [11], and it provided a stable platform for our markerless AR test suite implementation. Although Tango and Phab2 Pro have been discontinued, the presented ideas and results are valid in general for any mobile markerless-AR-capable device.

Figure 1 depicts the system support for the test suite implementation. While the Android Software Development Kit (SDK) provided system-level support, the 3D virtual object interactions were based on the Unity game engine framework. The Google Tango SDK depended on both the Android SDK and the Unity game engine framework as it presented the interface for markerless-AR-specific functionality. The software application for this study was implemented on top of all three libraries.
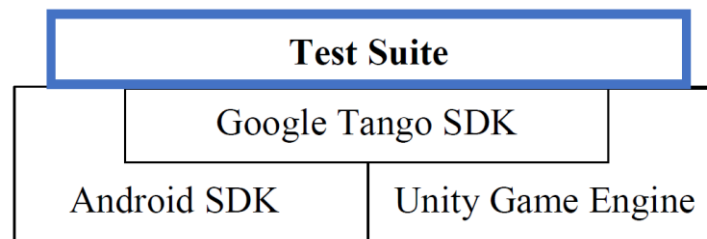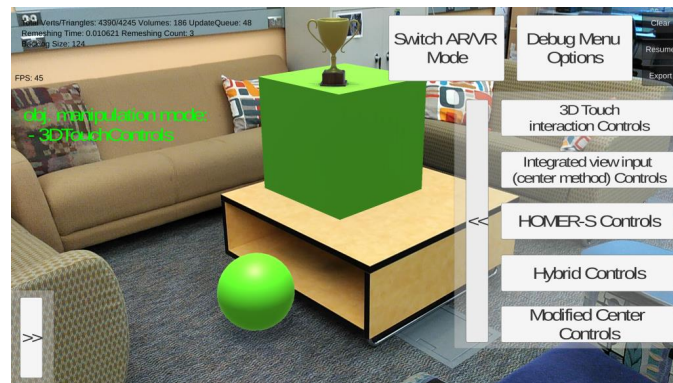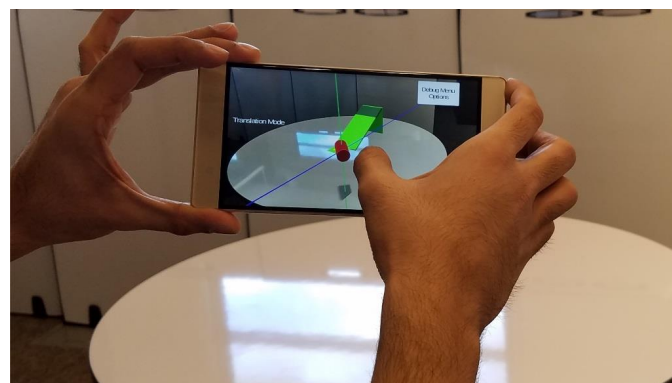

Figure 1. System support for the suite implementation.

Figure 2a shows the test runtime where the user could select from one of the five methods to be tested: 3DTouch, IVI, HOMER-S, hybrid, and MOD IVI. Figure 2b shows a user interacting with a test case (discussed in the next section).

(**a**)

(**b**)

Figure 2. The test suite implementation. (**a**) Initial user interface screen; (**b**) user running a test case.

5. Test Cases

The primary goal of this study was to understand the implications of adopting marker methods for markerless AR. Our tests therefore needed to include existing marker AR cases [13]. Additionally, it was essential to examine the flexibilities offered by markerless AR, i.e., larger augmented space and virtual object size, agnosticism toward inter-object occlusion, improved freedom of movement, and large variations in viewing angle.

Seven test cases were derived. The first three tests were adopted directly from the HOMER-S studies [13]. The last four tests were designed specifically for markerless AR.

