Computers & Graphics 95 (2021) 23–35

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

SlidAR+: Gravity-aware 3D object manipulation for handheld augmented reality[☆]

Varunyu Fuvattanasilp[a,*], Yuichiro Fujimoto[a], Alexander Plopski[b], Takafumi Taketomi[a], Christian Sandor[c], Masayuki Kanbara[a], Hirokazu Kato[a]

[a] *Nara Institute of Science and Technology, Nara, Japan*
[b] *University of Otago, North Dunedin, New Zealand*
[c] *City University of Hong Kong, Kowloon Tong, Hong Kong*

ARTICLE INFO

Article history:
Received 23 March 2020
Revised 22 December 2020
Accepted 18 January 2021
Available online 23 January 2021

Keywords:
Handheld augmented reality
Object manipulation
Gravity

ABSTRACT

Accurately placing virtual objects in a scene is a challenging tasks in handheld augmented reality (HAR). To add and arrange virtual objects in HAR, users must manipulate 6 degrees of freedom (DoFs) of the virtual object, namely: position (3) and orientation (3). However, it is difficult to manipulate all DoFs with the two-dimensional display of the handheld device. We present SlidAR+, a method for controlling the position and orientation of objects in HAR. SlidAR+ is an extension of SlidAR [1], a technique that allows users to control the position of a virtual object by manipulating only 1 DoF. We use the direction of gravity as a constraint to improve the user's control and reduce the time it takes to adjust the orientation. Upon comparing it with a state-of-the-art object manipulation method, using SlidAR+, user were able to complete the tasks faster under our expected conditions and were also preferred by most participants.

© 2021 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

In Augmented Reality (AR), virtual objects are superimposed over the real world on a display. AR is currently being utilized in industry [2], task support [3], education [4], and for medical applications [5]. An essential requirement of AR applications is the spatially consistent alignment of virtual and real objects [6].

Early AR applications were developed for head-mounted displays and desktop computers. However, with the spread of handheld devices, such as tablets and smartphones, more and more applications are taking advantage of their availability and technical capabilities. These devices combine a camera and a display in a single device, which makes them ideal for development of AR applications. Handheld AR (HAR) applications have the potential to be widely used by the public and can be accessed anytime and anywhere, thanks to the widespread adoption of the devices they were designed for.

Currently, many HAR applications utilize markers for tracking. However, this severely limits an application's ability to be used more generally. To address this, an increasing number of applications are utilizing visual simultaneous localization and mapping (vSLAM) [7] to track the pose of the device, without requiring users to set up fiducial markers or scan the environment beforehand. Some applications let users freely manipulate the virtual content in HAR applications [8,9]. However, in order to manipulate a virtual object, users have to adjust up to 6 degrees of freedom (DoFs) (3 DoFs for position, 3 DoFs for orientation), which is a difficult task as handheld devices are usually controlled through a 2D display. It is thus necessary to develop intuitive methods that allow users to control all 6 DoFs easily.

Polvi et al. developed SlidAR [1], a 3D positioning technique for HAR application that utilizes ray casting and epipolar geometry to simplify the procedure of positioning a virtual object in the real world. SlidAR allows users to adjust and reposition virtual content by performing only a slide gesture. Polvi et al. compared SlidAR with a device-centric positioning method [1] and their results showed that users can position virtual content faster with SlidAR.

[☆] This paper was recommended for publication by Mark Billinghurst.
[*] Corresponding author.
E-mail addresses: varunyu.fuvattanasilp.ut0@is.naist.jp (V. Fuvattanasilp), yfujimoto@is.naist.jp (Y. Fujimoto), alexander.plopski@otago.ac.nz (A. Plopski), taketomi.takafumi@ieee.org (T. Taketomi), christian@sandor.com (C. Sandor), kanbara@is.naist.jp (M. Kanbara), kato@is.naist.jp (H. Kato).

https://doi.org/10.1016/j.cag.2021.01.005
0097-8493/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

However, this technique can not be used for 3D object manipulation as it lacks the ability to control orientation. This is because SlidAR was developed for a text-based AR annotation application that does not need orientation control. To place 3D AR content, users must be able to manipulate position as well as orientation of the object.

Most HAR applications initially align objects with a world coordinate system. In general, the world coordinate system coincides with a fiducial marker, or is set via a vSLAM algorithm. In some cases, the orientation of the created object may not match the desired orientation and require manual adjustment by the user [10]. Users can manipulate the pose of the virtual content using buttons, gestures, or by adjusting the pose of the handheld device [11] itself. However, as mentioned previously, this is a difficult task and simple interaction techniques are needed to assist user. Bowman et al. [10], explained the benefit of using constraints in 3D user interfaces as they can simplify interaction while improving accuracy and user efficiency. Even though using constraints limits the amount of control the user can have, the simplification allows for a more intuitive interaction [12]. It can also reduce the number of DoFs that users have to control.

In this paper, we present SlidAR+, a method for controlling the rotation and position of virtual obtjecs in HAR applications. SlidAR+ extends SlidAR, with gravity-constrained rotation capabilities. This constraint is based on the observation that most common AR content will be placed relative to man-made objects in the environment that are mostly either horizontal or vertical. This observation has been previously utilized to improve the quality of surface tracking [13], interface design [14], and physically accurate rendering [15]. In SlidAR+, we first constrain the initialization of virtual objects to be either parallel or perpendicular to the gravity vector and then constrain one of the rotation axes to always align based on the gravity vector to allow for fast adjustments.

We conducted two experiments wherein we compared SlidAR+ with *Hybrid* [16], a state-of-the-art manipulation technique, to evaluate the usability and performance of SlidAR+ in 6DoF object manipulation and to investigate the effect of gravity-constrained rotation method on SlidAR+.

The main contributions of this paper are:

- We present a novel 6 DoF objection manipulation method for handheld devices, by integrating the 3D rotation control using gravity with the position technique of SlidAR.
- The insights gained from a user study that compares SlidAR+ with *Hybrid* to investigate the efficiency of using gravity information to pre-align and constrain the rotation of the virtual objects for 6 DoF tasks.

SlidAR+ does not require any special hardware and can be implemented in any handheld device with a build-in gyroscope or accelerometer that is capable of running AR applications. We believe that our experiment will be useful in the future development of in-situ virtual object placement in HAR application.

2. Related work

We divided this section into two main parts: (1) Overview of object manipulation, and (2) Object manipulation in HAR. The first part is to briefly discuss and summarize the overview of existing methods outside handheld devices. In the second part, we discuss existing methods in HAR as comparisons to the goal of this research work.

2.1. Overview of object manipulation

Object manipulation for virtual environments is already a thoroughly researched field dating back in the earlier days. Many researchers have proposed ideas to improve manipulating virtual objects on a desktop computer, during that time mostly utilize the mouse to directly manipulate the virtual objects [17–19]. Zeleznik et al. utilize two hands to control two independent cursors in object manipulation [20]. Conner et al. introduced the idea of manipulating 3D objects using widgets [21].

As virtual environments moved from desktop computer to head mounted display(HMD), it has opened many new possible applications for Augmented reality and Virtual Reality. However, due to the lack of physical input mechanics on the HMD itself, many researchers have also proposed many ideas to overcome that problem such as using a Mid-Air gesture in front of the camera [22]. Some famous examples are the one-handed Wireframe Cube technique and Chaconas et al. work of two hand gesture in HoloLens [23]. Some research used an external equipment such as the ChordGloves [24], HMD controller [25], tablet used for input and control [26], or a marker in table-top AR environment [27].

2.2. Object manipulation in HAR

Existing methods for object arrangement in HAR can be divided into: (1) methods that automatically align the virtual content with the physical world, and (2) methods that let users manually adjust the pose of the virtual content.

2.2.1. Automatic alignment of AR content

Automatic alignment methods extract features from the environment to adjust the pose of virtual content, without any input from the user. Many methods align content with fiducial markers that are placed in the scene. The pose of the virtual object can thus be adjusted by adjusting the pose of the marker. In recent years, more and more methods have been developed for marker-less tracking. These systems use the surfaces of the detected environment to constrain and align pose of the virtual content [13,28–30].

However, these methods depend heavily on the accuracy of surface detection and any errors can cause misalignment of the virtual object. In particular, if the target area has a complicated shape, image-based shape measurement techniques such as vSLAM may not be able to recover a surface accurately.

2.2.2. Manual alignment of AR content

To correct erroneous placement, users can manually adjust the pose of the virtual content. Over the past years, a large variety of techniques have been devised to simplify this process. In general, they can be sub-divided into four categories: button-based, screen-based, device-centric, and gesture-based.

Button-Based Manipulation

Button-based manipulation methods utilize physical and virtual buttons on the handheld devices to position and orient virtual objects. Herrysson et al. [11,31] used a smartphone's physical buttons to control each DoF parameter with a distinct button. In a similar manner, Castle et al. [32] used virtual buttons on the device's display. These methods require at least two buttons to control 1 DoF parameter (one to increase and another to decrease). In all, they require 12 buttons to control the position and orientation of a virtual object. Bai et al. [33] combine button and screen-based manipulation wherein users freeze a frame and select the DoF they would like to control with virtual buttons located at the edge of each translation and rotation axis. Afterward, the parameter can be adjusted via finger scrolling on the screen.

Button-based methods can change only one parameter at a time, which takes a lot of time, needs a large number of buttons, and is difficult to operate when controlling multiple DoFs simultaneously.

