2017 IEEE International Symposium on Mixed and Augmented Reality Adjunct Proceedings

[POSTER] Usability Analysis of an off-the-shelf Hand Posture Estimation Sensor for Freehand Physical Interaction in Egocentric Mixed Reality

Andreea-Dalia Blaga* Maite Frutos-Pascual† Maadh Al-Kalbani‡ Ian Williams§

DMT Lab, School of Computing and Digital Technology
Birmingham City University

Abstract

This paper explores freehand physical interaction in egocentric Mixed Reality by performing a usability study on the use of hand posture estimation sensors. We report on precision, interactivity and usability metrics in a task-based user study, exploring the importance of additional visual cues when interacting. A total of 750 interactions were recorded from 30 participants performing 5 different interaction tasks (Move, Rotate: Pitch (Y axis) and Yaw (Z axis), Uniform scale: enlarge and shrink). Additional visual cues resulted in an average shorter time to interact, however, no consistent statistical differences were found in between groups for performance and precision results. The group with additional visual cues gave the system and average System Usability Scale (SUS) score of 72.33 (SD = 16.24) while the other scored a 68.0 (SD = 18.68). Overall, additional visual cues made the system being perceived as more usable, despite the fact that the use of these two different conditions had limited effect on precision and interactivity metrics.

Index Terms: H.5.2 [User Interfaces]; I.3.6 [Methodology and Techniques]

1 Introduction

Optical See-Through (OST) devices and Head-Mounted Displays (HMDs) have experienced an unprecedented growth in consumer availability, leading to the development of new and innovative Augmented Reality (AR) applications for general public use [10].

AR interaction uses a wide variety of interfaces and wearable options, however, freehand-based interaction is often considered the most intuitive, natural and affordable, not requiring cumbersome wearable input devices [20].

When creating new freehand interactive systems in AR, several studies rely on predefined sets of gestures for natural interaction [5, 24, 25]. However, there is a growing trend for the use of physical freehand interaction, defined as the interaction that act physically on the virtual object as if it was a real object [15].

New generation HMD and OST devices such as the Microsoft HoloLens include a suite of integrated sensors and tracking cameras for 3D modeling the surroundings and freehand gesture recognition [6]. However, these devices currently offer a limited freehand interaction similar to mouse clicks on a desktop. To overcome this limitation, researchers are currently combining hand posture estimation sensors with Head Mounted Displays (HMDs) for hand posture calibration in AR [10, 11] and for immersion and interaction enhancement in Virtual Reality (VR) environments [8, 13]. However, the usability, interactivity and precision of hand posture estimation sensors in an egocentric AR/MR setting for physical freehand interaction has not been fully addressed yet.

Following Piumsomboon *et al.* [14] definitions, we define the interactivity as the users' ability to manipulate virtual objects and the precision as the level of control the user has when interacting. An example measure of precision would be how accurately the user can rotate or move an object to match a target, while we reported on interactivity as the perceived usability of the system.

This paper presents a prototype evaluation of the hand posture estimation sensor Leap Motion using an egocentric perspective. It reports on its usability and precision during a task-based interaction test. We contribute to the AR/MR field in the evaluation of the potential of this off-the-shelf sensor to be incorporated to fully immersive freehand AR/MR environments using physical natural interaction.

The paper is structured as follows, Section 2 discusses relevant studies related to freehand interaction, Section 3 describes the grasping metrics used, while Section 4 presents the experiment design. The results and discussion of the findings are reported in Sections 5 and 6.

2 Related Work

Freehand gesture is the most common human way of communication and controlling objects in the real world, underpinning powerful interactions due to hands multiple degrees of freedom [17]. This interaction paradigm has been linked to ease of access and naturalness in the literature due to the absence of constraints imposed by wearable devices and its potential in delivering natural, intuitive and effective interaction [7, 9, 12, 16, 20].

