Computers & Graphics 55 (2016) 33–43

Contents lists available at ScienceDirect

Computers & Graphics

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

Technical Section

# SlidAR: A 3D positioning method for SLAM-based handheld augmented reality ☆

Jarkko Polvi [a,*], Takafumi Taketomi [a], Goshiro Yamamoto [a], Arindam Dey [b], Christian Sandor [a], Hirokazu Kato [a]

[a] Nara Institute of Science and Technology, Japan
[b] Worcester Polytechnic Institute, United States

ARTICLE INFO

*Article history:*
Received 28 March 2015
Received in revised form
21 August 2015
Accepted 28 October 2015
Available online 21 November 2015

*Keywords:*
Handheld augmented reality
3D manipulation
3D positioning
SLAM
User evaluation

ABSTRACT

Handheld Augmented Reality (HAR) has the potential to introduce Augmented Reality (AR) to large audiences due to the widespread use of suitable handheld devices. However, many of the current HAR systems are not considered very practical and they do not fully answer to the needs of the users. One of the challenging areas in HAR is the in-situ AR content creation where the correct and accurate positioning of virtual objects to the real world is fundamental. Due to the hardware limitations of handheld devices and possible restrictions in the environment, the correct 3D positioning of objects can be difficult to achieve we are unable to use AR markers or correctly map the 3D structure of the environment.

We present SlidAR, a 3D positioning for Simultaneous Localization And Mapping (SLAM) based HAR systems. SlidAR utilizes 3D ray-casting and epipolar geometry for virtual object positioning. It does not require a perfect 3D reconstruction of the environment nor any virtual depth cues. We have conducted a user experiment to evaluate the efficiency of SlidAR method against an existing device-centric positioning method that we call HoldAR. Results showed that SlidAR was significantly faster, required significantly less device movement, and also got significantly better subjective evaluation from the test participants. SlidAR also had higher positioning accuracy, although not significantly.



## 1. Introduction

AR refers to a combination of real world and virtual computer-generated objects where virtual objects are registered in 3D and can be interacted with real time [1]. HAR means AR on handheld devices such as smartphones, tablet computers, and ultra-mobile computers. The fast technical advancement of handheld devices has increased the interest of HAR among researchers and developers [2]. A vast amount of AR applications already exist for various handheld devices [3].

Currently, HAR provides the best means to introduce AR to the mass consumer market due to the widespread use of suitable handheld devices [4]. However, many of the existing HAR applications are not considered very practical due to insufficient functionality and they do not fully answer to the needs of the users [5,6]. Many design and technical challenges still remain and easy in-situ AR content creation is one of them.

In order for HAR to become widely accepted, the users must be able to create AR contents by positioning virtual objects in the real environment [7,8]. Furthermore, the potential HAR users want to create AR contents in various indoor and outdoor environments [9]. The basic 3D manipulation [10] of virtual objects is fundamental in HAR content creation and 3D positioning is the first subtask of virtual object manipulation.

HAR systems often utilize AR markers to track the environment, but the use of markers can be impractical or restricted in many use environments. Markerless tracking technologies, such as SLAM, track the environment without the need for adding any physical objects to the environment. In order to enable the 3D positioning of virtual objects, the markerless tracking based HAR system needs to reconstruct a 3D map of the environment. However, due to the insufficient processing capabilities of modern handheld devices and the vast amount of possible use environments, the correct 3D mapping of the environment might not always be possible. This can make the accurate 3D positioning of virtual objects very difficult.

In this paper, we present a SLAM-based HAR 3D positioning method called SlidAR (Fig. 1) that uses 3D ray-casting and epipolar geometry. This method enables accurate 3D positioning of virtual objects to the real environment, which 3D structure is not

☆This article was recommended for publication by Dirk Reiners.

* Corresponding author. Tel.: +81 90 6007 7611.
*E-mail addresses:* jarkko-p@is.naist.jp (J. Polvi), takafumi-t@is.naist.jp (T. Taketomi), goshiro@is.naist.jp (G. Yamamoto), adey@wpi.edu (A. Dey), sandor@is.naist.jp (C. Sandor), kato@is.naist.jp (H. Kato).

http://dx.doi.org/10.1016/j.cag.2015.10.013

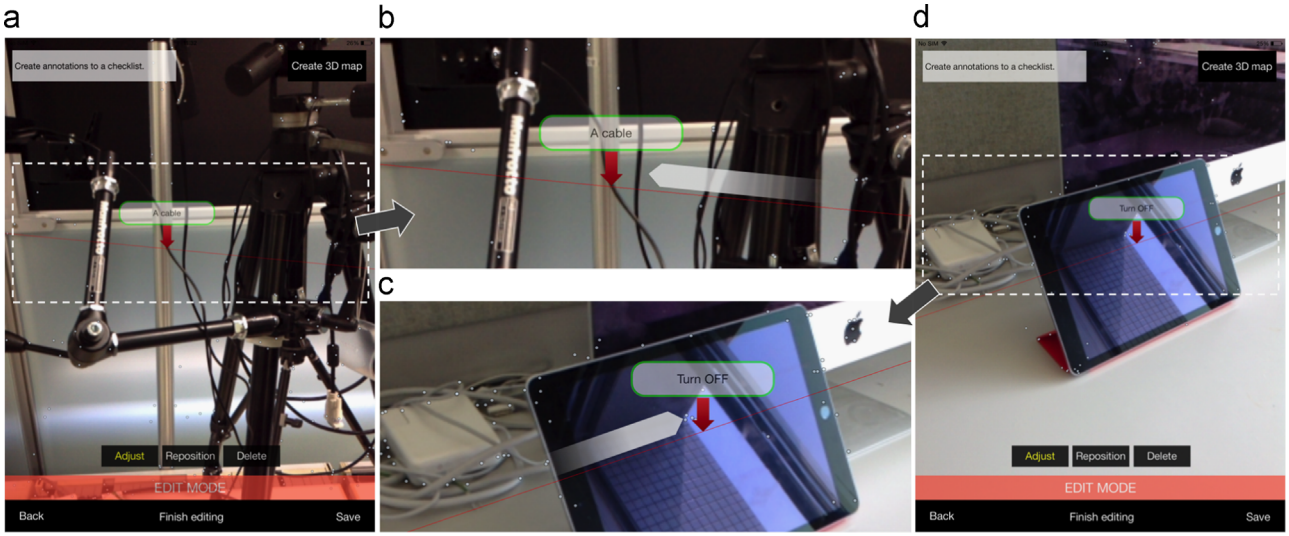


**Fig. 1.** The SlidAR method and possible practical scenarios: full view (a,d) and (b,c) close-ups. If we need to annotate challenging objects in the environment, it might not be possible to position virtual objects directly to the desired target position. The position needs to be adjusted. (a,b) A virtual object is positioned precisely to a thin cable. (c,d) A virtual object is positioned to a reflective surface. The white arrows (b,c) represent slide gestures along the red epipolar line. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

correctly mapped by the HAR system. The initial position is determined by tapping to the representation of the real environment on the handheld device's display. A ray is cast from the device's camera to the tapped initial position. The object's position can then be adjusted along the epipolar line. SlidAR does not use virtual depth cues and it also enables the positioning of virtual objects in mid-air. We have also implemented another SLAM-based 3D positioning method called HoldAR, which is similar the device-centric method first introduced by Henrysson et al. [11]. In HoldAR, a virtual object can be freely positioned by fixing it's position to the handheld device and physically moving the device. Virtual depth cues are displayed on a ground plane.

We conducted a user experiment to evaluate the efficiency of SlidAR against HoldAR. We asked the participants to position virtual objects to the real environment. The results showed that SlidAR was significantly faster and required significantly less device movement. The subjective feedback on SlidAR was also rated significantly higher. Although not significant, we observed that the positioning accuracy was also higher when using SlidAR.

*The main contribution* of this paper is the SlidAR 3D positioning method for SLAM-based HAR systems. SlidAR does not require special hardware and it could be implemented to a vast variety of consumer handheld devices suitable for AR. Even though the method is developed for SLAM, it can be applied to marker-based HAR and it can be useful in any scenario where accurate 3D positioning of virtual objects is required. We proved the efficiency of SlidAR in a user experiment and we believe this is the first HAR 3D positioning experiment to have virtual objects being associated accurately to real world objects. The insights acquired from our experiment can be helpful in the design of future HAR systems and user experiments.

The rest of the paper is organized as follows: Related work is discussed in Section 2. Section 3 describes the details of the two positioning methods we used in the experiment. Sections 4 and 5 explain the design of our experiment and the results, respectively. Finally, the results are discussed in Section 6 and future work in Section 7.