Screen-based manipulation

Screen-based manipulation methods utilize hand or finger gestures while interacting directly with the screen to manipulate the pose of the virtual object. Most screen-based manipulation techniques share the same basic idea of assigning one type of gesture (e.g., a vertical or horizontal slide) to control one DoF parameter. This allows users to control more than one parameter at the same time. With ARCBALL [19], users can adjust the rotation of the object by sliding with a single finger into the direction they want to rotate the object. Similar ideas have been explored to control 6 DoFs with two-finger gestures [34–38]. Martinet et al. [39] developed the z-technique to control the depth of a virtual object wherein while one finger is touching the virtual object, the second finger moves horizontally across the screen to move the object farther from or closer to the user. They later expanded their work to control all 6 DoFs and depth with the Depth-Separated Screen-Space (DS3) method [40]. They used different types of gestures to control the direction and orientation, thereby minimizing any mistakes due to similar gestures. The Shallow-Depth 3D technique developed by Hancock et al. [41] extends DS3 to three-finger gestures to control all 6 DoFs. However, increasing the number of fingers to be used for gestures also increases the cognitive load demand and complexity of the interaction.

Screen-based manipulation methods are accurate and do not require the user to move the device. Many studies [16,31,35] have also shown that these methods are most suitable for rotation and scaling tasks, as the user is able to control these parameters very accurately. However, as all parameters are controlled on the screen, the increasing number of gestures users have to learn and the number of fingers involved in each gesture affect the intuitiveness and ease of manipulation.

Device-centric manipulation

Device-centric movement methods utilize the movement of handheld devices to control the position and orientation of the virtual object. By adjusting the 6 DoFs pose of the device, users can adjust the position and rotation of the virtual object simultaneously. To prevent unintentional adjustments, users can trigger when to start the adjustment. Henrysson et al. [31] developed a grasping technique where the user can manipulate a virtual object by continuously pressing the screen or a button while moving the device [31]. Mossel et al. [35] support the manipulation of the virtual object by highlighting its axes as virtual depth cues. Marzo et al. [16] combined a device-centric movement grasping technique with screen-based gestures to control position and orientation, respectively, to improve the accuracy and speed of object manipulation.

Device-centric methods have been found to be the fastest among the 4 types of methods discussed here. However, as they use the movement of the device, it is difficult to control individual parameters accurately [11,31,35,42].

Gesture-based manipulation

Gesture-based manipulation methods use the device's camera to detect and track hand gestures performed in front of the camera to manipulate 6 DoFs of the virtual object.

Users can perform a variety of gestures, such as pushing, grabbing, or twisting [11,43–46] to manipulate the virtual object. Alternatively, these gestures can also be performed by other devices, like a pen [47].

Although users can control all 6 DoFs at the same time through gestures, these methods have been shown to be less effective than device-centric movement methods in practical scenarios [11,43,48].

2.3. Motivation

Most previous studies focused on the efficiency with which users can control all 6 DoFs. On the other hand, we designed SlidAR+ to minimize the number of parameters that users have to control in order to adjust the pose of a virtual object. For positioning, users can place an object in the scene by adjusting only 1 DoF with SlidAR. For orientation, we utilize gravity information to constrain the initial pose of the virtual object to the most probable orientation (either parallel or perpendicular to the gravity vector), and to provide users with an option to perform gravity-constrained rotation. As previously discussed, most of planar surfaces in man-made are aligned either parallel or perpendicular to the gravity vector (wall, table, pillar, etc.). We made an assumption that in a general AR content placement task, the virtual contents are also likely to be placed aligned with the direction of gravity (parallel or perpendicular); hence, users would have to adjust only 1 DoF.

In that sense, SlidAR+ combines the features of automatic alignment during the initial placement of AR object phase with the ability to pre-align the virtual object to the physical world constraint and manipulability (screen-based). By using gravity information to control the pre-alignment, SlidAR+, allows the user to have more control over the initial pose. This can help to avoid the misalignment caused by the tracking system, which is a common issue in SLAM-based applications. Normally, the virtual object is set to align to the system coordinate. However, in SLAM-based tracking, it is hard to predict how the system coordinate will be initialized, i.e., whether the coordinate system will be aligned with the real world or not. In our approach, the initial pose is now fixed based on the direction of gravity instead of the system coordinate. SlidAR+ required an IMU sensor that is available in most common smartphones nowadays. However, it does not require any additional real-world sensing (such as using computer vision technique) or special hardware (such as a depth camera or a 3D construction of the environment).

3. Design of SlidAR+

SlidAR+ extends the capability of SlidAR [1] by adding the ability to orient virtual objects in HAR applications. We followed the original motivation behind SlidAR and aimed to reduce the number of DoFs that users have to control while manipulating a virtual object, especially when users want to place objects parallel or perpendicular to the gravity vector.

For the experiments discuss in Sections 5 and 6, we implemented SlidAR+ in Unity3D[1] and used marker-based tracking from Vuforia SDK[2]. In a practical scenario, SlidAR+ can be used with marker-less tracking such as Apple ARKit[3] or Android ARCore[4] platforms. The current system simulates a task where the user has to place 3D AR models in the real environment and adjust their pose. We divide this task into two phases (Fig. 1): (1) initialization phase and (2) adjustment phase. The system also provides a set of 3D assets with an axis aligned to gravity.

3.1. Initial placement of AR object

In the initial placement phase, a user first selects the desired type of annotation. After that, the user selects the alignment-type to determine whether the object should be parallel or perpendicular to the direction of gravity and presses the desired location on the screen of the handheld device. The system places the virtual

[1] https://unity.com/.

[2] https://developer.vuforia.com/.

[3] https://developer.apple.com/augmented-reality/.

[4] https://developers.google.com/ar.

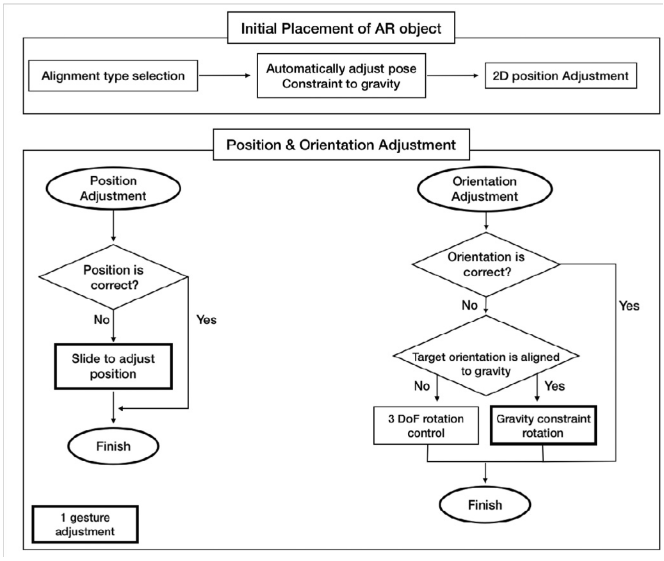


Fig. 1. Two-phase workflow of SlidAR+ is divided into two phases: (1) In the initialization phase, users align the selected object with the direction of gravity. (2) During the adjustment phase, the position can be adjusted with SlidAR and the orientation can be corrected using gravity-constrained orientation control.

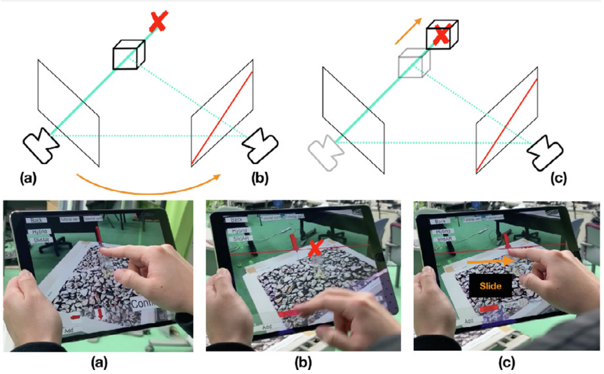


Fig. 2. Positioning with SlidAR. (a) Upon creation of a virtual object, the user inputs the 2D initial position. (b) After moving to another viewpoint, the position of the virtual object is misaligned because the depth information cannot be input during object creation. (c) The user corrects the position by sliding the object along the red epipolar line with a slide gesture.

object at a predefined depth (fix distance from the front camera instead of using vSLAM to remove the effect of inaccuracy of vSLAM on the experimental results). Then aligns it with the desired direction according to the gravity information from the sensors built into the handheld device. During this phase, the user can adjust the position by pressing or moving their finger on the screen. This phase lasts until the user taps the ”Confirm” button to finished the initial phase.

3.2. Positioning using SlidAR

SlidAR utilizes ray casting and epipolar geometry to adjust the position of virtual objects (Fig. 2(a–c)). This process of SlidAR can be divided into 2 steps: (1) setting the 2D position and (2) adjusting the depth information. The 2D positioning process is performed during object initialization phase. When the user presses the desired location, the object appears correctly aligned with the target position from the current perspective. When the user presses the ”Confirm” button, the system casts a ray from the current camera pose to the created object to create the epipolar geometry.

When the user views the scene from a different viewpoint, the epipolar line is rendered as a 2D red line and is used to represent the depth information. The user can adjust the position of the annotation (depth information) by moving it along the epipolar line with a one-finger slide gesture. During the slide gesture, the user does not have to match the finger position with that of the virtual object.

However, if the 2D position is incorrect, the user can not use SlidAR to position the annotation to the desired location because the epipolar line does not intersect with it. In this case, the user can used the two-finger gesture to re-adjusting the 2D position by pressing on the desired location and then releasing the finger to recreate the epipolar geometry.