However, these interactions also present some limitations and challenges; gestures may not be easy to remember and long interactions may result in fatigue, as interaction with no physical support is tiring [17].

To overcome these limitations, researchers are evolving gesture interactions into physical interactions, thus, completely mimicking the interaction performed in the real world, into AR/MR environments [19]. Several studies have explored the easiness and naturalness of user defined gesture interaction [15, 21, 23]. Recent literature in this field includes the analysis in Mixed Reality (MR) [1, 18, 19]. However, the use of grasping metrics in an egocentric task based environment has not been fully explored.

2.1 Grasping Metrics

Limitations and challenges when addressing a natural grasping interaction in MR were addressed by Al-Kalbani *et al.* in their recent studies [1, 2]. Grasp measurements used in this study have been inspired by those presented by Al-Kalbani *et al.* in their study [1].

3 Experiment Design

Five different tasks were reported in this experiment to quantify the usability and precision of the Leap Motion sensor as a HMD complement. Two different conditions were explored for each of the five tasks, one with an overlaid 3D model of the hand and one without, as shown in Figure 2. We used a 5 x 5 x 2 repeated

*e-mail: andreea.blaga@mail.bcu.ac.uk
†e-mail:maite.frutos@bcu.ac.uk
‡e-mail:muadh.alkalbani@bcu.ac.uk
§e-mail:ian.williams@bcu.ac.uk

978-0-7695-6327-5/17 $31.00 © 2017 IEEE
DOI 10.1109/ISMAR-Adjunct.2017.25

measures (between-subjects design), with 5 different tasks (Move, Rotate: Pitch (Y axis) and Yaw (Z axis), Uniform Scale: Enlarge and Shrink) 5 repetitions per task and 2 conditions (hand overlaid, no hand overlaid).

3.1 Participants

30 right-handed participants from a population of university students and staff members were recruited to take part in this study.

Participants were divided into two balanced groups of 15 for each experiment (No Overlaid and 3D Hand Overlaid, as it is shown in Figure 2) . Participants completed standardized consent form and were not compensated. Visual acuity of participants was measured using a Snellen chart, each participant was also required to pass an Ishiara test to exclude for color blindness. Participants with color blindness and/or visual acuity of < 0.80 (where 20/20 is 1.0) were not included in this study.

No Overlaid group was formed of 15 participants (7 female) with a mean age of 25.20 (SD = 8.79); the 3D Hand Overlaid group was formed of 15 participants (7 female) with a mean age of 27.46 (SD = 11.91).

3.2 System Architecture

The system in this study comprises of a Leap Motion sensor featuring an egocentric view and a feedback monitor. Freehand interaction was implemented using the Leap Motion Orion SDK, Unity game engine and *C#* programming language.

3.3 Task Selection

Each task contains two 3D objects on display, one being the interaction object and the other the target object. Target objects were not interactive and they were displayed in a different color to avoid confusion.

Tasks were selected based on the survey presented in [15], where Piumsomboom *et al.* categorized the most common AR interaction tasks into 6 categories. This study analyzed a subset of tasks extracted from the Transform category, each task is described below:

- *Move - Short distance*: Participants were required to successfully locate, grasp and move a cube to its target location in the 3D space. See figure 1(a).
- *Rotate - Pitch (Y axis)*: Participants were required to successfully locate, grasp and match the Y axis rotation of the interaction cube with the target cube. See figure 1(b).
- *Rotate - Yaw - (Z axis)*: Participants were required to successfully locate, grasp and match the Z axis rotation of the interaction cube with the target cube. See figure 1(c).
- *Scale - Uniform scale (enlarge)*: Participants were required to successfully locate, grasp and match the size of the interaction cube with its target by increasing it. See figure 1(d).
- *Scale - Uniform Scale (shrink)*: Participants were required to successfully locate, grasp and match the size of the interaction cube with its target by decreasing it. See figure 1(e).