## 2. Related work

Bowman et al. [10] have designated three basic virtual object manipulation tasks for Virtual Reality (VR) and AR: selection, positioning and rotation. Authors define positioning as changing the 3D position of a virtual object. In this paper, we focus only on the HAR positioning task. In this section, we introduce AR positioning methods specific for handheld devices, and methods that utilize ray-casting applied in hardware other than handheld devices.

### 2.1. Handheld devices

Different manipulation methods for HAR have been widely studied. In related research, a single method is commonly implemented and evaluated for more than one manipulation task. For example, a method that combines positioning and rotation is often proposed. We present methods that have been designed solely for positioning or for more than one manipulation task including positioning. Here, the previous methods have been roughly divided into three groups: (1) buttons and touch-screen gestures, (2) mid-air gestures, and (3) device movement.

#### 2.1.1. Buttons and touchscreen gestures

Button-based positioning uses either the physical or the touchscreen buttons of a handheld device to position virtual objects. Henrysson et al. [11] have utilized smartphone's physical buttons for positioning where different buttons are mapped for different Degrees Of Freedom (DOF). Castle et al. [12] have applied touchscreen buttons in tablet computer HAR system to position objects in three DOF. In the work of Bai et al. [26], the positioning in two DOF is conducted in a freezed AR view using a combination of buttons and gestures.

Touch gestures have become a standard for 2D manipulation on touchscreen handheld devices [13] and they have been used extensively in HAR 3D manipulation as well. Jung et al. [14] have developed a system where virtual objects can be positioned in 3D by controlling one DOF at a time with a single or multitouch drag gestures. The controlled DOF is based on the pose of the device relative to a ground plane. Marzo et al. [15] have used the DS3 technique [16] for 3D multitouch gesture positioning on a smartphone. Their method displayed a shadow on the ground plane below the virtual object as a depth cue. Mossel et al. [17] have developed a method where the positioning is done with a slide gesture. The controlled DOF is based on the pose of the device relative to a ground plane. Kasahara et al. [18] have developed a

tablet system where positioning is done only by tapping to the desired location on the device's display. The position of a virtual object is determined by the feature points detected from the live AR view, which is then compared to an image database. Touchscreen gestures have also been utilized in commercial HAR applications like Minecraft Reality [19], Junaio [20] and the Ikea Catalog [21].

3D manipulation in VR has been widely studied and gesture-based positioning methods have also been applied for handheld VR systems. For example, Telkenaroglu et al. [22] and Tiefenbacher et al. [23] have experimented on 3D positioning in VR using touchscreen gestures. Interaction in handheld VR positioning shares similarities with HAR positioning, but there are also great differences related to scene navigation, etc. Thus, we will not discuss about handheld VR positioning methods more thoroughly.

#### 2.1.2. Mid-air gestures

Mid-air gesture HAR positioning methods utilize the user's mid-air finger movement in front of the device's camera. Virtual objects can be positioned by moving the finger while the system tracks the finger movement. Henrysson et al. [24] have developed a 2D and a 3D mid-air gesture positioning methods using the front-facing camera of a smartphone. In the 2D method, the positioning of a virtual object is done in a freezed AR view. After freezing the AR view, the object's position is translated in two DOF by moving the finger in front of the camera. A small colored dot on the user's finger is tracked. In the 3D method, an AR marker is attached to the user's finger allowing a three DOF positioning in a live AR view. In the method presented by Hürst et al. [25], positioning is done with different finger gestures in front of the back-facing camera of a smartphone while colored dots on user's fingers are tracked. Objects can be pushed with one finger or grabbed and moved with two fingers. Bai et al. [26] have also developed a finger gesture method where different axes of the objects position can be controlled by moving the finger in front of the back-facing camera.

#### 2.1.3. Device-centric movement

Device-centric methods utilize the movability and small form-factor of a handheld device. Virtual objects are positioned by moving the device while the object's position is fixed relative to the device. Henrysson et al. [11] have developed a one-handed and bimanual device-centric methods to a smartphone system using AR markers. The object's position can be controlled by pressing a physical button from phone's keypad and moving the device. Mossel et al. [17] have implemented the same method for a modern touchscreen smartphone. In their method, virtual lines based on axes are used as depth cues. Marzo et al. [15] have also implemented a similar method for a touchscreen smartphone and they use a virtual shadow below the object as a depth cue. Hürst et al. [25] have implemented the device-centric method for a smartphone system that uses only sensor based (a gyroscope, an accelerometer, and a compass) tracking. Guven et al. [27] have developed a device-centric method that uses a PDA and an external camera attached to it. Their method allows the AR view to be freezed while virtual objects are being positioned.

### 2.2. 3D ray-casting

A 3D ray-casting for positioning is utilized widely in Head-Mounted Display (HMD) AR systems. Reitmayr and Schmalstieg [28] have presented a mobile AR system for outdoors that utilizes HMD and a handheld device. Positioning is done by using the handheld device for casting a 3D ray through a crosshair displayed in the HMD. The system uses a predetermined 3D model of the buildings in the environment and the ray is intersected with the geometry of the buildings. Bunnun et al. [29] have developed an AR 3D modeling tool that uses 3D ray-casting and epipolar geometry to define the vertices of a plane. The system uses a handheld interface similar to a computer mouse with track wheels and buttons. A small camera is attached to this mouse-like interface, and the image is sent to a separate display.

Wither et al. [30] have developed a mobile AR system with a mouse input interface. The ray is cast using the first person view of the HMD and the mouse interface. The target position is determined based on the intersection point of the ray and geometry of the buildings recognized from the aerial images. Later, Wither et al. [30] developed another positioning method using similar kind of hardware, but instead of aerial images, their method used a single-point laser attached to an HMD. Lastly, Reitmayr et al. [31] have developed a SLAM-based method that allows pointing without pre-knowledge of the environment by detecting planar surfaces from the camera image.

### 2.3. Summary

Because SlidAR is developed for HAR and utilizes ray-casting, Table 1 shows a summary of existing HAR or AR ray-casting based positioning methods. The main difference between SlidAR and other HAR positioning methods is that, unlike previous methods, it utilizes ray-casting and does not require virtual depth cues nor AR markers. Using markers is not always possible and there are limitations on what kind of 3D structures and surfaces can be tracked from the environment. Previous 3D ray-casting based methods use HMD or other types of hardware and and they have not been implemented to a consumer handheld device. Some existing positioning methods do not use depth cues either, but their efficiency has not been confirmed in user experiments. Few methods utilize slide gestures, but those are not based on epipolar geometry.

As we can see from the previous user experiments and as stated by Bowman et al. [10], one manipulation method is not necessarily suitable for all three basic manipulation tasks. On the other hand, the combination of different methods for two or more manipulation tasks can be beneficial [17]. Buttons and touch gestures methods have been proven to be very efficient for rotation and scaling tasks, but they have difficulties in positioning tasks [11,17,15]. Mid-air finger gestures have been evaluated to be more suitable for entertainment purposes rather than for practical use [24,25]. Device-centric methods have been the most efficient for HAR 3D positioning [11,24,17,15]. We chose to compare SlidAR method against the conventional device-centric method, because device-centric positioning has been the most efficient 3D positioning method in previous experiments.

## 3. Positioning methods

We implemented two positioning methods, SlidAR and HoldAR [11], for a markerless SLAM-based HAR iPad system. The main difference between the two methods is that SlidAR relies on ray-casting and touchscreen gestures where HoldAR uses physical movement of the device. Our SLAM system uses PointCloud SDK [33] for markerless feature point detection and tracking of the environment. The PointCloud SDK uses images and internal sensor information of the handheld device. We divide a 3D positioning task with HAR into two phases: (1) initial positioning and (2) position adjustment. The initial positioning in both SlidAR and HoldAR is determined by tapping to the desired location on the representation of the real environment on the handheld device's display. The required level of accuracy in the initial positioning

**Table 1**
The summary of the HAR and ray-casting based positioning methods and their attributes from the related work. Several related publications present more than one method and those are marked with abbreviations: D = a device-centric method, B = a button or gesture based method, and M = a mid-air gesture based method.