3.3. Orientation control

The main idea behind SlidAR+ is to use the gravity information to assist in orientation control. We use the gravity information to assist in 2 processes: 1) setting up the initial orientation, and 2) allowing the user to perform gravity constrained rotation.

During the initial placement process, the system automatically aligns virtual objects with the gravity vector based on user’s selection, thereby effectively pre-determining 2 DoFs. The last DoF is the rotation around the gravity vector. The user can then adjust the remaining DoF (rotation around the gravity vector) with a one-finger horizontal sliding gesture (Fig. 3a).

Users can also manipulate all 3 rotational DoFs, if necessary, by using the two-finger vertical and horizontal sliding gestures for ARCBALL [19] rotation and a two-finger twist gesture [16] to rotate around the device’s *x*-, *y*- and *z*-axis, respectively (Fig. 3(b–d)).

4. Evaluation of SlidAR+

To evaluate the efficiency of SlidAR+ in 6 DoFs object manipulation tasks, we performed a user study to compare SlidAR+ with a state-of-the-art method, *Hybrid*. Marzo et al. [16] found that *Hybrid* is the most efficient object manipulation method compared to other device-centric and screen-based methods.

4.1. Methodology

Hybrid [16] combines device-centric movement and screen-based manipulation techniques. It uses the device-centric movement to control an object’s position and screen-based gestures to control orientation. This allows the user to control all 6 DoFs at the same time without requiring the need to switch between the two methods.

We found Mazo et al.s *Hybrid* suit our requirement very well as our task is to focus on manipulating 6 DoFs using tablet or 2D input devices.Hybrid efficiently controls the position of the object using the device movement which means users have to input only 3 DoFs for orientation through the 2D display. In their work, they compare *Hybrid* with device-centric movement and screen-based technique in 6 DoFs tasks. The results show that Hybrid performs the best among the three methods[16].

However, *Hybrid* did not include object placement and creation functionality. Therefore, we have added this capability in our experiments. *Hybrid*’s workflow can also be divided into two 2 phases: 1) initial placement of the AR object, and 2) object manipulation.

4.1.1. Initial Placement of AR object

This is the same as in SlidAR+, as described in Section 3.1. The only difference is that instead of aligning the AR object to the gravity vector, *Hybrid* aligns it to the system coordinate.

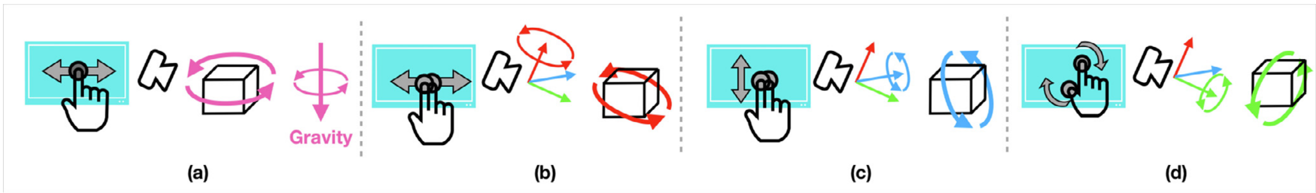


Fig. 3. Orientation control in SlidAR+: (a) One-finger horizontal slide gesture to perform gravity constrained rotation. (b) Two-finger horizontal and (c) vertical slide gesture to rotate around the camera's x- and y- axes. (d) Two-finger twist gesture to rotate the object around the camera's *z*-axis.

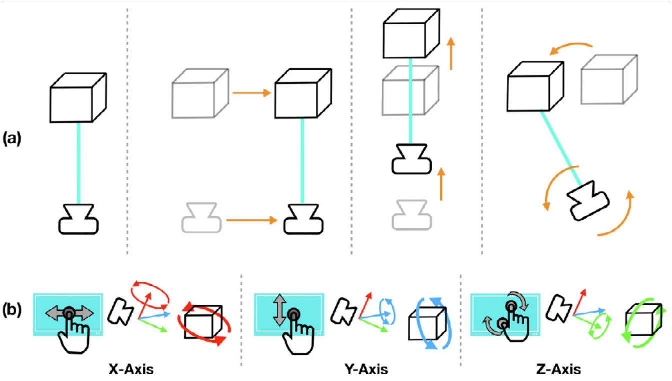


Fig. 4. Object manipulation control in *Hybrid*: (a) Device centric movement position control (blue line indicates the fix distance between object and camera). (b) Orientation control in Hybrid: Horizontal and vertical slide gesture to rotate around the camera's *x*- and *y*- axes. Two-finger twist gesture to rotate the object around the camera's *z*-axis.

4.1.2. Object manipulation: positioning and orientation

In *Hybrid*, a user can control both position and orientation at the same time without needing to switch modes, as in SlidAR+. To manipulate the object, the user first aligns the center of the screen with the intended object and taps with a finger anywhere on the screen. This fixes the object in the device's coordinate system. Now its position can be manipulated by moving the device (Fig. 4a).

Orientation in *Hybrid* is controlled by combining ARCBALL and the two-finger twist gesture (Z-Rot). Both techniques rotate a virtual object relative to the camera axis. In ARCBALL, the user performs vertical and horizontal sliding gestures along the screen to rotate around the x- and y-axes while the two-finger twist gesture rotates the object around the *z*-axis (Fig. 4b).

4.2. Experiments design

There are two main points we want to evaluate in this study: 1) the overall efficiency of SlidAR+ in 6 DoFs manipulation task, and 2) the affect on performance when the gravity control feature is added to SlidAR.

We therefore conducted two experiments: 1) a "positioning task" and 2) a "6 DoFs task". The first experiment will be used to confirm the performance of SlidAR as it has been implemented by us and evaluate it under condition that have not been covered in the previous study, as described in Section 4.2.1. The second experiment is the evaluates of SlidAR+ in a task that involves manipulating all 6 DoFs (positioning + orientation) tasks. By performing both the experiments, we can compare the change in performance of SlidAR+ before and after adding orientation control to see the affect of our orientation control feature.

4.2.1. Depth cue using shadow

One of the problems in object manipulating AR object with handheld devices is the lack of depth information as most of devices use a monoscopic display. To solve this problem, researchers have used the shadow cast by the virtual object onto the planar or ground surface to aid object placement.

When placing any AR content in the real-world scene, we can divided into two scenarios: 1) AR content is placed or attached on a real objects on the planar surface, or 2) AR content is placed on the mid-air or on a non-planar surface; for an example placing on an object protruding out of the wall. The difference between these two is the difficulty in using shadows to get depth cue information. In the first case, a user can easily get the depth information using the real-world object as a reference by directly matching the shadow with the base of the object on the planar surface. However, in the second case, the shadow might be projected on a difference surface, make it more difficult to to place the object correctly.

In the original SlidAR experiment [1], the effect of shadows could only be partially understood from the subjective user feedback. However, we would like to objectively confirm and verify the effect of shadows and depth cue on the performance of SlidAR+.

We therefore have performed the experiment under two conditions with different level of depth cue difficulties: 1) an "easy condition", and 2) a "hard condition". In the easy condition, the AR content is always placed on an object on the planar surface. In hard condition, the AR content is always placed mid-air or a non-planar surface.

4.2.2. Orientation related condition

In this part, we would like to explain the main factors that affect the orientation in our experiment. From our discussion in Section 2.3, our condition are related to the relationship between factor that affect the orientation and the gravity direction.

Target pose

Target pose is the position and orientation of the AR content that the user wants it to be. In our study we refer to it as the pose that the participants have to manipulate the AR content to match. Normally, target pose is aligned to a surface of the physical world environment. As described in Section 2.3, we also assume that the target pose is aligned to the gravity vector in most cases. In summary, we have two conditions: 1) target pose is aligned (parallel or perpendicular) to the gravity, and 2) target pose is not aligned to the gravity.

System coordinate

System coordinate refers to the coordinate created by a SLAM based AR application for the initial localization as discussed in Section 1, this coordinate is defined by detecting the physical world (mostly planar surface) and is used to determine the initial orientation of AR object. Normally, we would want the initial orientation to be as close to the target pose as possible or at least aligned to the same coordinate so as to reduce the number of angles or DoFs that may need to be adjusted. However, in some cases the system coordinate is not aligned properly, such as in case of an error during SLAM system initialization or if the planar surface is oblique to the alignment of target pose. This will increase the

Table 1
Summary of hypothesis: (a) Translation-related condition. (b) Orientation-related condition.

Condition		
Easy	Similar performance	
Hard	SlidAR+ is better	
	(a)	
	Target pose aligned to gravity	Target pose NOT aligned to gravity
System coordinate aligned to gravity	SlidAR+ is better	Similar performance
System coordinate NOT aligned to gravity	SlidAR+ is better	Similar performance
	(b)	

Table 2
Overall hypothesis.

Condition		Target pose aligned to gravity	Target pose not aligned to gravity
System coordinate aligned to gravity	Easy	SlidAR+ is better	Similar performance
	Hard	SlidAR+ is better	SlidAR+ is better
System coordinate not aligned to gravity	Easy	SlidAR+ is better	Similar performance
	Hard	SlidAR+ is better	SlidAR+ is better

amount of time and effort needed to adjust the pose. In summary, we have 2 conditions: 1) the system coordinate is aligned to the gravity, and 2) system coordinate is not aligned to the gravity.

4.3. Hypotheses

The hypotheses we are evaluating were created based on the condition affecting the 6 DoFs, which can be divided into two groups: 1) position and 2) orientation.