5.1. Small-Scale Tests: Tests 1 to 3

Figure 3 shows the three adopted tests where the size of the physical marker was an important design decision. Figure 3a shows Test 1, 2D translation, where the user was expected to move a 5-cm cube along the 2D table top to touch the small square. Figure 3b illustrates Test 2, 3D translation, where the user had to move the 5-cm cube from the table top to rest on a 30-cm-tall column. Figure 3c depicts Test 3, 3D translation and rotation, where the user was expected to orient and translate a 15-cm-long cylinder in 3D space. The cylinder needed to be moved to the top of a ramp in such an orientation that the physics of the test would cause the cylinder to roll down the ramp and connect with a small square sitting at the base of the ramp.
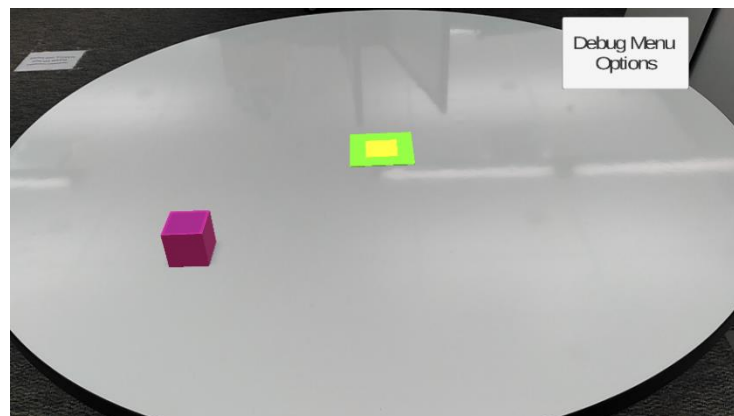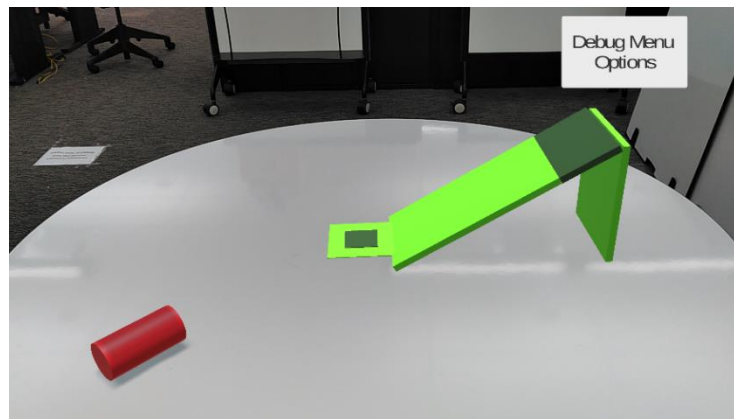

(**a**)
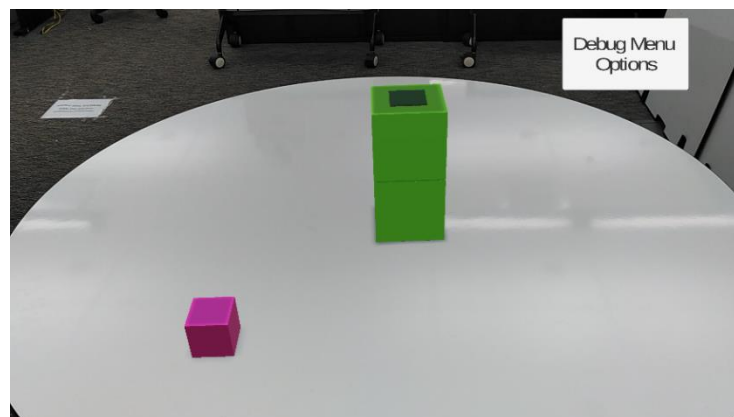
(**b**)

Figure 3. *Cont.*

(**c**)

Figure 3. Small-scale tests. (**a**) Test 1: 2D translation; (**b**) Test 2: 3D translation; (**c**) Test 3: 3D translation and rotation.

5.2. Large-Scale Tests: Tests 4 to 7

Tests 4 to 7 were designed specifically to take advantage of virtual object manipulation in the absence of the visual marker constraint. Tests 4 and 5 roughly corresponded to Tests 1 and 2 with the distinction that both the augmented space and object dimensions were about an order of magnitude larger.