| Method | Utilizes ray-casting | Usable on a handheld device | Does **NOT** require AR markers | Does **NOT** require preknowledge | Does **NOT** require virtual depth cues | Evaluated |
|---|---|---|---|---|---|---|
| Henrysson et al. [11]: B & D | | X | | | X | X |
| Reitmayr et al. [28] | X | | X | | X | |
| Wither et al. [32] | X | | X | | | X |
| Henrysson et al. [24] | | X | | | | X |
| Reitmayr et al. [31] | | | X | X | X | X |
| Bunnus et al. [29] | X | | X | X | X | |
| Castle et al. [12] | | X | X | X | X | |
| Wither et al. [30] | X | | X | X | X | |
| Hürst et al. [25]: M | | X | X | X | | X |
| Hürst et al. [25]: B | | X | X | | X | X |
| Bai et al. [26]: B | | X | | X | | X |
| Bai et al. [26]: M | | X | | X | | X |
| Jung et al. [14] | | X | | X | | |
| Kasahara et al. [18] | | X | X | | X | |
| Mossel et al. [17]: B & D | | X | | X | | X |
| Marzo et al. [15]: B & D | | X | | X | | X |
| **The SlidAR** | **X** | **X** | **X** | **X** | **X** | **X** |

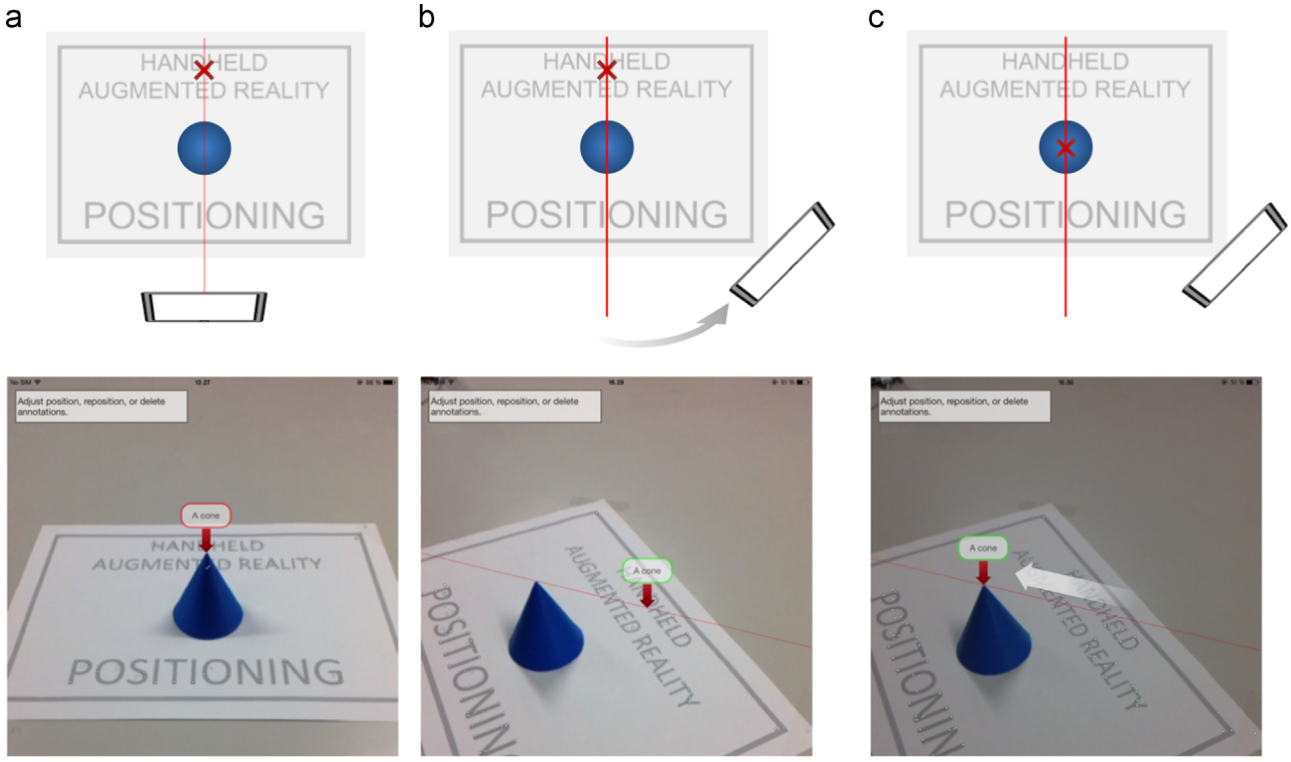


**Fig. 2.** The SlidAR method: top-down (above the dotted line) and display (below the dotted line) views. A virtual object (a white bubble and an arrow) is being positioned to the tip of a blue cone (the target position). (a) The object's position is perceived incorrectly from the first viewpoint. (b) A new viewpoint exposes the correct position of the object. (c) A ray from the device to the initial position intersects the target position and adjustment along the red epipolar line can be conducted with a slide gesture (shown as a white arrow). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

depends on the method used. The depth of the initial position is determined by the average depth $d_{average}$ of the surrounding feature points:

$$d_{average} = \frac{1}{|\boldsymbol{W}|}\sum_{i \in \boldsymbol{W}} d_i. \quad (1)$$

where $\boldsymbol{W}$ represents a set of natural feature points around the tapped area and $d_i$ represents a depth value for each feature point. The position adjustment in both methods is explained separately in Sections 3.1. and 3.2.

### 3.1. SlidAR

SlidAR utilizes 3D ray-casting and epipolar geometry for virtual object positioning (Fig. 2). After the initial positioning is conducted, a ray is cast from the handheld device's camera to the object's initial position. A ray can only be cast after initial positioning is done, because it requires camera pose and the ray direction information. This geometrical relationship between camera pose and a 3D point (in our case, a initial position) is known as epipolar geometry.

If the ray between the camera and the object's initial position intersects the target position (Fig. 2(a)), the object can be adjusted to the target position along the epipolar line using a slide gesture (Fig. 2(c)–(b)). The epipolar line is visualized as a 2D red line. We chose to use a slide gesture, because it gives smooth and precise control over the position of the virtual object. Furthermore, while conducting the slide gesture user's finger does not have to be directly on top of the epipolar line. This allows the adjustment to be done precisely without occlusion caused by the finger.

If the initial positioning is done incorrectly and the ray does not intersect with the target position, the ray must be recast by conducting the initial positioning again with a cut & paste function. The 3D position of the virtual object $\boldsymbol{p}_j$ is represented by the camera position $\boldsymbol{c}_i$, 3D ray direction $\boldsymbol{r}_j(|\boldsymbol{r}_j|=1)$, and the distance from the camera position to the object's position $l_j$. The relationship between these parameters is as follows:

$$\boldsymbol{p}_j = l_j \boldsymbol{r}_j + \boldsymbol{c}_i \tag{2}$$

The $l_j$ is changed during the adjustment phase, where the epipolar line defined by $\boldsymbol{c}_i$ and $\boldsymbol{r}_j$ is first projected onto the current image using the current camera pose $\boldsymbol{M}_t$ and the intrinsic camera parameters $\boldsymbol{K}$. The current camera pose is estimated by the SLAM algorithm. In the position adjustment phase, a new 3D position of the annotation $\boldsymbol{p}_i'$ can be calculated based on the object's position on the epipolar line.

The main difference between SlidAR and other ray-casting based positioning methods is the used hardware. Ray-casting with HMD has different ergonomical and perceptual issues because the ray is cast based on the head orientation instead of a handheld device's viewpoint. The actual adjustment in previous ray-casting based methods is done by using either special hardware or prior knowledge of the 3D structure of the environment. The positioning is easier this way, but our aim was to develop a method that is suitable for widespread adoption using low-cost hardware. Thus, we used only a handheld device without any prior knowledge of the environment, such as predetermined 3D model or aerial images. The system developed by Bunnus et al. [29] is closest to our work because it also uses ray-casting epipolar geometry. However, this system is not used for positioning virtual objects, but for making 3D models out of real world objects in AR. Furthermore, the hardware they used is different and it requires an external display.

### 3.2. HoldAR

A device-centric positioning method for HAR was first introduced by Henryson et al. [11] and we chose it for comparison, because it has been the most efficient for 3D positioning tasks in previous experiments. We call our SLAM-based implementation of this method HoldAR. Despite the different tracking technology, the interaction metaphor in HoldAR is similar to the marker-based device-centric methods introduced in the related work. With HoldAR, the position of virtual object is controlled by physically moving the device (Fig. 3). Unlike with SlidAR, the initial positioning can be done anywhere in the environment (Fig. 3(a)).

