An Evaluation of Smartphone-Based Interaction in AR for Constrained Object Manipulation

Kristoffer Waldow
Computer Graphics Group, TH Köln
Köln, Germany
kristoffer.waldow@smail.th-koeln.de

Martin Misiak*
Computer Graphics Group, TH Köln
Köln, Germany
martin.misiak@th-koeln.de

Ursula Derichs
Computer Graphics Group, TH Köln
Köln, Germany
ursula.derichs@th-koeln.de

Olaf Clausen†
Computer Graphics Group, TH Köln
Köln, Germany
olaf.clausen@th-koeln.de

Arnulph Fuhrmann
Computer Graphics Group, TH Köln
Köln, Germany
arnulph.fuhrmann@th-koeln.de

ABSTRACT

In Augmented Reality, interaction with the environment can be achieved with a number of different approaches. In current systems, the most common are hand and gesture inputs. However experimental applications also integrated smartphones as intuitive interaction devices and demonstrated great potential for different tasks. One particular task is constrained object manipulation, for which we conducted a user study. In it we compared standard gesture-based approaches with a touch-based interaction via smartphone. We found that a touch-based interface is significantly more efficient, although gestures are being subjectively more accepted. From these results we draw conclusions on how smartphones can be used to realize modern interfaces in AR.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**; *User studies*;

KEYWORDS

Augmented Reality, User Interface Design, Smartphone, Study

ACM Reference Format:
Kristoffer Waldow, Martin Misiak, Ursula Derichs, Olaf Clausen, and Arnulph Fuhrmann. 2018. An Evaluation of Smartphone-Based Interaction in AR for Constrained Object Manipulation. In *VRST 2018: 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), November 28-December 1, 2018, Tokyo, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3281505.3281608

1 INTRODUCTION

Augmented Reality (AR) applications presented in head mounted displays (HMD) allow the user to experience virtual content embedded in their surrounding environment wherever they are. While navigation is becoming less of an issue, due to precise tracking technology, interaction is still being achieved by a variety of different interfaces. The most common are hand and gesture-based interactions, which integrate well into current HMDs and theoretically allow for a more natural interaction with the virtual environment. Recent approaches have also experimented with touch-based interfaces for AR systems. Budhiraja et al. [1] described several possible techniques to integrate handheld devices (HHDs) as inputs for selection tasks. Further research shows the efficiency of HHDs in combination with AR for unconstrained object transformation [3] and object selection [1] as well as in combination with VR for text input [5] and level editing [7]. Although it is possible to manipulate up to 7-DOF with a HHD [2], we constrain our transformations to 2-DOF since this kind of transformation is used in a wide range of available AR applications and was previously not investigated. Therefore, we evaluate a gesture-based interface against a touch-based interface for an object transformation task in an AR scenario.

For reference to established SOTA, we also included a gesture-based approach with gaze selection, which is designed to support transformations with additional DOF. For constrained object manipulation however, we expected this method to be less efficient.

2 EXPERIMENT

The purpose of this experiment was to understand the effectiveness of a touch-based interface compared to two gesture-based approaches. Therefore, the participants wore a *Microsoft HoloLens*, a self-locating optical see-through head mounted display (HMD). Additionally, a *Samsung Galaxy S8* was used as touch-based interface. A stationary server communicated wirelessly with the HMD and smartphone to log the task status and to switch seamlessly between the different conditions.

2.1 Task

The task was to scale and rotate a virtual character to match the transform of a reference character, the target, shown next to it. If the task was fulfilled, the virtual character turned green. For each iteration, the target was randomly rotated around the y-axis and scaled uniformly. In order to maintain the same task complexity throughout the iterations, the transformation difference towards the target was kept constant.

2.2 Conditions

The participants were exposed to three conditions in a counter balanced Latin square order.

*Also with HCI Group, University of Würzburg.
†Also with LCG, Federal University of Rio de Janeiro.

VRST '18, November 28-December 1, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6086-9/18/11.
https://doi.org/10.1145/3281505.3281608

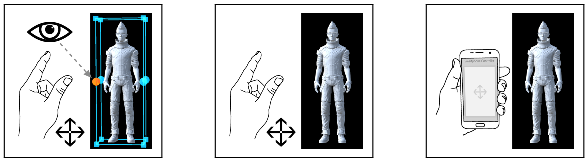

Figure 1: The three different interaction variants: *Left:* UI Handles, *middle:* Gestures and *right:* Smartphone.

The *UI Handles (H)* condition is based on *Microsofts HoloLens* transformation method with gaze, gestures and a bounding box with virtual handles (cf. Figure 1) [4]. By selecting one of these handles via gaze the user is able to perform the *HoloLens* air tap gesture to initialize either rotation or scaling. During the closed air tap gesture, the user can move the hand vertically to scale and horizontally to rotate the object. Thereby only one action is performed at the same time. The manipulation process is stopped once the tap gesture is opened again or if the hand is no longer tracked. The participant gets an acoustic and visual feedback about the current transformation state. The *Gestures (G)* condition is similar to the *UI Handles* , except there is no gaze selection and no bounding box. The transform selection is directly identified by the users' gestures once their hand starts moving.

In the *Smartphone (S)* condition, the participant uses touch-input for the transformation. The input movements themselves are consistent to the gesture-based methods and are similarly detected. By swiping vertically or horizontally on the display, the object is scaled or rotated.