3.4 Procedure

Participants were asked to remove wristwatches, sleeves and bracelets to avoid the detection of extraneous objects as interaction hands by the sensor. Participants were seated 40 cm away from the feedback monitor with the Leap Motion placed in a stand facing their hands. All participants in this study completed the same test, which consisted of a set of 5 task-based interactions, as explained in subsection 3.3. Participants were then instructed to use their dominant hand (right) to perform the tasks. Each participant was shown a demonstration video along with scripted instructions.

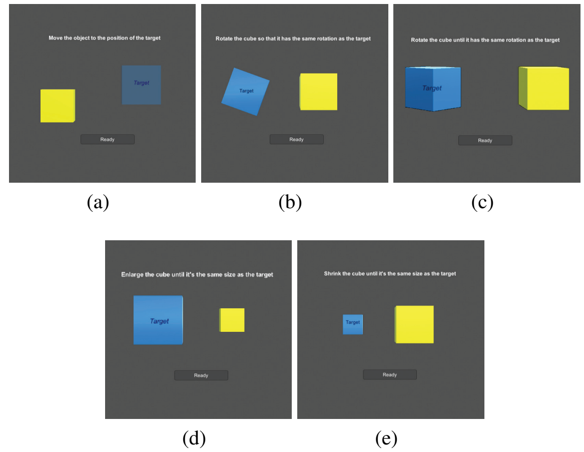
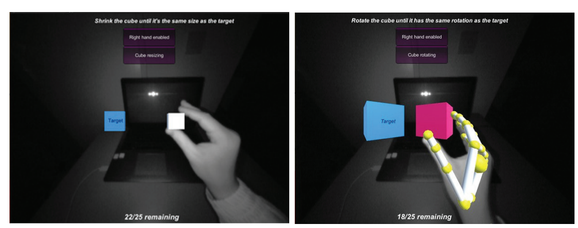


Figure 1: Tasks performed by participants: 1(a) Move. 1(b) Rotate pitch. 1(c) Rotate yaw. 1(d) Scale enlarge. 1(e) Scale shrink.

Participants were instructed to locate and grasp the virtual interaction object, and then perform the required task to match the virtual object with its target in the most accurate way in the shortest time possible. During the experiments, time spent by participants in triggering the interaction and performing the task was recorded, together with the position, size, and rotation of the virtual interaction object in every frame.

The tasks were randomly presented one after the other, in order to reduce any bias or lurking variable that may have an influence on the outcome of the study. Each task was repeated five times.

Half of the participants performed the task with a 3D modeled hand overlaying theirs, providing further depth reference. The other half could only see the real hand from the Leap Motion's cameras with no additional visual cues. These two conditions are showcased in Figure 2.



(a) (b)

Figure 2: Study conditions. 2(a) No Overlaid condition 2(b) 3D Hand Overlaid condition

4 Results

A total of 750 interactions were generated from the 30 participants divided in two different groups performing the 5 different tasks 5 times each. Repetitions were treated as additional data points. The data collected for each user included raw tracking collected from the hand posture estimation, together with user's subjective data (SUS questionnaire and post-test interview), and transcripts taken from a think-aloud protocol. We report on precision (time and accuracy) and interaction (perceived interactivity and usability) metrics.

Table 1: Average time to interact and completion time per task and per condition

	Move	Rotate - Pitch	Rotate - Yaw	Scale - Enlarge	Scale - Shrink
Time to interact *(in seconds)*					
No Overlaid	5.36 (SD = 2.99)	3.39 (SD = 4.90)	3.21 (SD = 2.28)	3.19 (SD = 2.49)	3.47 (SD = 3.26)
3D Hand Overlaid	4.01 s (SD = 2.47)	2.89 s (SD = 1.82)	2.92 s (SD = 2.45)	3.5 s (SD = 2.15)	3.22 s (SD = 1.83)
Completion time *(in seconds)*					
No Overlaid	16.58 (SD = 10.67)	16.99 (SD = 13.88)	13.84 (SD = 10.24)	19.62 (SD = 11.69)	24.3 (SD = 18.49)
3D Hand Overlaid	18.93 (SD = 10.98)	14.55 (SD = 7.88)	12.79 (SD = 9.93)	23.39 (SD = 21.05)	22.8 (SD = 18.06)

4.1 Precision metrics

Precision metrics were defined as follows: *time to interact* as the time it took to locate and start the interaction with the object, *completion time* as the time it took to complete the task, and *accuracy* as the difference between the target and the interaction cube at the end of the task.