When a tap-and-hold gesture is performed on the handheld device's display, the position of the virtual object is fixed in the camera coordinate system and the object can be adjusted by moving the handheld device (Fig. 3(b) and (c)). When the tap-and-hold is released, the position of the object is set to the final adjusted position in the world coordinate system. The HoldAR shows two virtual depth cues: (1) a shadow (D=5 cm, alpha value=0.8) directly below the object on a ground plane and (2) a line between the object and the shadow. If the initial position is unclear or too far away from the target position, the initial positioning can be conducted again with a cut & paste function. In

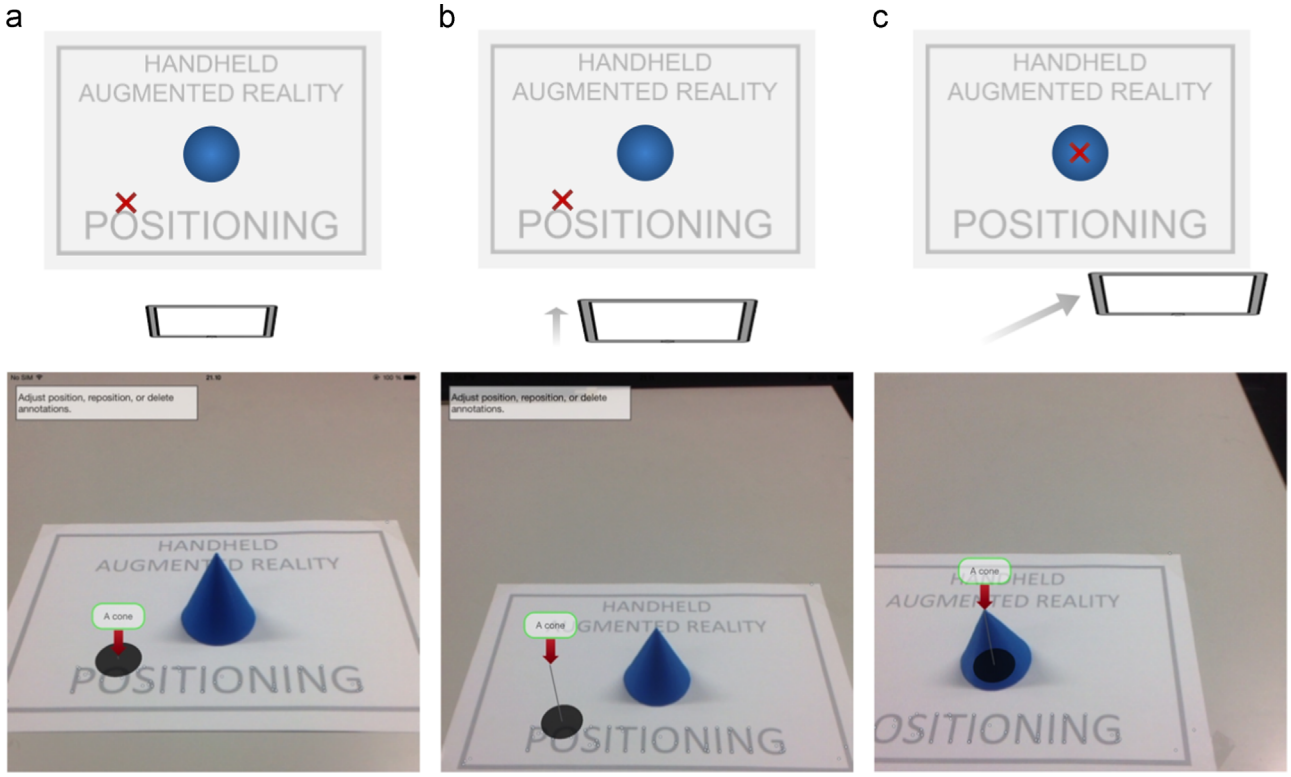


**Fig. 3.** The HoldAR method: a top-down (above the dotted line) and display (below the dotted line) views. A virtual object (a white bubble and an arrow) is being positioned to the tip of a blue cone (the target position). (a) The initial positioning is conducted near to the target position. A shadow is visualized below the object and a line between these two. (b) While taping and holding the device's display, the device is moved up and the object also moves up. (c) Again, the device is moved left and the object moves to same direction. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

order to visualize the depth cues correctly, a ground plane below the virtual object needs to be detected.

## 4. Positioning experiment

We conducted a user experiment to evaluate the efficiency of use [34] and subjective feedback of SlidAR against HoldAR. In our experiment, the efficiency consists of three objective quantifications: (1) the average time needed to complete a task, (2) the average magnitude of positioning errors (accuracy), and (3) the average amount of device movement needed to complete a task. In addition, we observed the usage of the positioning methods during the experiment sessions.

### 4.1. Experiment design

We conducted the experiment in a laboratory scenario instead of a practical application driven scenario, such as creating AR annotations to machines inside a factory or to medical equipment inside a hospital. HAR 3D positioning, and manipulation in general, has a large amount of possible application domains where it can be needed. We did not choose a certain application domain, because we wanted to focus on a single problem in HAR 3D positioning that is similar to all domains. Furthermore, different domains can have specific use environment related issues that would affect to the generalization of the results. Thus, we chose a laboratory scenario for easier generalization of the positioning task itself and for better controllability. In addition, we had to take into account the requirement of depth cues for HoldAR. That is, in order for the comparison to be fair, we needed a test scenario that has a ground plane.

We used a within-group factorial design that included two independent variables ($2 \times 2$): the positioning method (*SlidAR, HoldAR*) and the test task difficulty (*Easy, Hard*). The dependent variables were task completion time, positioning accuracy, device movement, and subjective feedback. Four conditions were evaluated and counterbalanced measures were taken (counterbalanced condition orders and breaks between conditions) to prevent possible learning effects. A total of 23 graduate school students (16 male and 7 female; mean age, $29 \pm 5$ years; age range, 22 to 41; mean height, $167.5 \pm 12.8$ cm) were recruited as test participants. All of them successfully completed the experiment. On a 7-point Likert scale (1=not familiar at all and 7=very familiar), participants estimated their previous experience with touchscreen handheld devices (M=6.4, SD=0.9), AR (M=4.2, SD=1.4), HAR (M=3.7, SD=1.5), and 3D user interfaces (M=4.6, SD=1.4).

We used a 4th generation iPad [35] as a test device. The 4th gen. iPad has a 1.4 GHz dual-core processor and a 9.7-inch screen with the native resolution of $1536 \times 2048$ pixels. In our HAR system, the resolution of the camera's video output was set to $360 \times 380$ pixels due to performance limitations of the iPad and the PointCloud SDK. The system was usable only in a portrait orientation. The SLAM maps of the test environment were created in advance and the detection of additional feature points was disabled during the experiment. The detected feature points were not visible to the participants. Every participant used the same SLAM maps.

### 4.2. Experiment tasks

In both tasks, participants had to position virtual objects relative to real world objects (Fig. 4). Here, with a virtual object we refer to a short 2D virtual text annotation. Textual information is 2D by nature and there is no need to present it in 3D [36]. This withdraws rotation and other manipulation tasks from the scope of this experiment. The participants were asked to number the virtual objects using the device's touchscreen keyboard.

Both tasks contained eight target positions at the top of eight Lego structures, and featured a predetermined order for conducting the eight positioning tasks. Because each participant did both tasks twice, two equally difficult versions of the predetermined positioning order were prepared. The structures were placed on a small table (length=80 cm, width=80 cm, and height=70 cm). The participants were allowed to move around the table if they felt it necessary. The picture on the surface of the table served as a ground plane for the depth cues in HoldAR. The Lego structures on the table were not part of the SLAM maps, which means that participants had to always conduct the position adjustment.

In both tasks, the eight target positions were on top of four low (height=16.32 cm) and four high (height=31.68 cm) Lego structures. The hard task was more dense because the structures were placed in closer proximity to each other and four faux structures were added. The faux structures did not have target positions. The purpose of the hard task was to investigate the effect of higher object density. The HAR system did not inform when the positioning was accurate enough and the level of accuracy was based on the participants' own perception. The target positions were located at the top most blocks in Lego structures and in order to avoid the ambiguity in accuracy measurement, they did not have a volume (Fig. 5). We used real world target positions instead of virtual ones in order to simulate a practical scenario.