Figure 4a shows Test 4, the large-scale translation test. In this case, the user had to move a 50-cm cube from the left to the right table, which was a distance of 4.26 meters (14 feet) away. The distance between the tables was governed by the 4.57 meter (15 feet) effective distance of the Lenovo Phab 2 Pro device. A physical barrier divided the space between the two tables, where the user had to navigate around the physical barrier to complete the test. This test was completed when the cube was moved onto the target on the right table.

Figure 4b illustrates Test 5, the large object placement test. This test examined the manipulation of a large object with full freedom of movement, where a 50-cm cube had to be moved from the right to the left table (the reversed left/right movement was to avoid pattern of movements) and stacked on top of a column of two existing 50-cm cubes. The combined height of the table with the stack of cubes was two meters above the ground. The relatively large change in height was designed to enforce large viewing angle variations. The starting position for the user was in the middle of the two tables facing the cube to be moved. This test was completed when the cube was stacked without falling off.
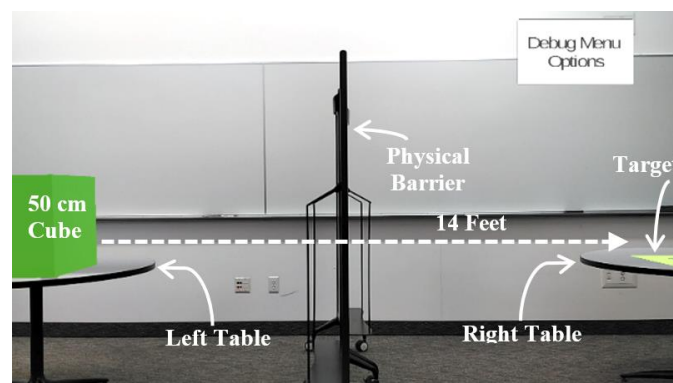

(**a**)

Figure 4. *Cont.*
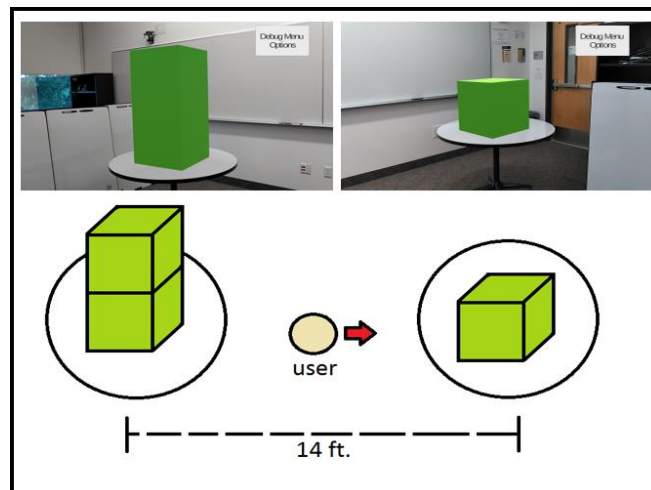
(**b**)

Figure 4. Large-scale object manipulation. (**a**) Test 4: translation; (**b**) Test 5: object placement.