- Translation-related condition In the easy condition, both SlidAR+ and *Hybrid* should have similar task completion times because with a depth cue *Hybrid* should be able to perform as fast as SlidAR+. In the hard condition, SlidAR+ should be able to complete the task faster than *Hybrid* because *Hybrid* requires the shadows for depth information. When shadows become difficult to observe, participants would have to spend more time adjusting depth and position. However, this will not affect SlidAR+ as it does not rely on shadows to obtain the depth information (Table 1 (a)).
- Orientation related condition In case of both system coordinate and target pose are aligned to the gravity, SlidAR+ should show a faster completion time than *Hybrid* due to the gravity constrainted rotation feature of SlidAR+, which will allow users to complete the task faster than *Hybrid*'s ARCBall, even though both methods will have the same number of DoFs and angles to adjust. In the case where the target pose is aligned to gravity but system coordinate is not, we expect *Hybrid* to perform slower than system coordinate is aligned since the system coordinate will affect the initial pose in *Hybrid* and cause extra angle and DoFs needed to adjust. But this system coordinate will not affect SlidAR+ as it fix the initial pose to aligned to the gravity.
  In the condition where target pose not aligned to the gravity vector, both methods should show similar completion times as SlidAR+ has no direct advantage over *Hybrid* in such a case (Table 1 (b)).

From the above, we have 3 main hypotheses (Table 2):

- **H1**: SlidAR+ will perform better than *Hybrid* when target pose is aligned to the gravity vector.
- **H2**: SlidAR+ will perform better than *Hybrid* under the hard condition when the target pose is not aligned to the gravity vector.
- **H3**: SlidAR+ and *Hybrid* will have a similar performance under easy condition when the target pose is not aligned to the gravity vector.

5. Experiment 1: positioning task

In Polvi et al.'s [1] experiment, the user had to place the AR content on a Lego structure. In this experiment, we wanted like to explore and evaluated the performance of SlidAR in the case where the mid-air object placement is needed.

We conducted this experiment following the same design as Polvi et al.'s experiment [1]. We measured efficiency on the basis of two aspects: 1) the average time taken to complete the task and 2) the average distance of device movement. We used a screen recorded to study the participants' behavior during the experiment.

5.1. Experiment design

Our experiment simulated a task support scenario, where participants were asked to place 3D annotations in the scene. We conducted the experiment in a laboratory environment with marker-based tracking for better control over all variables. This experiment used a within-subject design with two independent variables: 1) object manipulation methods, and 2) difficulty.

5.1.1. Independent variables

- **Object manipulation method** We compared two object manipulation methods in this experiment: SlidAR and *Hybrid*. Both methods provide to create a 3D arrow annotations from the 3D assets provided by the system. All participants utilized both methods in a counterbalanced order.
- **Difficulty** This variable is used to described the difficulty in viewing and using shadow or depth cue while positioning the object. As described previously in Section 4.2.1, we defined two conditions based on difficulty: an "easy condition" and a "hard condition". In the easy condition, all of AR targets were placed on top of virtual pillars (height: 4–14 cm) that connect to the ground. Participants could adjust the position easily by matching the shadow of the AR object with the virtual pillar's base. Under hard condition, AR targets were placed on the top of floating pillars (height: 10 to 25 cm from ground). In this case, participants could not easily use the shadows to guess the position (Fig. 5).

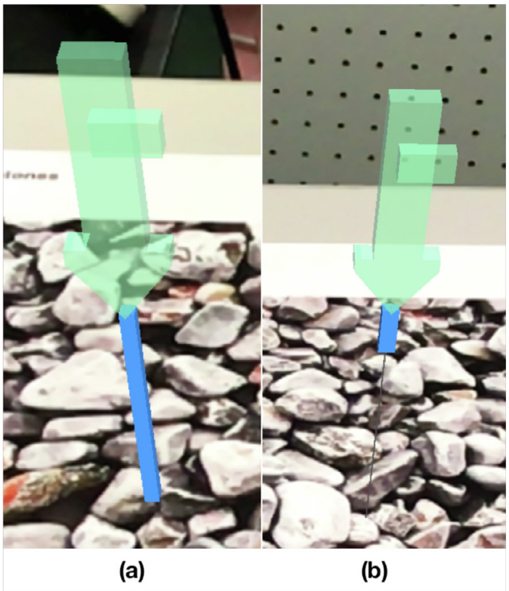


Fig. 5. Picture of difficulty setup: (a) a long AR pillar which connect to the ground. (b) A floating AR pillar (Black line illustrate the high of pillar which is not show during experiment). Small green rectangle on arrow used to represent the direction of the arrow.

5.1.2. Experiment platform and setup

The handheld device used in this experiment was an Apple iPad Pro(2017) with a 1668 × 2224 pixels 10.5 inch display, Apple A10X CPU, and a weight of 477 g. The reason why we choose a tablet is because one of our goals is to apply this interface to an industrial AR application. In an industrial AR, the tablet is more preferable as it can provide more information on the larger screen at the same time than a smartphone. The system was usable only in portrait orientation with the back camera in the top-left corner. Furthermore, we used an AR marker for tracking the device pose and for defining the system coordinate in the AR application.

All tasks were conducted with the same setup, i.e., AR marker (80 × 60 cm) placed on a table (length = 80cm, width = 80cm, and height = 70cm). Participants were encouraged to walk and look around the table and tasks area from different angles and viewpoints. This setup were used in both the experiments.

5.2. Hypotheses

We have three hypotheses for this first experiment.

- **E1-H1**: SlidAR and *Hybrid* should have similar completion times under easy condition.
- **E1-H2**: SlidAR will be faster than *Hybrid* under the hard condition.
- **E1-H3**: SlidAR will require less device movement than *Hybrid*.

Hypotheses **E1-H1** and **E1-H2** are based on the hypotheses shown in Table 1 (a). In **E1-H3**, we expected SlidAR to utilize smaller device movements to accomplish the tasks as participants using *Hybrid* would have to move the device in order to positioning the object, whereas SlidAR participants only need to move the device once to the change viewpoint before they begin positioning.

5.3. Experiment tasks

The participants were asked to create and place an AR arrow (represented as a red AR arrow similar to the one shown in Fig. 2b) and move it to the correct position. Out of all the virtual objects, only the created AR arrow cast a computer-generated shadow. Participants can receive a depth cue by trying to match the shadow from the created AR arrow with the base of the virtual pillar. Each participant completed two tasks per method (four tasks per participant in total). Each task consisted of five trials or five target AR arrows that were highlighted as a translucent green AR arrow presented one at a time. To completed each trial, participants had to create an AR arrow and align its position with the shown target. The system automatically checked the alignment of the user created arrow with target AR arrow by comparing their positions. The task was completed if the difference between the current object position and target position was within a set margin (2 cm) for 1 second. When one trial was completed, both arrows disappeared and the next trial was started. Participants received a notification once they completed all five trials. In all, participants had to align 20 target annotations.

Table 3
HARUS Questions.

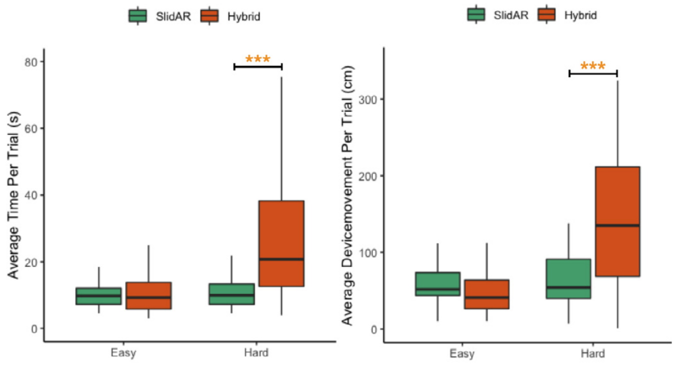
Manipulability:
Q1. I think that interacting with this application requires a lot of body muscle effort.
Q2. I felt that using the application was comfortable for my arms and hands.
Q3. I found the device difficult to hold while operating the application.
Q4. I found it easy to input information through the application.
Q5. I felt that my arm or hand became tired after using the application.
Q6. I think the application is easy to control.
Q7. I felt that I was losing grip and dropping the device at some point.
Q8. I think the operation of this application is simple and uncomplicated.
Comprehensibility:
Q9. I think that interacting with this application requires a lot of mental effort.
Q10. I thought the amount of information displayed on screen was appropriate.
Q11. I thought that the information displayed on screen was difficult to read.
Q12. I felt that the information display was responding fast enough.
Q13. I thought that the information displayed on screen was confusing.
Q14. I thought the words and symbols on screen were easy to read.
Q15. I felt that the display was flickering too much.
Q16. I thought that the information displayed on screen was consistent.

5.4. Experimental procedure

The experiment took approximately 40–60 min to complete per participant. First, each participant was tutored for up to 10 minutes (explanation and practice level) on the first method (depending on the order of each participants). We instructed participants on the following steps: (1) how to create an arrow, (2) how to adjust and correct the position, and (3) a way to use each method effectively. Participants are free to grasp and hold the device as they are favored and comforted.