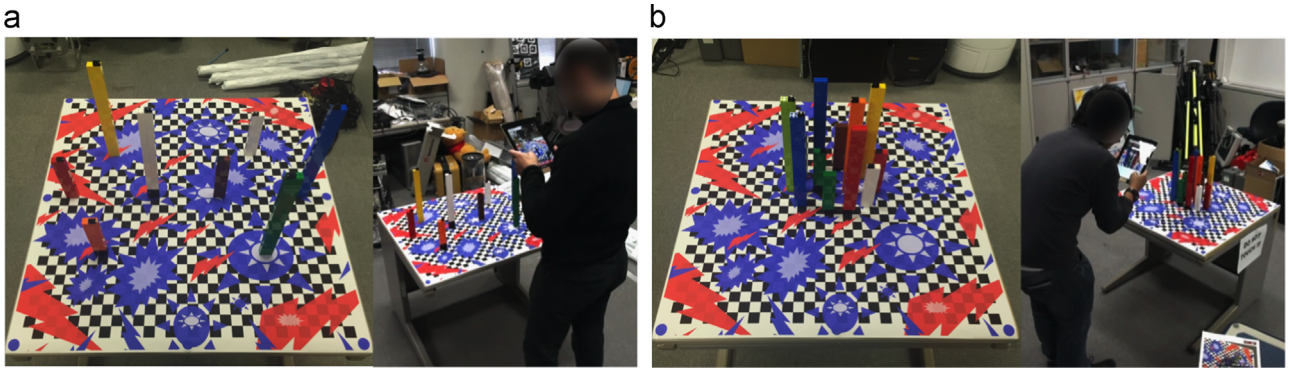


**Fig. 4.** The experiment tasks. (a) The easy task with eight target positions on top of eight Lego structures and a participant conducting the task. (b) The hard task with eight target position on top of eight Lego structures, four faux Lego structures, and a participant conducting the task.

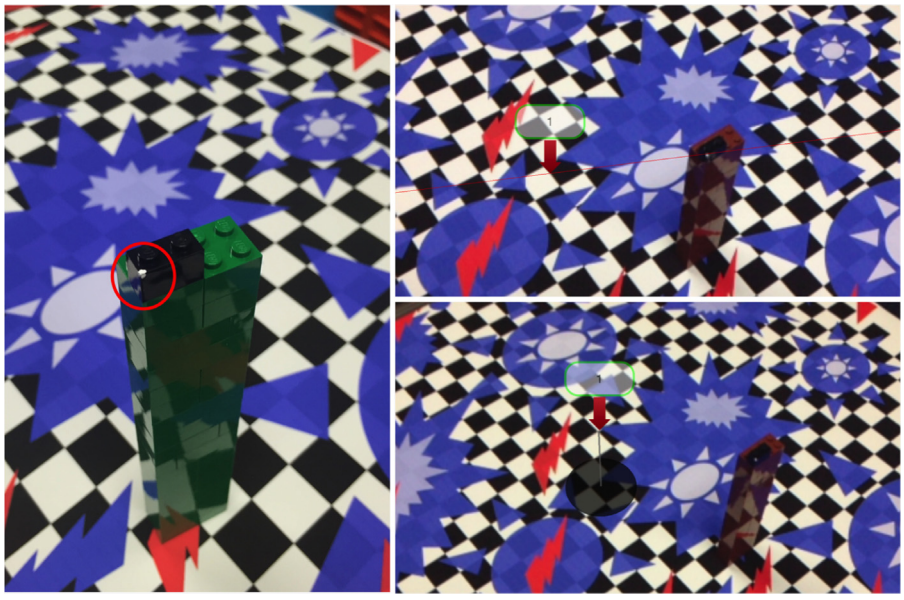

**Fig. 5.** A target position on the top most block of the Lego structure and positioning being conducted with both methods.

### 4.3. Experiment procedure

The user experiment consisted of a pre-questionnaire followed by the all four conditions and a post-questionnaire. The whole experiment took approximately 80–90 min per participant. After the pre-questionnaire, instructions (a slide presentation and a video demonstration) to both methods were given. Finally, participants were able to practice the methods in a tutorial tasks sequentially. Feedback was given to participants during the tutorial tasks. In the tutorial for SlidAR, we emphasized two main points: (1) The initial positioning should be done as accurately as possible; (2) In order to conduct the adjustment, the viewpoint needs to be changed from the initial viewpoint. For HoldAR, the following three main points were instructed: (1) The initial positioning can be anywhere in the environment. (2) The shadow is always directly below the virtual object on the ground plane; and (3) The movement of the device also moves the virtual object similarly.

The participants were instructed to position the virtual objects as accurately as possible, and move on whenever they felt the positioning was accurate enough or that they could not conduct it more accurately. We also instructed how to use the cut & paste function in situation where initial positioning was not done correctly. This was important especially with SlidAR where the position could be adjusted only along the epipolar line. The participants were told that the Lego structures are not part of the SLAM maps. They were also encouraged to check the position of the objects from different viewpoints. After each condition, there was a four minute break. During the break, the participants were reminded of the main points of the positioning method in the next condition, but did not receive any further feedback on their performance. In case of tracking failures, the system instructed participants to return to a marked starting point in front of the table and to initialize the tracking again.

### 4.4. Hypotheses

We formulated the following four hypotheses for the positioning experiment. H1–H3 address the different quantifications of efficiency of use and H4 deals with the effect of the task difficulty to HoldAR. Because the device movement required in SlidAR is more consistent and fewer DOFs are controlled at the same time, we hypothesize that it should performed significantly better against HoldAR (H1–H3). We assume that the environment has a higher effect to the effiency of HoldAR compare to SlidAR (H4), because HoldAR relies heavily to the virtual depth cues.

- **H1**: SlidAR has a lower task completion time compared to HoldAR.
- **H2**: SlidAR has a lower error rate in positioning accuracy compared to HoldAR.
- **H3**: SlidAR requires less device movement compared to HoldAR.
- **H4**: HoldAR has a higher efficiency in the easy task than in the hard task.

**Table 2**
The results from the objective measurements. N=23.

| Method | Task | Task time (s) | | Positioning error (mm) | | Device movement (m) | |
|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | Mean | SD |
| SlidAR | Easy | 361.04 | 122.84 | 11.5 | 16.0 | 35.8 | 15.6 |
| SlidAR | Hard | 403.65 | 172.04 | 12.0 | 20.2 | 37.4 | 13.4 |
| HoldAR | Easy | 488.61 | 248.04 | 14.0 | 11.0 | 53.3 | 23.9 |
| HoldAR | Hard | 601.96 | 255.73 | 14.3 | 11.1 | 65.7 | 31.4 |

## 5. Results

In this section, we describe the results of each objective and subjective measurement separately. Table 2 shows the summary of results for the objective measurements.

### 5.1. Task completion time

Fig. 6(a) shows the average task completion times. The measurement included all eight target positions in each task. The participant started the timing and stopped it after the task was completed. We noticed a significant difference between the methods in terms of overall task time from both tasks. A repeated-measure ANOVA showed that SlidAR (M=382, SD=149) was

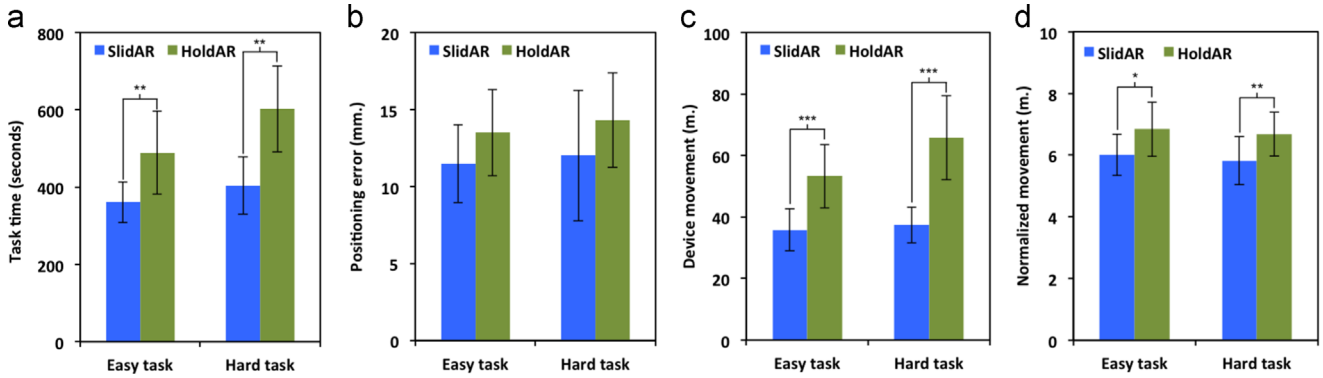


**Fig. 6.** The results from the objective measurements. (a) The average task completion times in seconds. (b) The average positioning errors in millimeters. (c) The average amount of device movement in meters. (d) Normalized device movement per minute in meters. Connected bars represent significant differences between means (*=significant at 0.05 level, **=significant at 0.01 level, ***=significant at 0.001 level). N=23 and error bars=+−95% CI.

significantly faster than HoldAR (M=545, SD=256) method: F (1, 22)=28.08, $p < .001$, p.e.s.=.56 (Fig. 2). Similarly and expectedly, we noticed a significant effect of the task difficulty on the completion time: F (1, 22)=16.61, $p$=.001, p.e.s.=.43. We did not notice any significant interaction effect of Method × Test task. The results support the **H1**, but not the **H4**.