2.3 Procedure

A total of $n = 21$ subjects (19 male, 2 female, average age 25) participated in the study. 57% had previous experience in VR or AR. At first, a short oral introduction was given to the participants, in which the task was clarified. For each condition, the participants received written instructions of the interaction method followed by a two-minute training session. Subsequently, five task iterations were performed with their completion times recorded. Finally, the participants had to answer four questions on a 7 point Likert scale (1 = very positive, 7 = very negative):

Q1 How good could you learn the interaction in the given time?
Q2 How easy was the interaction?
Q3 How intuitive was the interaction for scaling?
Q4 How intuitive was the interaction for rotating?

This procedure was repeated for the remaining two conditions. In a final question, the participants selected a favorite interaction method and justified their decision.

2.4 Analysis of the Results

Since the recorded measurements are time based, the produced results do not follow a normal distribution (*Shapiro-Wilk*: $\alpha = 5\%, p_H = 0.207, p_G = 0.299, p_S < 0.001$). A *Friedman* significance test shows an effect in the task completion time ($\alpha = 5\%$, $\chi^2(2) = 36.86, p < 0.001$). The smartphone interaction is significantly faster (post-hoc *Wilcoxon* comparison with *Bonferroni* correction: $Z_{HS} = -4.015, p_{HS} < 0.001; Z_{GS} = -3.215, p_{GS} < 0.004$) with medians of $Mdn_S = 3.91s$, $Mdn_G = 6.17s$ and $Mdn_H = 21.68s$ (cf. Figure 2, *left*). In the questionnaire, the answers are not equally distributed (*Shapiro-Wilk*: $\alpha = 5\%, p_{Q1-Q4} < 0.05$). A *Friedman* significance test yield significant variation among conditions ($\alpha = 5\%$, $\chi^2_{Q1}(2) = 34.85, \chi^2_{Q2}(2) = 34.76, \chi^2_{Q3}(2) = 23.40, \chi^2_{Q4}(2) = 18.03, p_{Q1-Q4} < 0.001$). A post-hoc *Wilcoxon* test with *Bonferroni* correction only reveals a significant difference for Q2 for the *Smartphone* and *Gestures* condition ($Z_{GS} = -2.86, p_{GS} < 0.013$). For all answers, the *UI Handle* method differentiates significantly to all other conditions which was to be expected (cf. Figure 2, *right*). In the final question, the *Gestures* condition was selected significantly more often as favorite interaction type ($\chi^2(2) = 8.95, p < 0.012$).

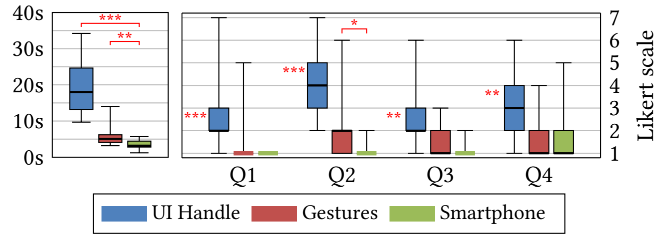

Figure 2: *Left:* The task completion time during the 5th iteration for each condition. *Right:* Given answers of the questionnaire.

3 DISCUSSION AND FUTURE WORK

In our study, we observe that the usage of a touch-based interaction in AR is more time efficient in constrained transformation tasks compared to both gesture-based methods. A possible explanation for this could be that a smartphone interaction is more common and better known. The most surprising aspect of the result is that the subjects rated the *Gestures* condition more likable even though they tend to perform the interaction with a smartphone significantly faster. This greater acceptance of gestures was also observed by van Beurden et al. [6]. Multiple participants associated with gestures that they were grabbing the virtual character with their real hands which was perceived as a natural way of manipulating the character.

In conclusion, a smartphone-based interaction in AR performs significantly faster compared to gestures for constrained object transformation even though the gestures are more enjoyable. For future work, we suggest combining smartphones and gestures for an efficient and pleasant user interface. Also we plan to extend the user interface to more complex interactions like text input.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research (BMBF) under grant number 02K16C232 (Retail 4.0).

REFERENCES

[1] Rahul Budhiraja, Gun Lee, and Mark Billinghurst. 2013. Using a HHD with a HMD for Mobile AR Interaction. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*.

[2] Jerônimo Gustavo Grandi, Henrique Galvan Debarba, Luciana Nedel, and Anderson Maciel. 2017. Design and Evaluation of a Handheld-based 3D User Interface for Collaborative Object Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5881–5891.

[3] Alexandre Millette and Michael J. McGuffin. 2016. DualCAD: integrating augmented reality with a desktop GUI and smartphone interaction. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 21–26.

[4] Yoon Park and David Kline. 2018. MixedRealityToolkit-Unity -Bounding Box & App Bar. (2018). https://github.com/Microsoft/MixedRealityToolkit-Unity

[5] Sebastian Pick, Benjamin Weyers, Bernd Hentschel, and Torsten W. Kuhlen. 2016. Design and Evaluation of Data Annotation Workflows for CAVE-like Virtual Environments. *IEEE Trans. Vis. Comput. Graph.* 22, 4 (2016), 1452–1461.

[6] Maurice HPH van Beurden, Wijnand A Ijsselsteijn, and Yvonne AW de Kort. 2011. User experience of gesture based interfaces: a comparison with traditional interaction methods on pragmatic and hedonic qualities. In *International Gesture Workshop*. Springer, 36–47.

[7] Jia Wang, Owen Leach, and Robert W. Lindeman. 2013. DIY World Builder: an immersive level-editing system. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. 195–196.