Test 6 was the movement/occlusion test, which tested the impacts of freedom of movement and occluded line of sight to a virtual object being manipulated. As illustrated in Figure 5a, the user had to move a 50-cm cube from the left to the right table while being constrained to a distance of at least 2 meters away. In between the two tables was a 2-m-wide barrier designed to occlude the user's line of sight. The user was free to move left and right to negotiate partial virtual object occlusion.
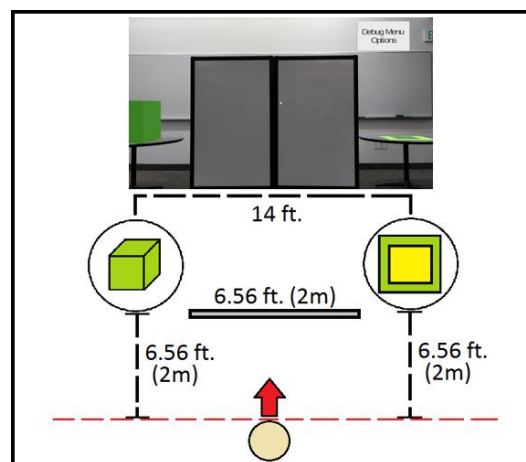
(**a**)
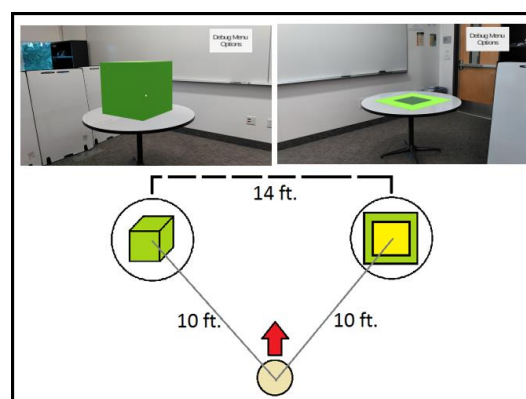
(**b**)

Figure 5. Working with constraints. (**a**) Test 6: view occlusion; (**b**) Test 7: restricted movement.

Test 7 was the user movement test, which mimicked the scenario where a user's movement is fixed in the physical environment. As illustrated in Figure 5b, the user had to move a 50-cm cube from the left to the right table while standing at a fixed location of 10 feet away from the two tables.

6. User Testing

User testing began after initial pretesting for refinements to the test cases and the testing protocol. Inadvertently, testing was divided into two phases. The analysis of initial results offered clear insights and optimization pointers that resulted in a modified second phase of testing.

A total of 34 testers participated in this study. The testers were exclusively university computer science undergraduate and graduate students. Half of the testers were women, and the other half were men. None reported any prior experience with AR devices, and all reported to be experienced with mobile devices. Phase One involved 14 participants, and Phase Two involved 20 different participants. There was no overlap of users in the two phases. In the rest of this paper, "users" and "testers" are used interchangeably.

6.1. Testing Protocol

Our testing protocol emulated that of the HOMER-S study [13], i.e., at the beginning of a test, the object manipulation method was explained to the user, and the user was given five minutes to practice. This testing protocol was consistent between the two phases.

When a user was ready, the objective of the test was explained, and the user was given the opportunity to walk around and examine the virtual objects through the AR device before they began. During this observation period, the user was not allowed to manipulate the virtual objects. However, they were encouraged to strategize approaches. The user was instructed to begin testing from a predefined starting location for each test. The test suite kept track of the testing time; the user had to click a virtual start button on the AR device to begin, and the timer was stopped automatically when a task completion condition was met.

Users were given a posttest questionnaire after they had completed all seven test cases for a given virtual object manipulation method. The questionnaire was used to collect qualitative feedback on how users felt during the tests as well as how they felt about the virtual object manipulation methods they tested. The posttest questionnaire was adopted from the HOMER-S study [13] with minor modifications for our test cases.

6.2. Data Analysis

The collected timing information were analyzed with Microsoft Excel using the Microsoft Excel Analysis ToolPak add-on [20].

We used the unequal variance version of the *t*-test in our analysis with an alpha of 0.05. In the Phase Two result analysis, the one-tailed *t*-test was used to verify differences in the performances of methods (e.g., MOD IVI vs. IVI), whereas the two-tailed *t*-test was used to analyze similarities in the timing results (e.g., 3DTouch and HOMER-S).

7. Phase One Testing and Results

To avoid exhaustion, for each user, two of the five object manipulation methods were randomly selected. For each of the methods, the user had to complete all seven test cases in one setting. All data set analyses were treated uniformly due to the randomized testing order. Table 1 shows the minimum, maximum, and mean test completion data for all seven tests by all users. The red highlights identify the shortest average completion time for each test. For example, the shortest average completion time for Test 1 was 7.3 seconds for the IVI method. Three observations could be made.