### 5.2. Positioning accuracy

We calculated the average positioning errors in order to determine the overall accuracy (Fig. 6(b)). We measured the errors by calculating the distance between the positioning done by the participants and the target positions (Fig. 5). The 3D coordinates of the SLAM maps and absolute coordinate system were registered by manually specified corresponding points. Although SlidAR (M=14.8, SD=7.98) caused less error than HoldAR (M=18.3, SD=6.71) (Fig. 2), we did not notice a significant difference between the two: F(1,22)=2.66, $p$=.117, p.e.s.=.11. Similarly, the hard task did not cause more errors than the easy task: F(1,22)=2.81, $p$=.113, p.e.s.=.01. There was no significant interaction effect either. The results do not support **H2** nor **H4**.

### 5.3. Device movement

Fig. 6(c) shows the average amount of movement during the task. We measured the overall trajectories of the device's movement based on the pose of the device's camera related to the tracked environment. The camera pose information was saved 30 times per second and the trajectories between each pose were added together. The movement was calculated only while the environment was tracked and the extra movement caused by the loss of tracking was not included in the overall trajectories.

We analyzed the movement data using a repeated measures ANOVA. The analysis revealed that overall, when using SlidAR (M=36.60, SD=14.42), participants had to move the display significantly less than HoldAR (M=59.50, SD=28.31) F(1,22)=31.47, $p < .001$, p.e.s.=0.59 (Fig. 2). Expectedly, during the easy task (M=44.54, SD=21.83) participants had moved the device significantly less than during the hard task (M=51.56, SD=27.85) F (1,22)=18.04, $p < .001$, p.e.s.=0.45. We have also noticed a significant interaction effect of Method × Test task F(1,22)=4.4, $p < .05$, p.e.s.=.17. Both of the methods required less display movement for the easy task than the hard task. This decrease in movement was significantly more in the case of HoldAR than SlidAR.

Additionally, we analyzed device movement data normalized by time, i.e. device movement per minute 6(d). Our analysis revealed that SlidAR (M=5.91, SD=0.25) had significantly less device movement per minute than HoldAR (M=6.76, SD=0.27); F (1,22)=11.91, $p$=.002, p.e.s.=0.35. Interestingly, we did not notice a significant effect of task on normalized device movement. The device movement results support **H3** and **H4**.

**Table 3**
The HARUS statements.

| | |
|---|---|
| | Manipulability statements |
| S1 | I think that interacting with the positioning method requires a lot of body muscle effort |
| S2 | I felt that using the positioning method was comfortable for my arms and hands |
| S3 | I found the device difficult to hold while operating the positioning method |
| S4 | I found it easy to manipulate information through the positioning method |
| S5 | I felt that my arm or hand became tired after using the positioning method |
| S6 | I think the positioning method is easy to control |
| S7 | I felt that I was losing grip and dropping the device at some point |
| S8 | I think the operation of the positioning method is simple and uncomplicated |
| | Comprehensibility statements |
| S9 | I think that interacting with the positioning method requires a lot of mental effort |
| S10 | I though the amount of information displayed on screen was appropriate |
| S11 | I though that the information displayed on screen was difficult to read |
| S12 | I felt that the information display was responding fast enough |
| S13 | I though that the information displayed on screen was confusing |
| S14 | I though the words and symbols on screen were easy to read |
| S15 | I felt that the display was flickering too much |
| S16 | I though that the information displayed on screen was consistent |

### 5.4. Subjective feedback

We collected subjective feedback with the Handheld Augmented Reality Usability Scale (HARUS) [37] and written freeform comments. We also asked participants which method was their overall preferrence.

#### 5.4.1. Questionnaire

The HARUS (Table 3) measures participants' overall opinion about the manipulability (Table 3(a), S1–S8) and comprehensibility (Table 3 (b), S9–S16) of HAR on a 7-point Likert scale. The manipulability and comprehensibility statements consider different ergonomic and perceptual issues common to HAR, respectively. To analyze HARUS data we used paired two-tailed $t$-tests for the HARUS scores. For manipulability, SlidAR (M=70.83, SD=10.69) was significantly easier to handle than HoldAR (M=48.57, SD=18.54); t(22)=−4.82, $p < .001$.

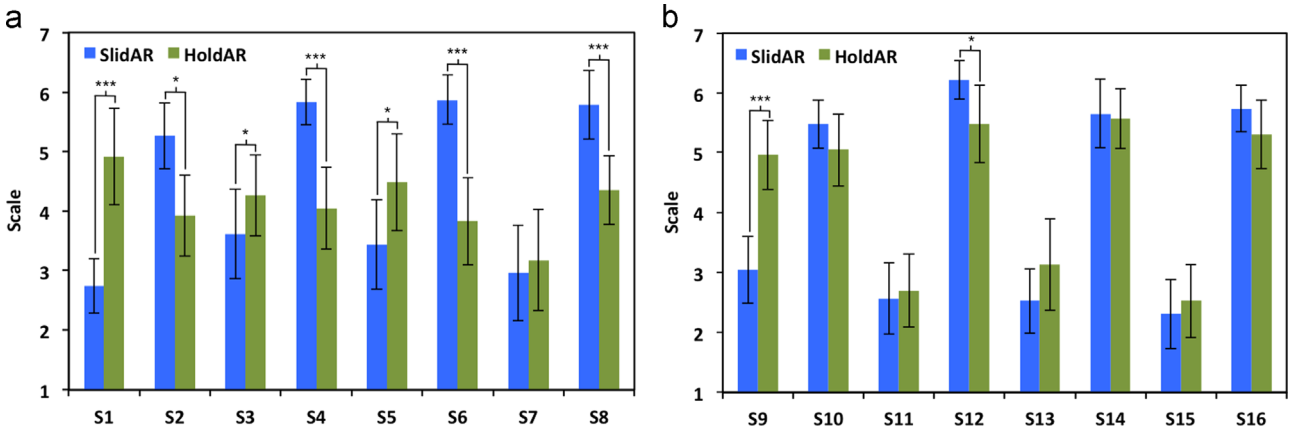


**Fig. 7.** Subjective feedback results from the HARUS in a 7-point Likert scale: (a) Manipulability statements and (b) Comprehensibility statements. S1–S16 represent statements from Table 3. Connected bars represent significant differences between means (*=significant at 0.05 level, ***=significant at 0.001 level). N=23 and error bars=+−95% CI.

For comprehensibility, SlidAR (M=76.3, SD=10.83) was significantly easier to understand than HoldAR (M=66.96, SD=15.71); $t(22)=-2.61$, $p=0.02$. Overall, SlidAR (M=73.57, SD=6.54) was significantly more usable than HoldAR (M=57.76, SD=15.39); $t(22)=-4.54$, $p<.001$. Fig. 7 illustrates the results of individual statements. A significant differences with $p<.001$ where found from S1, S4, S6, S8, and S9. A significant difference with $p<.05$ were found from S2, S3, S5, and S12.

#### 5.4.2. *Freeform comments*

Overall, 14 participants preferred SlidAR, seven preferred HoldAR, and two could not say. SlidAR was seen straightforward and fast. It did not require participants to move a lot, because in most cases the viewpoint had to be changed only once: from the initial viewpoint to a new viewpoint to conduct the adjustment. The drawback of SlidAR was the unclear visualization of the epipolar line. Moreover, the initial positioning was considered difficult because it had to be very precise. Even though it was not necessary to keep the fingers directly on top of annotation while conducting the slide gesture, some participants mentioned that their fingers sometimes block the view to the target position. Furthermore, participants noted that holding the device with one hand while conducting the initial positioning and position adjustment can be tiring.

The initial positioning with HoldAR was reported as fast because it did not have to be accurate. The position adjustment was seen as intuitive, but a precise matching of the virtual object to the target position was difficult. The simultaneous use of 3 DOF for adjustment was considered as unwanted because the method was sensitive to small movements and requires very fine adjustments and steady hands. The adjustment was seen more difficult in the hard task because the real objects were often occluded and it was difficult to perceive the position of the shadow correctly. Some participants felt that it was more intuitive to conduct the initial positioning precisely to the target position, similar to SlidAR, instead of positioning it freely to the close proximity of the target position.