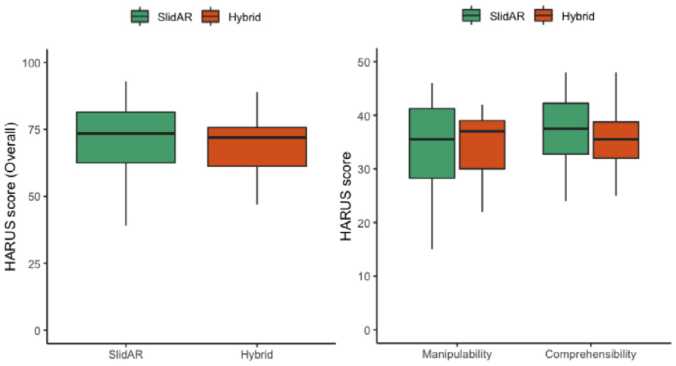
Next, each participant spent approximately 5–10 min (depending on individual skill) to complete all tasks using the first method (approximately 2–5 min per task). After the experiment, the participant was asked about their opinion of the method, and their answeres were recorded on a HARUS (Handheld Augmented Reality Usability Scale) questionnaires [49] (Table 3), that recorded subjective feedback in two aspect: manipulability and comprehensibility, and free-form written comments. This process took approximately 5 minutes, followed by a small break. Then, the tutorial for the second method started with the same procedure as the first.

All measurement data were captured automatically by the system. And we screen-recorded all of the operation data for each trial. As for the time data, we divided it into two parts: (1) Authoring time: the time participant spent in adjusting the initial 2D position. This was recorded after participant created an AR arrow until the end of initial phase. (2) Editing time: the time participant spent adjusting an AR arrow to the correct position. This was auto-



(a) Average completion time per trial. (b) Average device movement per trial.

Fig. 6. Result of objective measurements: (a) task completion time and (b) device movement. Connected bar represents significant difference (* = significant at 0.05 level, *** = significant at 0.001 level).



(a) HARUS total score. (b) HARUS individual score.

Fig. 7. Result from HARUS questionnaire: (a) total score and (b) manipulability and comprehensibility score.

matically recorded at the time between the end of initial phase until the trial was completed. We also measured the device's movements based on the relative position of the device's camera to the marker. Every 30 frames, the trajectory between the current and previous poses was added to the total device movement.

5.5. Participants

We recruited a total of 12 participants from the local university (8 males and 4 females; average age: 24 years (SD = 1.4); range: 23-27 years). We asked the participants about their experiences with AR applications and 8 participants reported having using an AR application previously whereas 4 had never used any AR application before. We also asked the participants about their pre-existing knowledges in 3D manipulation: 5 participants were familiar, 2 had moderate knowledge, and 5 had no experience.

5.6. Results

For our analysis, we first ran a normality test on the data. The results from Shapiro-Wilk test showed that the data violated normality ($p < 0.05$). So, we used non-parametric Wilcoxon test for the analysis of the data.

We noticed that SlidAR (Mdn = 9.94) completed tasks significantly faster than *Hybrid* (Mdn = 21.71) under the hard condition, $z = 5.072,\ p < 0.001,\ r = 0.655$. However, we found no significant difference between SlidAR (Mdn = 9.76) and *Hybrid* (Mdn = 9.23) under the easy condition, $z = 0.611,\ p = 0.54,\ r = 0.078$ (Fig. 6a). Next, we investigated the experiment results under the easy condition and found that SlidAR (Mdn = 2.92) required significantly more time for 2D positioning than *Hybrid* (Mdn = 0.1), $z = 6.14,\ p < 0.001,\ r = 0.793$. However, SlidAR (Mdn = 6.22) showed a significantly less time in editing mode than *Hybrid* (Mdn = 8.91), $z = 2.88,\ p = 0.003,\ r = 0.372$.

For the device movement, we can not find any significant difference between SlidAR+ (Mdn = 51.72) and *Hybrid* (Mdn = 41.035) in easy condition, $z = 1.899,\ p < 0.057,\ r = 0.245$. However, SlidAR (Mdn = 54.31) recorded significantly smaller device movements than *Hybrid* (Mdn = 179.92), under the hard condition, $z = 5.33,\ p < 0.001,\ r = 0.688$) (Fig. 6b).

Upon analyzing the subjective feedback, we could not find any significant difference in the total score, manipulability, or comprehensibility between SlidAR and *Hybrid* (Fig. 7). In the free-form written feedback, 7 participants preferred SlidAR and 5 preferred *Hybrid*.

5.7. Discussion

Under the easy condition, there was no significant difference in completion time between SlidAR and *Hybrid*, which support the **E1-H1**. This happened because most participants spent a lot of time in positioning the arrow in SlidAR whereas in *Hybrid* they could just tap on screen instantly and adjust the position after that. However, SlidAR performed better in the editing mode(3D position) as SlidAR can correct an object's position just via a single slide gesture whereas *Hybrid* requires adjustment of all 3 DoFs.

Results under the hard condition were clearer and we could see that SlidAR performed better and required less device movements which supports **E1-H2**. A plausible reason of this is that SlidAR does not require any additional depth information whereas participants take a longer time to move and check the current position or guess by looking at the shadows while using *Hybrid*.

Our results show that SlidAR was significantly faster than *Hybrid* only under the hard condition and not in easy condition, which negates the **E1-H3**. One of the reason for this is that under the easy condition, participants can match the shadow with the base of pillars in order to get depth information instead of moving to other viewpoints when using *Hybrid*. Whereas under the hard condition, participants need to move to other viewpoints to get depth information and this process may need to repeated many times during one trial. However, in SlidAR participants have to move only once.

Upon analyzing the subjective feedback, we did not find any significant differences between SlidAR and *Hybrid* in term of any category of HARUS questionnaire, which was different from the results of the Polvi et al.'s study [1]. We believe this might happen due to the duration of task, weight and size of device causing a negative effect on SlidAR more than in *Hybrid*. There were 2 main reasons why 5 participants of the 12 participants still preferred *Hybrid* over SlidAR. First, even through many participants comment that SlidAR is easier to use, *Hybrid* offered more freedom of control and they felt more engaged and entertained when using it. The second reason was related to the difficulty and fatigue induced when holding the device. From observation all participants in each method have the following manner. During the initial placement process, both methods utilized a single hand holding gesture where participants had to use one hand to hold the device and the index finger of the other hand to interact with the device. Participants spent less time in this process in *Hybrid* as it does not require an accurate 2D position as SlidAR. For the process of moving and changing the viewpoint to observe and check the correctness, participants were likely to hold the device with two hand gestures as it was more comfortable in both methods.

For the position adjustment, SlidAR used the single hand gesture with the other hand as support to position the AR content. While in *Hybrid*, participants could operate by holding the device with two hands while using their thumbs to interact with the screen. As the single hand holding causes more physical fatigue on big and larger devices than with two hands holding, resulting in more physical fatigue in SlidAR than *Hybrid*. However, many participants felt that the tasks were too short to feel any difference in terms of fatigue between both methods, but this might change if the task become longer. We also received comments about using smaller devices such as mobile phone, rather than a tablet.

Overall, we found that the results supported **E1-H1** and **E1-H2** but not **E1-H3**.

6. Experiment 2: 6 DoFs task

From the previous experiment, we found that SlidAR is significantly faster than *Hybrid* under the hard condition but both have similar performance under the easy condition. However, the task in this experiment, *Hybrid* has an advantage in that it allows control over both position and rotation at the same time, while SlidAR+ controls them separately.

6.1. Experiment design

The basic design of this experiment is similar to the previous experiment. Participants had to create an AR arrow, place it into the scene, and adjusting its pose to match a target (position and orientation). We used the same platform and equipment setup as in the previous experiment, except with an additional variable that effects orientation. In this experiment we used a within-subject design with three independent variables: object manipulation methods, target pose, and coordinate system as we described in Section 4.2. For evaluation, we measure same efficiency in terms of time, device movement, and subjective feedback just as in the previous experiment.

Independent variables

- **Target pose** This variable describes the alignment of the target pose relative to the direction of gravity: 1) target pose is parallel or perpendicular to the direction of gravity vector or 2) target pose is not parallel or perpendicular to the direction of gravity vector.
- **The system coordinate** This variable is the relationship between the the system coordinate of the AR application and the gravity direction as we described in Section 4.2.2. There are two possibilities for this variable are as follows: 1) The system coordinate of the AR application is aligned to the gravity vector, or 2) The system coordinate of the AR application is not aligned to the gravity vector.
- **Object manipulation methods** This variable is the method participants used to perform the task. We have 2 methods in this experiment, SlidAR+, and *Hybrid*.

The dependent variables are objective results consisting of task completion time (seconds) and device movement distance (cm). We also collected subjective feedback following the experiment using HARUS [49] and free-form written feedback.

Experimental conditions

In the experiment, we have four conditions (target pose × coordinate system) for each manipulation method.

Condition 1: The system coordinate is aligned to the gravity vector and the target annotation is either parallel or perpendicular to gravity.

Condition 2: The system coordinate is aligned to the gravity vector and the target pose is not parallel or perpendicular to gravity.

Condition 3: The system coordinate is not aligned to the gravity vector and the target pose is either parallel or perpendicular to gravity.

Condition 4: The system coordinate is not aligned to the gravity vector and the target pose is not parallel or perpendicular to gravity.

6.2. Hypotheses

On the basis of the design of SlidAR+ and *Hybrid*, we had three hypotheses for this experiment as discussed in Section 4.3:

- **H1**: SlidAR+ will perform better than *Hybrid* when target pose aligns to the gravity vector.
- **H2**: SlidAR+ will perform better than *Hybrid* when the target pose does not align to the gravity vector under the hard condition.
- **H3**: Both SlidAR+ and *Hybrod* will perform similarly when the target pose is not aligned to the gravity vector under the easy condition.