Due to the nature of the data being non parametric and not normally distributed, statistical significance was tested using a non parametric Mann Whitney-U test (5% alpha) comparing the two conditions (No Overlaid and 3D Hand Overlaid).

4.1.1 Results - Time to interact

Average time to interact per task and condition are reported in Table 1. Statistically significant differences were found between groups in the time it took to successfully locate and start to interact with the object in tasks move ($U = 1871$, $p < 0.05$) and scale enlarge ($U = 2328$, $p < 0.05$). No statistically significant differences were found in rotation, and scale shrink tasks ($p > 0.05$).

4.1.2 Results - Completion time

Average completion times per task and condition are reported in Table 1. The additional visual cues used in the 3D Hand Overlaid condition resulted in shorter completion times for the two Rotate and Scale Shrink tasks, however no statistically significant differences were found between the two user groups in any of the tasks for the two conditions ($p > 0.05$).

4.1.3 Results - Accuracy

Accuracy was defined as the difference between the target ideal and the interaction cube at the position left by the user after task completion. Depending on task this accuracy is reported based on: distance from the target, difference in size, difference in rotation.

- *Move - Short distance*: Euclidean distance was measured between the target and the interaction cube center points. An average euclidean distance of 4.54 cm (SD = 0.03 cm) was reported for the 3D Hand Overlaid condition while an average euclidean distance of 4.55 cm (SD = 0.03 cm) was found in the No Overlaid condition. No statistically significant difference was found between groups in this task ($U = 2288$, $p < 0.05$).

- *Rotate - Pitch (Y axis)*: Ideal target rotation in the Y axis was 70°. Average rotation for the 3D Hand Overlaid condition was 65.20° (SD = 8.43°) while for the No Overlaid the average value was reported as 69.73° (SD = 7.11°). A statistically significant difference was reported between conditions ($U = 1966$, $p < 0.05$).

- *Rotate - Yaw - (Z axis)*: Ideal target rotation in the Z axis was 50°. Average rotation for the 3D Hand Overlaid condition was 72.48° (SD = 6.45°) while for the No Overlaid condition the average value was reported as 71.49° (SD = 10.68°). No statistically significant differences were reported between conditions ($p > 0.05$).

- *Scale - Uniform scale (enlarge)*: The target size for the enlarge task was 6.30 cm. Participants were requested to duplicate the size of the interaction cube, which was presented with a 3.10 cm size. Average size for the 3D Hand Overlaid condition was 5.40 cm (SD = 1.44 cm) while for the No Overlaid condition the average value was reported as 5.93 cm (SD = 0.74 cm).A statistically significant difference was found between conditions ($U = 2298$ $p < 0.05$).

- *Scale - Uniform Scale (shrink)*: The target size for the enlarge task was 3.20 cm. Participants were requested to halve the size of the interaction cube, which was presented with a 6.20 cm size. Average size for the 3D Hand Overlaid condition was 3.58 cm (SD = 0.81 cm) while for the No Overlaid condition the average value was reported as 3.48 cm (SD = 0.45 cm). No statistically significant differences were reported between conditions ($p > 0.05$).

4.1.4 Analysis

The use of additional visual cues were not determinant in obtaining a more accurate performance. The 3D Hand Overlaid group took shorter time in locating and triggering the interaction with the object, however no conclusive significant differences were found.