### 5.5. *Observations*

The observations were conducted based on the video recordings of the device's display. In SlidAR, the 2D visualization of the epipolar line caused issues because participants were not always sure of the direction of the line. If the participants forgot the viewpoint of the initial positioning, they sometimes tried to conduct the position adjustment from the initial viewpoint. In HoldAR, the bad visibility of the shadow was sometimes an issue. If the ground plane had same coloring as the shadow, the shadow can get lost to the environment. This caused participants to perceive the depth incorrectly. The participants often adjusted the virtual objects to directions where they did not want objects to be adjusted because they were controlling all 3 DOF at once. With HoldAR, the positioning had to be confirmed from multiple viewpoints.

## 6. Discussion

In this section we discuss about the findings from the user experiment and how they could be applied to the field. In Section 6.1, we discuss about the objective and subjective data in our test scenario. In Section 6.2, we talk about how our experiment and the findings are applicable to practical scenarios.

### 6.1. *Test scenario*

We assume that SlidAR was faster mainly due to very specific target positions. Even though the accurate initial positioning took some effort, the position adjustment was quick and accurate because only 1 DOF was controlled. There was no need to constantly change the viewpoint and the adjustment was not affected by the unintentional movement of the device. The initial positioning with HoldAR was fast, but the position adjustment was time consuming because 3 DOF were controlled. This made the adjustment vulnerable to unintentional movement and perceptual errors.

A direct tap gesture is very intuitive for initial positioning, but it has problems regarding the ambiguity caused by user's fingers blocking the screen and the shakiness of the handheld devices. issue in SlidAR if target positions in the real world are very small in which case initial positioning has to be very precise. The initial positioning with a tap gesture could be improved with view freezing [27] or with a combination of view freezing and Shift [38].

Unlike SlidAR, HoldAR does not require a precise initial positioning because target position does not need to be on the ray cast from the camera. However, according to participants' comments the use of HoldAR requires more mental effort if they have to determine the initial position based on how effectively they can translate the virtual object from the initial position to the target position.

The perceptual issues [2,39] can have a considerable effect on positioning accuracy when target positions are real instead of

virtual. The combined average error rate in all conditions (M = 12.8 mm, SD = 1.3 mm) can be due to the issues in perception and the participants' judgment of the sufficient level of accuracy. A small positioning error can be very difficult to detect if the position is not checked from several viewpoints and at a close distance. Furthermore, the low resolution ($480 \times 640$ pixels) of the video output in our implementation and the 2D representation of virtual objects can affect the accuracy in both methods. The large amount of variation (Fig. 6(b)) in the positioning errors of SlidAR can be explained with the threshold of adjusting the objects position away from the epipolar line. Because an arbitrary adjustment with SlidAR was impossible, the virtual object had to be first moved with the cut & paste function and then adjusted again along the new epipolar line. Some participants may have settled with a certain level of accuracy due to the required effort in repositioning, even if they were aware that the position was not accurate enough.

The overall and normalized device movement needed was significantly higher because the position had to be adjusted and confirmed several times with HoldAR. The movement required while using SlidAR was more consistent. Furthermore, the adjustment was done with a finger gesture without the need to move the handheld device. The significant difference in movement between the easy and hard task with HoldAR can be associated with perceptual issues in understanding depth cues. The viewpoint had to be changed if the position of the object and it's shadow was unclear. We did not find significant differences between easy and difficult tasks. As such, based on our observations, the efficiency of SlidAR was not dependent on the environment's complexity.

The subjective results from HARUS strongly correlate to the results from the objective measurements. Completing the tasks with SlidAR took less effort in terms of time and movement, which is reflected to overall manipulability scores. The comprehensibility scores were also significant, but this was mainly due to S9 and S12, which are related to the difficulties in controlling and perceiving the position accurately. The remaining comprehensibility statements were expectedly not significantly different, because both positioning methods were implemented to the same HAR system and their user interfaces were very similar.

Although the experiment results only support **H1** and **H3** but not **H2**, we argue that the SlidAR was more efficient in our test scenario. It can achieve the same level of accuracy with significantly less time and less effort compared to the HoldAR method. The **H4** was supported only partially, but it shows that the environment can affect those methods that require virtual depth cues to be displayed in the environment.

### 6.2. Practical scenarios

In our test scenario, we considered two important aspects that are often missing from HAR positioning experiments in the related work: (1) We used real target objects (Lego structures, Fig. 5) instead of virtual ones (e.g. target zones visualized with virtual rectangles) to simulate a practical scenario where virtual objects are very often spatially dependent on the environment [40]. (2) We did not have predetermined initial positions for the virtual objects.

In practical scenarios, the initial positioning is a fundamental part of AR content creation and it should not be separated from the position adjustment. Especially in case of SlidAR, where position adjustment is highly dependent on the accuracy of initial positioning. Simply adjusting the position between two points can be unrealistic if we are unable to justify why user would have chosen the specific initial position. In addition, we forced participants to move around while doing the tasks instead of just standing still or sitting. This is important, because HAR is used in mobile context, where users often move around.

There are still few matter that should be considered when applying our findings to practical scenarios, such as creating AR annotations to machines inside of factory or to medical equipment inside a hospital. In the test scenario, the participants were aware that the real objects are not mapped by the SLAM-system. This was because we wanted to focus on the specific HAR positioning problem that can occur often, but not every time. We designed the test scenario in a way that the positioning problem occurs every time. In practical scenarios, users might not always what in the environment is mapped and what is not. Thus, we would not know if the virtual object's initial position going to be correct or is position adjustment also needed. If we use SlidAR, it is not necessary to know is the real world object mapped or not because the initial positioning is conducted in similar manner in both situations. With HoldAR, however, we might need to choose a initial position differently if it is too far away from the target position.

We did not limit the movement in any way and participants were allowed us to freely move around the scene. Even though neither method did not require users to move 360° around the target position, various environments in practical secenarios might set limitations to the movement. This could possibly affect the efficiency of both methods: SlidAR requires the user to move to a new viewpoint and HoldAR relies to movement entirely.

Our test scenario was ideal for HoldAR, because we used an easily trackable ground plane in order to correctly show the depth cues. The complexity and the structure of the environment can vary a lot depending on the scenario. This can make the correct visualization of depth cues more difficult. SlidAR does not require depth cues to be visualized on the environment, thus it can be easily used in various practical scenarios with different level of environmental complexities.

The required level of accuracy can also depend highly on the scenario and the size of the objects that are being annotated. Small objects, such as buttons or cables, require very precise positioning. Larger objects, like factory machinery, can allow more ambiguity. Positioning to a larger object is easier, regardless of the used method. Initial positioning would be easier with SlidAR and the position adjustment would be easier with the HoldAR.

We used a tablet device, but both methods could also be used on a smartphone. Tablets are beneficial because they provide more screen estate thereby easing perception and gesture-based interactions [41]. This can be beneficial in industrial or medical systems where it is necessary to view conventional 2D information, in addition to AR information. The form-factor of the device and the amount of movement needed can affect the usability of HoldAR because it relies on the physical movement. With SlidAR, the form-factor affects the initial positioning because the device had to be kept as still as possible in order to perform the position adjustment correctly. The initial positioning can be improved by adding view freezing discussed in the previous section.

We chose a generic test scenario instead of a practical one, because the positioning problem can occur in any kind of practical scenario. Conducting the experiment in a practical scenario, such as inside a hospital or a factory is risky, because the results could be affected by unique features of the scenario itself. This would steer the research focus away from the fundamental object positioning problem that is not specific for any type of scenario. A generic test scenario allowed us to focus more closely to the positioning problem and it gave us a solid implications regarding the efficiency of SlidAR. Furthermore, the whole experiment gave us important knowledge about the positioning of virtual annotations to real world objects. Practical scenarios might have some differences compared to our test scenario, but these are rather minimal. Furthermore, we believe that in practical scenarios SlidAR would provide even greater efficiency over HoldAR, because HoldAR requires more movement and virtual depth cues. Despite

the possible differences between our test scenario and practical scenarios, we strongly argue that our results can be applied to various scenarios, because the 3D positioning of virtual object is a requirement and fundamental part of any kind of practical scenario where we want to create AR content to the real world.

## 7. Conclusions and future work

We have developed SlidAR, a positioning method for SLAM-based HAR systems. SlidAR utilizes ray-casting and epipolar geometry. We have evaluated and proven it to be more efficient against a conventional device-centric positioning method that we call HoldAR. The results showed us that SlidAR was significantly faster, required significantly less device movement, and had significantly better subjective feedback compared to HoldAR method. SlidAR method also had higher positioning accuracy, although not significantly. The experiment confirmed the efficiency of SlidAR method.