Table 1. Phase One completion time (seconds) with highlight of shortest completion time.

Tests	IVI			MOD IVI			3DTouch			HOMER-S			Hybrid		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Test 1	5.1	9.4	**7.3**	6.4	13.3	8.8	24.6	45.4	36.9	15.9	21.5	18.9	61.3	86.8	71.9
Test 2	4.9	12.3	8.8	5.5	12.2	**8.7**	24.8	32.7	28.7	17.0	18.8	17.8	70.1	85.9	75.6
Test 3	8.2	21.9	16.3	9.4	15.0	**13.4**	40.4	136.6	78.1	44.2	108.7	79.2	121.6	198.6	145.4
Test 4	63.9	77.8	68.4	63.6	75.7	67.5	20.9	40.0	32.2	11.5	25.0	**18.3**	100.7	136.4	123.4
Test 5	116.2	149.3	129.5	101.3	153.8	132.8	26.6	136.6	62.0	7.3	79.8	**27.7**	119.7	397.7	182.9
Test 6	64.0	83.5	74.2	63.7	79.8	69.8	23.8	35.6	27.5	14.1	29.9	**19.2**	106.5	162.5	125.1
Test 7	30.6	39.2	35.6	24.2	33.8	29.1	16.2	24.0	21.4	7.1	11.5	**9.3**	104.7	114.4	127.7

7.1. Performance of the Hybrid Method

It can be seen from the hybrid (last) column of Table 1 that, for every test case, the mean completion time of the hybrid method was drastically worse. In fact, except for Test 5, the minimum completion time of the hybrid method was worse than the maximum completion time of the other four methods. This consistently worse performance was unexpected because the method is an attempt to intuitively combine device movement and touch screen control.

Users reported in posttest feedback that it was awkward and clumsy to switch from physical one-to-one proxy motion control to logical touch-screen rotation manipulation. In addition, the z-axis indicator of the selected object was often obstructed by other virtual objects, which caused further confusion.

7.2. Method Sensitivity to Size of Space and Objects

Without considering the hybrid results, it can be observed from the mean completion time in Table 1 for Tests 1 to 3 that IVI and MOD IVI significantly outperformed 3DTouch and HOMER-S; the maximum completion time of the former two was shorter than the minimum completion time of the latter two. It can be further observed that, excluding the anomaly of Test 5 results, the method and completion time relationship was swapped for results from Tests 4 to 7, where Touch3D and HOMER-S significantly outperformed IVI and MOD IVI.

Recalling that Tests 1 to 3 were designed for small-scale objects whereas Tests 4 to 7 were for larger-scale augmented space and virtual objects, these results suggest that users welcomed the stateless simplicity and one-to-one proxy manipulation of IVI and MOD IVI for small-scale objects. However, users reported difficulties in maintaining finger contact with the touch screen while moving their entire body to simultaneously translate and rotate objects in large-scale tests.

In contrast, 3DTouch and HOMER-S allowed users to separate virtual object manipulation into distinct and logical steps: first, select an object; second, choose a transformation mode; and finally, focus on the actual manipulation of the currently selected object.

7.3. Problem with Design of Test 5

Test 5 required stacking a cube on top of a 2-meter-tall column. It was easy to verify that the viewing angle was not a concern during the manipulation. However, the completion time of this task turned out to be highly dependent on the user's height. This design flaw was a result of our concerns for the impact of the viewing angle on user interaction. This test case was kept in Phase Two because of the project team's interest in observing potentially noteworthy trends from the results.

8. Phase Two Testing and Results

Phase Two was optimized accordingly, i.e., the hybrid method was dropped, IVI and MOD IVI methods were only performed on Tests 1 to 3, and 3DTouch and HOMER-S methods were only performed on Tests 4 to 7. The results from this phase would identify the most suitable methods for manipulating small- and large-scale virtual objects.