6.3. Experiment tasks

As previously mentioned, the task involved in this experiment were similar to experiment 1, which include placing and adjusting an AR arrow to the target pose (position and rotation). Each participant had to completed four tasks per method (total eight tasks per participant). And each task consisted of six trials or six target arrows, with the first three target arrows were set on long pillars (”easy trial”) and the last three targets were set on floating pillars (”hard trial”), as in the previous experiment. In order to complete the task, participants had to complete all trials by aligning the created arrow with the target one. The system determined whether both the user created and target arrows are aligned or not. The setting margin for the correctness of pose in this experiment were set at 2 cm for translation and 12° for orientation. In all, participants had to align 48 target arrows.

6.4. Experimental procedure

We divided the experiment into two sections, one per each manipulation method. Each section took approximately 50 to 60 minutes per participant. First, a participant spent time up to 20 min being tutored (explanation, presentation, and practice level) on the first method (depend on the order of each participants). We instructed participants on the following steps: (1) how to create an arrow, (2) how to adjust, and correct the position, and (3) a way to use each method effectively. Again, we did not specify the way to grasp and hold the device, participants can operate as they wish.

Next, the participant spent approximately 20–50 min (depending on individual skill) completing all tasks using the first method (approximately 5–10 min per task) followed by a 5 min break between the tasks. After the experiment, the participants were asked about their opinion of the method; they also answered an questionnaires that recorded their subjective feedback and free-form written comments. Next, the participant took a small break. This process took approximately 15–20 min and then the whole process was repeated for second section.

All data were measure automatically by the system, and we screen-recorded all of the operation data for each trial. We divided time data into 2 parts: (1) Authoring time: time spent by participant in adjusting the 2D initial position. This was recorded after a participant created an AR arrow until the end of initial phase. (2)

Editing time: time spent by a participant adjusting an AR arrow to the correct position. This was automatically recorded as the time between the end of initial phase until the trial was completed. We also measured the device movement based on the position of the device's camera relative to the marker. Every 30 frames, the trajectory between current and previous pose was added into total device movement. Finally, we collected subjective feedback after the experiment using HARUS [49] and free-form written feedback.

6.5. Participants

We recruited a total of 16 participants for this experiment (11 males and 5 females; average age: 24 years (SD = 1.4); range: 23-27 years) We asked the participants about their experience with AR application and 13 participants reported having used an AR application previously but 3 had never used an AR application before. We also asked the participants about any pre-existing knowledge in 3D manipulation: 10 participants were familiar, 1 had moderate knowledge, and 5 had no experience. 10 participants out of the 16 participants had participated in the previous experiment as well, however, there was at least a 1 day break between each experiment.

6.6. Results

Shapiro-Wilk test showed that the data in this experiment 2 also violated normality ($p < 0.05$). Therefore, we used non-parametric Wilcoxon test to analyze the data. We consider results significant for $p < 0.05$.

Task efficiency

In the overall completion time (easy + hard trials), we found the significant differences in Condition 1 and 3 wherein SlidAR+ completed the tasks faster than *Hybrid* (Con1: Mdn = SlidAR+ **(S)** 16.63(s) vs *Hybrid* **(H)** 44.45,$z = 7.32$, $p < 0.001$, $r = 0.74$; Con3: Mdn = **(S)** 16.84 vs **(H)**35.29,$z = 6.05$, $p < 0.001$, $r = 0.61$).

For a detailed analysis of the average completion time data, we focused only on the performance between SlidAR+ and *Hybrid* in the same condition with the same trials settings, and we did not compared the data between difference condition or trials. We do not report the results of authoring and editing time as the time spent in authoring mode is very small and similar to experiment 1. The results of editing time are also similar to the overall task completion time.

In easy trials, SlidAR+ performed significantly faster than *Hybrid* when the targets were aligned to the gravity vector (Con1: Mdn = **(S)** 16.1 vs **(H)** 32.52, $z = 7.32$, $p < 0.001$, $r = 0.74$; Con3: Mdn = **(S)** 17.04 vs **(H)** 25.81, $z = 6.05$, $p = 0.009$, $r = 0.61$). However, in Conditions 2 and 4, SlidAR+'s performance was significantly slower than *Hybrid* (Con2: Mdn = **(S)** 42.89 vs **(H)** 31.63, $z = 2.3$, $p = 0.02$, $r = 0.03$; Con4: Mdn = **(S)** 47.33 vs **(H)** 28.84 $z = 3.81$, $p < 0.001$, $r = 0.015$). As for the hard trials, SlidAR+ completed tasks significantly faster than *Hybrid* in the Conditions 1, 3, and 4 (Con1: Mdn = **(S)** 17.65 vs **(H)** 57.52, $z = 5.52$, $p < 0.001$, $r = 0.79$; Con3: Mdn = **(S)** 16.45 vs **(H)** 53.19, $z = 5.48$, $p < 0.001$, $r = 0.79$, Con4: Mdn = **(S)** 3.45 vs **(H)** 51.57, $z = 2.64$, $p = 0.008$, $r = 0.38$). However, we could not find any significant difference between SlidAR+ and *Hybrid* under Condition 2 (Mdn = **(S)** 42.98 vs **(H)** 53.19, $z = 1.6$, $p = 0.109$, $r = 0.23$) (Fig. 8).

Analyzing the device movement results, we found that SlidAR+ required significantly less movement than *Hybrid* in the Conditions 1,3 and 4 (Con1: Mdn = **(S)** 89.08(cm) vs **(H)** 303.39, $z = 7.21$, $p < 0.001$, $r = 0.73$; Con3: Mdn = **(S)** 81.1 vs **(H)** 239.8, $z = 6.57$, $p < 0.001$, $r = 0.67$; Con4: Mdn = **(S)** 209.08 vs **(H)** 251.25, $z = 2.31$, $p = 0.02$, $r = 0.23$). But this was not the case with Condition 2 (Mdn = **(S)** 265.91 vs **(H)** 231, $z = 0.22$, $p = 0.8$, $r = 0.022$).

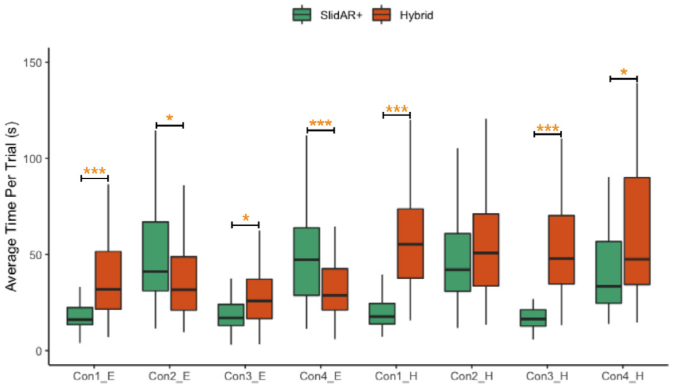


Fig. 8. Result of objective measurements, an average task completion time. Connected bar represents significant difference (* = significant at 0.05 level, *** = significant at 0.001 level).

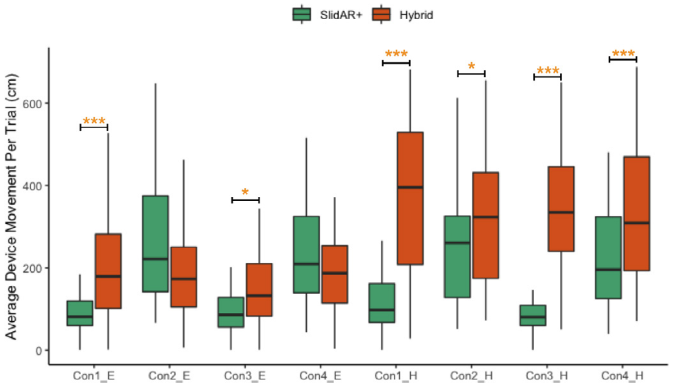


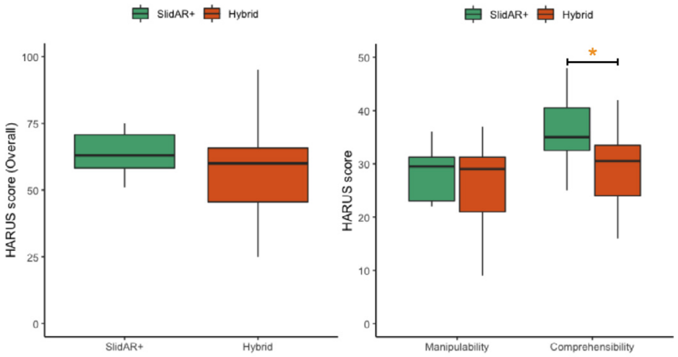
Fig. 9. Result of objective measurements, an average Device Movement. Connected bar represents significant difference (* = significant at 0.05 level, *** = significant at 0.001 level).

When analyzed the data in more detail, we found that under the easy trials, there was a significant difference between SlidAR+ and *Hybrid* wherein SlidAR+ required smaller device movement than *Hybrid* under Conditions 1 and 3 (Con1: Mdn = **(S)** 81.29 vs **(H)** 189.12, $z = 4.27$, $p < 0.001$, $r = 0.61$; Con3: Mdn = **(S)** 85.93 vs **(H)** 132.57, $z = 2.08$, $p = 0.03$, $r = 0.3$) but not under Condition 2 and 4 (Con2: Mdn = **(S)** 244.81 vs **(H)** 173.36, $z = 1.92$, $p = 0.053$, $r = 0.27$; Con4: Mdn = **(S)** 213.31 vs **(H)** 186.84, $z = 1.63$, $p = 0.1$, $r = 0.23$). Under the hard trials, SlidAR+ required significantly smaller device movement than *Hybrid* in every condition (Con1: Mdn = **(S)** 97.97 vs **(H)** 444.39, $z = 5.65$, $p < 0.001$, $r = 0.81$; Con2: Mdn = **(S)** 272.34 vs **(H)** 378.53, $z = 2.03$, $p = 0.04$, $r = 0.29$; Con3: Mdn = **(S)** 80.43 vs **(H)** 372.97, $z = 5.93$, $p < 0.001$, $r = 0.85$, Con4: Mdn = **(S)** 202.7 vs **(H)** 415.42, $z = 3.89$, $p < 0.001$, $r = 0.56$) (Fig. 9).