For the future work, we can improve the initial positioning phase of SlidAR by adding the possibility to freeze the AR view. The visualization of the epipolar line can be improved by making it's direction more easier to understand. We should also consider techniques that allow the object's position to be translated away from the epipolar line. The next step in evaluation of SlidAR method is to use practical application driven scenarios that could possibly reveal new findings. We should also evaluate what is the required level of positioning accuracy for different types of real world objects in order for users' to perceive the spatial connection between the real object and the virtual annotation correctly.

## Acknowledgments

The authors work has been partly supported by Fujitsu Laboratories Ltd., Japan, and JSPS KAKENHI Grant number 26880016.

## References

[1] Azuma R. A survey of augmented reality; 1997.
[2] Kruijff E, Swan J, Feiner S. Perceptual issues in augmented reality revisited. In: 2010 9th IEEE international symposium on mixed and augmented reality (ISMAR); 2010. p. 3–12.
[3] Olsson T. User expectations and experiences of mobile augmented reality services, vol. 1085. Tampere, Finland: Tampere University of Technology Publication; 2012.
[4] van Krevelen DWFR, Poelman R. A survey of augmented reality technologies, applications and limitations. Int J Virtual Real 2010;9(2):1–20.
[5] Olsson T, Salo M. Online user survey on current mobile augmented reality applications. In: 2011 10th IEEE international symposium on mixed and augmented reality (ISMAR); 2011. p. 75–84.
[6] Grubert J, Langlotz T, Grasset R. Augmented reality browser survey. Technical report, University of Technology Graz; 2011.
[7] Kurkovsky S, Koshy R, Novak V, Szul P. Current issues in handheld augmented reality. In: 2012 international conference on communications and information technology (ICCIT); 2012. p. 68–72.
[8] Langlotz T. Ar 2.0: social media in mobile augmented reality [Ph.D. thesis]. Graz, Austria: University of Technology Graz; 2013.
[9] Vaittinen T, Karkkainen T, Olsson T. A diary study on annotating locations with mixed reality information. In: Proceedings of the 9th international conference on mobile and ubiquitous multimedia. MUM '10. New York, NY, USA: ACM; 2010. p. 21:1–21:10.
[10] Bowman DA, Kruijff E, LaViola JJ, Poupyrev I. 3D user interfaces: theory and practice. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc.; 2004.
[11] Henrysson A, Billinghurst M, Ollila M. Virtual object manipulation using a mobile phone. In: Proceedings of the 2005 international conference on augmented tele-existence. ICAT '05. New York, NY, USA: ACM; 2005. p. 164–71.
[12] Castle R, Klein G, Murray, D. Video-rate localization in multiple maps for wearable augmented reality. In: 12th IEEE international symposium on wearable computers, 2008 (ISWC 2008); 2008. p. 15–22.
[13] 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. CHI '07. New York, NY, USA: ACM; 2007. p. 1147–56.
[14] 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 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry. VRCAI '12. New York, NY, USA: ACM; 2012. p. 17–20.
[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. SUI '14. New York, NY, USA: ACM; 2014. p. 13–6.
[16] Martinet A, Casiez G, Grisoni L. Integrality and separability of multitouch interaction techniques in 3d manipulation tasks. IEEE Trans Vis Comput Graph 2012;18(3):369–80.
[17] 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. VRIC '13. New York, NY, USA: ACM; 2013. p. 12:1–12:10.
[18] Kasahara S, Heun V, Lee AS, Ishii H. Second surface: multi-user spatial collaboration system based on augmented reality. In: SIGGRAPH Asia 2012 emerging technologies. SA '12; New York, NY, USA: ACM; 2012. p. 20:1–20:4.
[19] Minecraft reality. URL: ⟨http://minecraftreality.com⟩; 2013, last checked: 2015-03-10.
[20] Junaio. URL: ⟨http://www.junaio.com⟩; 2013, last checked: 2015-03-10.
[21] Ikea catalog. URL: ⟨http://www.ikea.com/gb/en/catalogue-2015/index.html⟩; 2014, last checked: 2015-03-10.
[22] Telkenaroglu C, Capin T. Dual-finger 3d interaction techniques for mobile devices. Personal Ubiquitous Comput 2013;17(7):1551–72.
[23] Tiefenbacher P, Pflaum A, Rigoll G. Touch gestures for improved 3d object manipulation in mobile augmented reality. In: 2014 IEEE international symposium on mixed and augmented reality (ISMAR), http:dx.doi.org/10.1109/ISMAR:2014:6948467; 2014. p. 315–6.
[24] 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 Asia. GRAPHITE '07. New York, NY, USA: ACM; 2007. p. 187–94.
[25] Hürst W, van Wezel C. Gesture-based interaction via finger tracking for mobile augmented reality. Multimed Tools Appl 2013;62(1):233–58.
[26] Bai H, Lee GA, 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 new zealand. IVCNZ '12. New York, NY, USA: ACM; 2012. p. 126–31.
[27] Guven S, Feiner S, Oda O. Mobile augmented reality interaction techniques for authoring situated media on-site. In: Proceedings of the 5th IEEE and ACM international symposium on mixed and augmented reality. ISMAR '06. Washington, DC, USA: IEEE Computer Society; 2006. p. 235–6.
[28] Reitmayr G, Schmalstieg D. Collaborative augmented reality for outdoor navigation and information browsing. In: Proceedings of the second symposium on location based services and TeleCartography. Wien, Austria: TU Wien; 2004. p. 53–62.
[29] Bunnun P, Mayol-Cuevas WW. Outlinar: an assisted interactive model building system with reduced computational effort. In: Proceedings of the 7th IEEE/ACM international symposium on mixed and augmented reality. ISMAR '08. Washington, DC, USA: IEEE Computer Society; 2008. p. 61–4.
[30] Wither J, Coffin C, Ventura J, Hollerer T. Fast annotation and modeling with a single-point laser range finder. In: 7th IEEE/ACM international symposium on mixed and augmented reality, 2008 (ISMAR 2008); 2008. p. 65–8.
[31] Reitmayr G, Eade E, Drummond TW. Semi-automatic annotations in unknown environments. In: Proceedings of the 2007 6th IEEE and ACM international symposium on mixed and augmented reality. ISMAR '07. Washington, DC, USA: IEEE Computer Society; 2007. p. 1–4.
[32] Wither J, Diverdi S, Hollerer T. Using aerial photographs for improved mobile ar annotation. In: Proceedings of the 5th IEEE and ACM international symposium on mixed and augmented reality. ISMAR '06. Washington, DC, USA: IEEE Computer Society; 2006. p. 159–62.
[33] Pointcloud sdk. URL: ⟨http://developer.pointcloud.io⟩; 2013, last checked: 2015-03-10.
[34] Nielsen J. Usability engineering. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1993.
[35] 4th generation ipad. URL: ⟨https://support.apple.com/kb/SP662?⟩; 2012, last checked: 2015-03-10.
[36] Gabbard JL, Hix D. Researching usability design and evaluation guidelines for augmented reality (ar) systems. URL: ⟨http://www.sv.vt.edu/classes/ESM4714/Studentupdate_elementProj/class00/gabbard/⟩; 2001, last checked: 2015-03-10.
[37] Santos MEC, Taketomi T, Sandor C, Polvi J, Yamamoto G, Kato H. A usability scale for handheld augmented reality. In: Proceedings of the 20th ACM symposium on virtual reality software and technology. VRST '14. New York, NY, USA: ACM; 2014. p. 167–76.
[38] Vincent T, Nigay L, Kurata T. Precise pointing techniques for handheld augmented reality. In: Kotzé P, Marsden G, Lindgaard G, Wesson J, Winckler M, editors. INTERACT (1), Lecture notes in computer science, vol. 8117. Springer; 2013. p. 122–39 ISBN 978-3-642-40482-5. URL: ⟨http://dblp.uni-trier.de/db/conf/interact/interact2013-1.html#VincentNK13⟩.
[39] Dey A, Sandor C. Lessons learned: evaluating visualizations for occluded objects in handheld augmented reality. Int J Hum-Comput Stud 2014;72(10–11):704–16.
[40] Wither J, DiVerdi S, Höllerer T. Technical section: annotation in outdoor augmented reality. Comput Graph 2009;33(6):679–89.
[41] Dey A, Jarvis G, Sandor C, Reitmayr G. Tablet versus phone: depth perception in handheld augmented reality. In: 2012 IEEE international symposium on mixed and augmented reality (ISMAR); 2012. p. 187–96.