Each of the recruited 20 users was randomly assigned either IVI or MOD IVI for completing Tests 1 to 3 and either 3DTouch or HOMER-S for Tests 4 to 7. In this way, all Phase Two testers only ran through all seven test cases once. The collected data was analyzed in a similar fashion as in Phase One.

8.1. Small-Scale Test Results

Table 2 shows the mean completion time and the corresponding standard deviation (SD) of IVI and MOD IVI methods. The last column shows the *p*-value of the one tailed *t*-test when comparing the results. Recalling that MOD IVI was proposed to improve the IVI method for manipulating large-scale objects, e.g., swiping to move objects farther/closer, these features were not relevant for small-scale tests, and it is not surprising that there were no significant differences observed between Tests 1 and 2.

Table 2. Phase Two: integrated view input (IVI) and Modified (MOD) IVI comparison.

Tests	IVI		MOD IVI		*p*-value
	Mean	SD	Mean	SD	
Test 1	6.7	1.6	7.2	2.2	0.240
Test 2	8.8	2.4	8.4	2.2	0.310
Test 3	18.8	11.	13.0	2.2	0.037

The 0.037 *p*-value of Test 3 suggests that varying the selected object translucency is an effective option for working with increasingly complex virtual environments.

8.2. Large-Scale Test Results

Table 3 shows the mean, standard deviation (SD), and two-tailed *p*-values when comparing results from the 3DTouch and HOMER-S methods.

Table 3. Phase Two: 3DTouch and HOMER-S comparison.

Tests	3DTouch		HOMER-S		*p*-value
	Mean	SD	Mean	SD	
Test 4	39.3	18.6	18.8	2.9	0.00043000
Test 5	55.0	28.1	30.4	17.3	0.00660000
Test 6	39.3	10.5	18.9	4.3	0.00000037
Test 7	23.3	10.2	11.4	2.4	0.00025000

3DTouch was slower than the HOMER-S method in all of the large-scale tests. While both methods allowed users to select an object and transformation mode, the actual manipulation was achieved by swiping on the touch screen for 3DTouch and one-to-one proxy manipulation for HOMER-S. The results from Table 3 clearly illustrate that users in our test cases uniformly preferred manipulation via intuitive device movement. It is interesting that this preference was consistent even for the ill-designed Test 5, although with a higher variability between users.

9. Qualitative Feedback Results

After testing, participants completed a Likert scale survey with a range between 1 (disagree) to 7 (agree). As detailed in Appendix A, users' opinions on the length of practice time, comfort level in using the device, and general intuitiveness of each of the object manipulation methods were collected.

Table 4 presents the average from the questionnaire. It is interesting that all responses were highly positive, with average values greater than 6 and *p*-values much lower than 0.05. These results show that the users believed 5 minutes of practice time was sufficient; they were in general comfortable working with mobile AR application; and that they did care about ease of use, speed, and accuracy when judging an object manipulation method.

Table 4. User feedback from posttest questionnaire.

Average	Practice Time	Comfort Level	Ease of Use	Speed	Accuracy
Average (max of 7)	6.15	6.49	6.12	6.37	6.47

In our posttest questionnaire, we also included questions on the usability and intuitiveness of the methods. We were cautious in interpreting the very high averages across all object manipulation methods in terms of usability and intuitiveness even when users struggled with some of the test cases. We speculate that the wow factor (https://dictionary.cambridge.org/us/dictionary/english/wow-factor) may have contributed to the overwhelmingly positive results. Our test subjects were notably impressed by the technology, which might have biased their response. For this reason, the conclusions will focus only on the quantitative timing results.

10. Conclusion and Future Work

We approached the study of object manipulation for markerless AR by examining the results from prior marker AR research and selecting potentially effective methods. MOD IVI was proposed to modify and enhance the IVI method, the latest and seemingly best marker AR method, to support the manipulation of large-scale objects.

In addition to adopting the testing protocol and test cases from previous marker AR studies, new test cases were derived to assess the methods in light of the significantly reduced constraints on virtual object size, augmented space dimension, inter-object occlusion, and freedom of user movement.