Subjective feedback

For subjective feedback, we measured user preference through the HARUS scale. We ran paired Wilcoxon signed rank test to analyze the data. Analyzing the results, we could not find any significant difference between score of SlidAR+ and *Hybrid* on the overall (Fig. 10a) and manipulability. However, SlidAR+ (Mdn = 34) was scored significantly higher than *Hybrid* (Mdn = 31) in comprehensibility ($z = 2.833$, $p = 0.004$, $r = 0.731$) (Fig. 10b). An in-depth analysis of each questions revealed a significant difference with $p < 0.05$ on Q4, Q9, Q11, and Q13.

In the free-form written feedback, 13 participants preferred SlidAR+ whereas only 3 preferred *Hybrid*. Many comments were similar to the previous experiment, wherein participants felt that



(a) HARUS total score. (b) HARUS individual score.

Fig. 10. Result of HARUS questionnaire: (a) total score and (b) manipulability and comprehensibility score. Connected bar represents significant difference (* = significant at 0.05 level, *** = significant at 0.001 level).

SlidAR+ was easier to use and understand. They could easily correct the position using fewer inputs by using SlidAR, and they did not have to move a lot. Many participants reported that they felt more engaged, entertained and had more freedom to control while using *Hybrid*. They also reported about the difficulty in holding the device while carrying out the tasks using SlidAR+, something that they did not experience with *Hybrid*.

6.7. Discussion

The results of the first experiment showed SlidAR+ having better performance than *Hybrid* under the hard condition but not under the easy condition. By including the orientation control in the tasks in the second experiment, we could see a change in terms of performance between SlidAR+ and *Hybrid* based on the conditions.

In overall performance (easy and hard trial), SlidAR+ showed a better completion time than *Hybrid* under Conditions 1 and 3, in which the targets were aligned to the gravity vector. In these conditions, the participants could easily correct the AR object's orientation using gravity constrainted rotation. Whereas in *Hybrid*, the participants could only perform rotation based on the device's camera perspective. However, we could not find any significant differences in completion time between the two methods when the targets were not aligned to the gravity vector (Condition 2 and 4), which support **H1**.

Under the easy trials, we observed a change in performance when compared to the first experiment, where in SlidAR+ now required significantly less time to completed the tasks than *Hybrid* under Conditions 1 and 3. However, *Hybrid* performed significantly faster than SlidAR+ under Conditions 2 and 4. In these conditions, SlidAR+ had no advantage over *Hybrid* in terms of rotaion control using gravity; thus, both methods had to manipulate the same amount of DoFs. One of the plausible reason is the separation of control scheme between the two methods. In *Hybrid*, the participant can control both position and orientation at the same time without need to switching the control mode, unlike in SlidAR+. Hence, **H3** is rejected.

In the trials under the hard condition, we found a significant performance difference with SlidAR+ performing faster than *Hybrid* in Condition 1, 3, and 4 but not in Condition 2. Hence, **H2** is rejected. We believe that the main reason SlidAR+ performed better in most of hard trials is because of the positioning, as we found in the first experiment. However, in the Condition 2 we could not find any significant difference between both methods. One of the plausible reason for this is in condition 2 the system coordinate was aligned to the gravity direction for both methods. So, the participant knew the initial orientation of created object and was able to predict the next step that need to be performed. Unlike in the Condition 4, the initial pose of the created targets in *Hybrid* appeared random or unpredictable for the participants, and this might have caused the performance of *Hybrid* to worsen under Condition 4 in comparison to Condition 3. However, we could not find any significantly evidence to support this and further investigation is required. Overall, the completion time results support only **H1**.

As for the device movement, SlidAR+ required significantly less movement in overall data (easy and hard trials) under Conditions 1,3 and 4. We also found the change in the results of easy trials under Conditions 1, 3, and 4. We also found the change in the results of the easy trials comparison to the first experiment, wherein we now found that SlidAR+ required significantly less device movement than *Hybrid* under Conditions 1 and 3 (where the target pose were aligned to the gravity vector). For SlidAR+, the participants needed to change their viewpoint only once to adjust the position. In the case of targets aligned to gravity vector, the participants can adjust the orientation without needing to move the camera/device to change the viewpoint. Under the hard trials, SlidAR+ required significantly less movement in every Condition which is similar to the results of the first experiment. This shows that the positioning process has a larger influence over device movement than orientation under hard trials.

The subjective results from HARUS showed that SlidAR+ is required significantly less mental effort and was easier to use than *Hybrid*, which is reflected by the comprehensibility score. But there was no significant difference between the two methods in terms of manipulability. We believe that this was mainly because the ways participants holding the device in each method have balanced out the physical effort to complete the task. For the translation, the holding behavior of participants is similar to the experiment 1 for both SlidAR+ and *Hybrid*. In SlidAR+ participants performed the task using one hand holding gesture for the whole process of rotation. However, in *Hybrid* participants can perform rotation using one or two hand holding gestures based on their preference. SlidAR+ might require less physical effort to input and operate but the way of holding in SlidAR+ might require more physical effort than one in *Hybrid*. One hand holds a big device such as an IPad Pro requiring more physical effort and less comfortable than two hand holding. Even though it is not significant, SlidAR+ likely has a higher average score in questions related to fatigue on arms and hands such as Q5 (SlidAR+: Avg 4.68, std 0.41; *Hybrid*: Avg 3.93, std 0.42). This is the reason we believe why the result of manipulability has no significant difference between two methods. Further analysis of HARUS scores revealed that participants found SlidAR+ to be simpler and easier to used as we observed a significant difference in Q4. SlidAR+ scored less in negative questions in Q9, Q11, and Q13 (Table 3), revealing that participants SlidAR+ required less mental effort than *Hybrid* to accomplish the same task.

To summarize, in the 6 DoFs tasks, SlidAR+ showed a faster completion time than *Hybrid* when the targets were aligned to the gravity vector. However, if they were not, SlidAR+ showed similar to *Hybrid*, but under hard trials only. It was slower in the easy trials. By combining the results of both the experiments, we can conclude that the gravity constrained orientation control feature can improve the performance of SlidAR when the targets are aligned to the gravity vector. However, it also worsens performance if the targets are not aligned.

We believe that SlidAR+ can be useful in the situations where it is hard to observe the depth cue/information. It can also be useful in situations where users have movement limitations in terms of size and control of the actual annotation space, as with remote collaboration scenarios. The subjective feedback and comments from the participants also suggest that SlidAR+ can help improve usability and reduce the mental effort required in placing 3D annotations.

6.8. Limitations

SlidAR+ was designed for users who want to place 3D content that is aligned either parallel or perpendicular to the direction of gravity. However, in cases where the target pose is not aligned to the direction of gravity, SlidAR+ still requires the user to control all 3 DoFs in orientation. SlidAR+ also requires the user to manually switch between position and rotation modes.

In the experiments, as we used a translucent object as the target, some of the participants found it difficult to see its orientation and this might have affected the results. Also, there were no real-world physical objects other than the pattern of the marker on the table that can be used as reference. The tracking quality and the error during the experiment might also effect the performance. Hence, the performance of both the methods may differ in other environment setups, the scale of the environment, and the number of object to manipulate at time same time might affect the results. Additionally, We do not record the amount of touch interaction in our experiment as we focus on the task completion time and user mental load rather than the number of interactions. However, SlidAR+ requires touch input to control both position and rotation. This makes SlidAR+ require more multiple combinations of touches compared to other methods that utilizes device-centric movement. This might cause SlidAR+ to require more learning time compared to methods with fewer touch inputs. Finally, the number of participants in our experiments are only 10 for experiment 1 and 16 for experiment 2.

The current SlidAR+ UI is not conducive for large devices because both of the participant' hands are needed to hold the device, which make it inconvenient to use SlidAR+. We believe that if we use a lighter and smaller device, the subjective feedback results for SlidAR+ will be improve.

7. Conclusions and future work

We presented SlidAR+ an object manipulation method for HAR applications. SlidAR+ utilizes ray-casting and epipolar geometry for positioning and gravity constrained orientation adjustment of virtual objects. SlidAR+ has been designed to minimize the number of inputs necessary to adjust the pose of the virtual object. Our experiments showed that SlidAR+ is more efficient than a state-of-the-art object manipulation method. It showed faster completion times, required smaller device movement when AR contents were to be placed aligned to the ground, and exhibited significantly better comprehensibility. We expect SlidAR+ to be used as an alternative choice for many in-situ AR object placement scenarios such as remote collaboration, navigation design, or virtual object manipulation for entertainment.

In future work, we would like improve the SlidAR+ UI in order to better support devices with large screens. We will try to find other techniques to rotate virtual object (when the target pose is not parallel or perpendicular to the direction of gravity). We also would like to explore SlidAR+ on other devices such as head-mounted displays.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Varunyu Fuvattanasilp: Methodology, Software, Formal analysis, Investigation, Writing - original draft, Visualization, Validation. **Yuichiro Fujimoto:** Writing - review & editing. **Alexander Plopski:** Writing - review & editing. **Takafumi Taketomi:** Conceptualization, Supervision. **Christian Sandor:** Supervision. **Masayuki Kanbara:** Supervision. **Hirokazu Kato:** Supervision.