Our testing results showed that users found it difficult to manipulate objects based on a combination of physical one-to-one proxy movements and logical touch screen control. However, in all cases, users were efficient when the one-to-one proxy manipulation was consistently involved: IVI (or MOD IVI) for the small-scale cases and HOMER-S for the large-scale cases. This result is not surprising because markerless AR attempts to naturally augment virtual objects. It is instinctive for users to attempt to pick up an object and perform the exact movement and rotation needed by treating the handheld device as a proxy.

When the scale of objects is small, users can efficiently pick up and perform object movement/ rotation as one continuous action. However, when object sizes are large, users prefer separating the manipulation into distinct selection, movement, and rotation steps. This difference in preference seems intuitive. Consider in real life, the cases of moving a small object, e.g., a pencil, versus a large and potentially heavy object, e.g., a sofa. In the former, it is likely that the pencil manipulation is performed as a single continuous action, whereas for the sofa, it is conceivable that the sofa movement and rotation are purposely planned and carried out as distinct steps.

The proposed MOD IVI method did not significantly outperform any existing methods. Recalling the lesson learnt from the hybrid test results, for the case of markerless AR, users found it awkward to combine physical device and logical touch screen manipulations. In alignment with this observation, the proposed touch screen modifications to the IVI methods were not welcomed by users. However, Test 3 results do indicate that changing the transparency of selected objects is an effective addition.

Although our testers were gender balanced, they were uniformly technologically inclined, eager, and excited by recent AR advancements. We believe that this uniformity resulted in heavily biased opinion poll results. Additionally, our augmented space was in a research lab, and the virtual objects were uniformly simple geometric shapes. It is difficult to generalize how users may strategize and carry out object manipulations in other settings.

The direct object manipulation methods studied and the associated conclusions are targeted at markerless AR applications. Currently, markerless AR capability is enabled via depth-sensing capability. It is important to note that our study is independent from the underlying depth-sensing implementation technology. The key issues are devices that are capable in sensing, approximating, and

augmenting the physical environment without explicit placements of markers as well as devices with the ability to support the selection and manipulation of virtual objects in the augmented environment.

The mobile consumer market continues to undergo rapid changes. Since our study, support for mobile AR devices seems to be moving away from the more accurate but resource-demanding depth-sensing hardware. Instead, recent device releases from major manufacturers suggest an embracing of algorithmic and software approximations of the physical environments [10,11].

As a future study, it would be interesting to examine direct object manipulation from the perspective of working with software-approximated physical environment. In all cases, the conclusions from our study are valid for both hardware and software depth-sensing technologies, i.e., next-generation markerless AR applications should consider the relative size of virtual objects when designing virtual object interactions, the manipulation of small-scale virtual objects should be directly mapped to device movements, and the manipulation of large-scale virtual objects should be supported by separate translation and rotation modes.

Lastly, ours is an early approach to understand object manipulation in upcoming markerless AR applications. An interesting observation is that markerless AR offers a more realistic and seamless virtual object augmentation, which can lead to object manipulation strategies that closely resemble users' real-life experiences. Future work should focus on discovering a unified method that supports effective virtual object manipulation independent of size or perceived weight, for example, automatic switching between manipulation methods based on target object sizes. Additionally, refined test cases and user testing can be based on environments that resemble the real world.

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

Appendix A

Posttest User Questionnaire

Object manipulation method tested:		
1. How adequate do you feel the time allotted for practice was?		
	Not Enough (0)	More Than Enough (7)
2. How comfortable were you with using a smartphone for task completion?		
	Very Uncomfortable (0)	Very Comfortable(7)
3. How would you rate the object manipulation technique in usability? Speed? Accuracy?		
Usability	Very Unusable (0)	Very Usable (7)
Speed	Very Slow (0)	Very Fast (7)
Accuracy	Very Inaccurate (0)	Very Accurate (7)
4. How would you rate the intuitiveness of the object manipulation technique for object movement, object rotation, object movement combined with rotation?		
Object Movement	Very Unintuitive (0)	Very Intuitive (7)
Object Rotation	Very Unintuitive (0)	Very Intuitive (7)
Object Movement & Rotation	Very Unintuitive (0)	Very Intuitive (7)
5. When determining if you liked using the object manipulation technique, how important in influence on your decision was ease of use? Speed? Accuracy?		
Ease of Use	Not Important (0)	Very Important (7)
Speed	Not Important (0)	Very Important (7)
Accuracy	Not Important (0)	Very Important (7)