References

[1] Polvi J, Taketomi T, Yamamoto G, Dey A, Sandor C, Kato H. SlidAR: a 3D positioning method for SLAM-based handheld augmented reality. In: *Computers and graphics*, 55; 2016. p. 33–43.

[2] Caudell TP, Mizell DW. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: *Proceedings of the Hawaii international conference on system sciences*, 2; 1992. p. 659–69.

[3] Polvi J, Taketomi T, Moteki A, Yoshitake T, Fukuoka T, Yamamoto G, et al. Handheld guides in inspection tasks: augmented reality vs. picture. In: *IEEE transactions on visualization and computer graphics*, 24; 2018. p. 2118–28.

[4] Kaufmann H, Schmalstieg D. Mathematics and geometry education with collaborative augmented reality. In: *Computers & graphics*, 27. Elsevier; 2003. p. 339–45.

[5] Shuhaiber JH. Augmented reality in surgery. In: *Archives of surgery*, 139. American Medical Association; 2004. p. 170–4.

[6] Azuma RT. A survey of augmented reality. In: *Presence: teleoperators & virtual environments*, 6. MIT Press; 1997. p. 355–85.

[7] Taketomi T, Uchiyama H, Ikeda S. Visual SLAM algorithms: a survey from 2010 to 2016. In: *IPSJ transactions on computer vision and applications*, 9; 2017.

[8] Marton F, Rodriguez MB, Bettio F, Agus M, Villanueva AJ, Gobbetti E. IsoCam: interactive visual exploration of massive cultural heritage models on large projection setups. In: *Journal on computing and cultural heritage*, 7; 2014. p. 12:1–12:24.

[9] 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 ACM SIGGRAPH symposium on video games*; 2009. p. 135–42.

[10] Bowman D, Kruijff E, LaViola Jr JJ, Poupyrev IP. *3D user interfaces: theory and practice*. Addison Wesley Longman Publishing Co, Inc; 2004.

[11] Henrysson A, Marshall J, Billinghurst M. Experiments in 3D interaction for mobile phone AR. In: *Proceedings of the international conference on computer graphics and interactive techniques in Australia and Southeast Asia*; 2007. p. 187–94.

[12] Apple. Apple augmented reality human interface guideline. 2018. Last check: Jun.14,2018; URL https://developer.apple.com/design/human-interface-guidelines/ios/system-capabilities/augmented-reality/.

[13] Kurz D, Benhimane S. Gravity-aware handheld augmented reality. In: *Proceedings of the IEEE international symposium on mixed and augmented reality*; 2011. p. 332–5.

[14] Orlosky J, Kiyokawa K, Takemura H. Dynamic text management for see-through wearable and heads-up display systems. In: *Proceedings of the international conference on intelligent user interfaces*; 2013. p. 363–70.

[15] Kurz D, Benhimane S. Handheld augmented reality involving gravity measurements. In: *Computers & graphics*, 36; 2012. p. 866– 883. Augmented Reality Computer Graphics in China

[16] 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 ACM symposium on spatial user interaction*; 2014. p. 13–16.

[17] Nielson GM, Olsen Jr DR. Direct manipulation techniques for 3d objects using 2d locator devices. In: Proceedings of the 1986 workshop on interactive 3D graphics; 1987. p. 175–82.

[18] Houde S. Iterative design of an interface for easy 3-d direct manipulation. In: Proceedings of the SIGCHI conference on human factors in computing systems; 1992. p. 135–42.

[19] Shoemake K. Arcball rotation control. In: *Graphics gems*; 1994. p. 175–92.

[20] Zeleznik RC, Forsberg AS, Strauss PS. Two pointer input for 3d interaction. In: Proceedings of the 1997 symposium on interactive 3D graphics; 1997. p. 115–20.

[21] Conner BD, Snibbe SS, Herndon KP, Robbins DC, Zeleznik RC, Van Dam A. Three-dimensional widgets. In: Proceedings of the 1992 symposium on interactive 3D graphics; 1992. p. 183–8.

[22] Bowman DA, Hodges LF. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Proceedings of the 1997 symposium on Interactive 3D graphics; 1997. p. 35–ff.

[23] Chaconas N, Höllerer T. An evaluation of bimanual gestures on the microsoft hololens. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE; 2018. p. 1–8.

[24] Mapes DP, Moshell JM. A two-handed interface for object manipulation in virtual environments. Presence 1995;4(4):403–16.

[25] Hayatpur D, Heo S, Xia H, Stuerzlinger W, Wigdor D. Plane, ray, and point: enabling precise spatial manipulations with shape constraints. In: Proceedings of the 32nd annual ACM symposium on user interface software and technology; 2019. p. 1185–95.

[26] Wang J, Lindeman RW. Object impersonation: towards effective interaction in tablet-and HMD-based hybrid virtual environments. In: 2015 IEEE virtual reality (VR). IEEE; 2015. p. 111–18.

[27] Kato H, Billinghurst M, Poupyrev I, Imamoto K, Tachibana K. Virtual object manipulation on a table-top ar environment. In: Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000). Ieee; 2000. p. 111–19.

[28] Reitmayr G, Eade E, Drummond TW. Semi-automatic annotations in unknown environments. In: *Proceedings of the IEEE and ACM international symposium on mixed and augmented reality*; 2007. p. 67–70.

[29] B N, E O, H B, D WA. SnapToReality: aligning augmented reality to the real world. In: *Proceedings of the ACM SIGCHI conference on human factors in computing systems*; 2016. p. 1233–44.

[30] Rui N, Nuno C. Magnetic augmented reality: virtual objects in your space. In: *Proceedings of the international working conference on advanced visual interfaces*; 2012. p. 332–5.

[31] Henrysson A, Billinghurst M, Ollila M. Virtual object manipulation using a mobile phone. In: *Proceedings of the international conference on augmented tele-existence*; 2005. p. 164–71. ISBN 0-473-10657-4.

[32] Castle RO, Murray DW. Object recognition and localization while tracking and mapping. In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*; 2009. p. 179–80.

[33] Bai H, Lee GA, Billinghurst M. Freeze View Touch and Finger Gesture Based Interaction Methods for Handheld Augmented Reality Interfaces. In: *Proceedings of the conference on image and vision computing New Zealand*; 2012. p. 126–31.

[34] Rousset E, Bérard F, Ortega M. Two-finger 3D rotations for novice users: surjective and integral interactions. In: *Proceedings of the international working conference on advanced visual interfaces*; 2014. p. 217–24.

[35] 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*; 2013. p. 12:1–12:10.

[36] Jung J, Hong J, Park S, Yang HS. Smartphone as an augmented reality authoring tool via multi-touch based 3D interaction method. In: *Proceedings of the ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*; 2012. p. 17–20.

[37] Telkenaroglu C, Capin T. Dual-finger 3D interaction techniques for mobile devices. In: *Personal and ubiquitous computing*, 17; 2013. p. 1551–72.

[38] Liu J, Au OK-C, Fu H, Tai C-L. Two-finger gestures for 6DOF manipulation of 3D objects. In: *Computer graphics forum*, 31. Blackwell Publishing Ltd.; 2012. p. 2047–55.

[39] Martinet A, Casiez G, Grisoni L. The design and evaluation of 3D positioning techniques for multi-touch displays. In: *Proceedings of the IEEE symposium on 3D user interfaces*; 2010. p. 115–18.

[40] Martinet A, Casiez G, Grisoni L. Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. In: *IEEE transactions on visualization and computer graphics*, 18; 2012. p. 369–80.

[41] Hancock M, Carpendale S, Cockburn A. Shallow-depth 3D interaction: design and evaluation of one-, two- and three-touch techniques. In: *Proceedings of the SIGCHI conference on human factors in computing systems*; 2007. p. 1147–56.

[42] Olsson T, Salo M. Online user survey on current mobile augmented reality applications. In: *Proceedings of the IEEE international symposium on mixed and augmented reality*; 2011. p. 75–84.

[43] Hürst W, van Wezel C. Gesture-based interaction via finger tracking for mobile augmented reality. In: *Multimedia tools and applications*, 62; 2013. p. 233–58.

[44] Chun WH, Höllerer T. Real-time hand interaction for augmented reality on mobile phones. In: *Proceedings of the international conference on intelligent user interfaces*; 2013. p. 307–14.

[45] Yousefi S, Kondori FA, Li H. Experiencing real 3D gestural interaction with mobile devices. In: *Pattern recognition letters*, 34; 2013. p. 912– 921.

[46] Bai H, Lee GA, Ramakrishnan M, Billinghurst M. 3D gesture interaction for handheld augmented reality. In: *Proceedings of the SIGGRAPH asia mobile graphics and interactive applications*; 2014. p. 7:1–7:6.

[47] Hürst W, Dekker J. Tracking-based interaction for object creation in mobile augmented reality. In: *Proceedings of the ACM international conference on multimedia*; 2013. p. 93–102.

[48] Bai H, Gao L, El-Sana J, Billinghurst M. Markerless 3D gesture-based interaction for handheld augmented reality interfaces. In: *Proceedings of the IEEE international symposium on mixed and augmented reality*; 2013. p. 1–6.

[49] Santos MEC, Taketomi T, Sandor C, Polvi J, Yamamoto G, Kato H. A usability scale for handheld augmented reality. In: *Proceedings of the ACM symposium on virtual reality software and technology*; 2014. p. 167–76. ISBN 978-1-4503-3253-8.