References

1. Feiner, S.; MacIntyre, B.; Hollerer, T. *Wearing It Out: First Steps Toward Mobile Augmented*; in the First International Symposium on Mixed Reality (ISMR'99); Springer-Verlag: Ohmsha, Tokyo, 1999; pp. 363–377.
2. Elder, S.; Vakaloudis, A. Towards uniformity for smart glasses devices: An assessment of function as the driver for standardisation. In Proceedings of the 2015 IEEE International Symposium on Technology and Society (ISTAS), Dublin, Lreland, 11–12 November 2015.
3. AR Experiments. Available online: https://experiments.withgoogle.com/ar (accessed on 22 March 2018).
4. Made with ARKit. Available online: http://www.madewitharkit.com/ (accessed on 22 March 2018).
5. Wagner, D.; Schmalstieg, D. First steps towards handheld augmented reality. In Proceedings of the Seventh IEEE International Symposium on Wearable Computers, White Plains, NY, USA, 21–23 October 2003; pp. 127–135.
6. 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.
7. Henrysson, A.; Marshall, J.; Billinghurst, M. Experiments in 3D Interaction for Mobile Phone AR. In Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast, Perth, Australia, 1–4 December 2007; pp. 187–194.
8. Billinghurst, M.; Clark, A.; Lee, G. A Survey of Augmented Reality. *Found. Trends Hum.-Comput. Interact.* **2015**, *8*, 73–272. [CrossRef]
9. Google Tango Platform. Available online: https://en.wikipedia.org/wiki/Tango_(platform) (accessed on 22 March 2018).
10. Apple AR Kit. Available online: https://developer.apple.com/arkit/ (accessed on 22 March 2018).
11. Google AR Core. Available online: https://developers.google.com/ar/ (accessed on 22 March 2018).
12. Bai, H.; Lee, G.A.; Billinghurst, M. Freeze View Touch and Finger Gesture Based Interaction Methods for Handheld Augmented Reality Interfaces. In Proceedings of the 27th Conference on Image and Vision Computing, Dunedin, New Zealand, 26–28 November 2012; pp. 126–131.
13. 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. 1–10.
14. Tanikawa, T.; Uzuka, H.; Narumi, T.; Hirose, M. Integrated View-input AR Interaction for Virtual Object Manipulation Using Tablets and Smartphones. In Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology, Iskandar, Malaysia, 16–19 November 2015; pp. 1–7.
15. 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.
16. Pucihar, K.Č.; Coulton, P. Exploring the Evolution of Mobile Augmented Reality for Future Entertainment Systems. *Comput. Entertain.* **2015**, *11*, 1–16. [CrossRef]
17. Garrido-Jurado, S.; Muñoz-Salinas, R.; Madrid-Cuevas, F.J.; Marín-Jiménez, M.J. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* **2014**, *47*, 2280–2292. [CrossRef]
18. Herout, A.; Zachariáš, M.; Dubská, M.; Havel, J. Fractal marker fields: No more scale limitations for fiduciary markers. In Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Atlanta, GA, USA, 5–8 November 2012.
19. Bowman, D.A.; Hodges, L.F. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In Proceedings of the 1997 Symposium on Interactive 3D Graphics, Providence, Rhode Island, USA, 27–30 April 1997; p. 35.
20. Use the Analysis ToolPack to Perform Complex Data Analysis. Available online: https://support.office.com/en-us/article/use-the-analysis-toolpak-to-perform-complex-data-analysis-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6 (accessed on 22 March 2018).