4.2 Interaction metrics

Interaction metrics were defined as the subjective metrics obtained from users using observation, the System Usability Scale (SUS) [4] and a post-test questionnaire. Results from these are reported in this section.

4.2.1 System Usability Scale (SUS)

The 3D Hand Overlaid group condition obtained an average SUS score of 72.33 (SD = 16.24) while the No Overlaid group scored a 68.0 (SD = 18.68). Scores can be labeled as 'Good' for the hand overlaid condition and 'OK' for the no overlaid group [3]. These results may be linked with the shorter time to interact scored by the 3D Hand Overlaid condition.

4.2.2 Post-questionnaire

Participants were asked a set of questions to gain a better understanding of the usability of using off-the-shelf hand posture estimation sensors to enhance MR freehand interaction.

- *Most intuitive task*: 8 participants in each group reported that the Move task was the most intuitive to complete. The remaining 7 participants were divided in between the other tasks.

- *Most difficult task*: No consensus was reached for the most challenging task. In the hand overlaid condition, 5 participants considered Shrink task the most difficult, while 3 considered it was either the Enlarge or Pitch rotation. The remaining 4 were divided between the Move task (2) and the Yaw rotation (2). The No Overlaid condition was segmented differently, with 5 participants reporting the Pitch rotation as the most challenging task, 4 it as the Shrink task and the rest divided between Yaw rotation (3), Enlarge (2) and Move (1).

- *Fatigue*: 3 participants in each group reported fatigue during the experiment.

4.2.3 Observations and feedback

Participants reported on the ease of use of the device, although majority agreed that the hand recognition sensor used had a learning curve. When questioned about the gestures and interaction, several participants report on the intuitiveness of the system, as it mimicked real world interactions. The interaction with the virtual objects was triggered by a realistic and intuitive precision grasp matching the virtual object size and shape.

5 Conclusion and discussion

This paper explores freehand physical interaction in AR by performing a feasibility study on the use of hand posture estimation sensors using an egocentric orientation. This paper evaluates their potential in enhancing OST and HMD hand recognition capabilities. We report on interactivity, usability and precision following Piumsomboon *et al.* definitions [14].

Even though the No Overlaid and the 3D Hand Overlaid conditions did not differ much when reporting on precision metrics (time to interact, completion time and accuracy), there was a reported difference for the additional interaction metrics. Additional visual cues made the system being perceived as more usable, despite the use of these two different conditions having limited effect on precision and interactivity metrics.

The tasks used in this study triggered a physical virtual object manipulation, overall users found the tasks intuitive and easy to perform. These results agree with the literature where a preference was found a for physical manipulation in freehand interaction [15, 21, 22]. One of the limitations of this study is the use of a simulated immersive environment, future work will include the replication of the current study with the use of an OST device as display.

The hand posture estimation sensor used, the Leap Motion, showed potential for being included as an off-the-shelf physical freehand interaction recognition device for enhancing user interaction in egocentric MR. Its potential for enhancing freehand interaction in HMD and OST devices is promising. Recommendations for the HCI community drawn from this study are the use of intuitive physical freehand interaction, providing a realistic interaction in the virtual world or with virtual objects in alignment with what would be expected in a real environment.

References

[1] M. Al-Kalbani, I. Williams, and M. Frutos-Pascual. Analysis of medium wrap freehand virtual object grasping in exocentric mixed reality. In *Mixed and Augmented Reality (ISMAR), 2016 IEEE International Symposium on*, pp. 84–93. IEEE, 2016.

[2] M. Al-Kalbani, I. Williams, and M. Frutos-Pascual. Improving freehand placement for grasping virtual objects via dual view visual feedback in mixed reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 279–282. ACM, 2016.

[3] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[4] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[5] D. Datcu and S. Lukosch. Free-hands interaction in augmented reality. In *Proceedings of the 1st symposium on Spatial user interaction*, pp. 33–40. ACM, 2013.

[6] R. Furlan. The future of augmented reality: Hololens-microsoft's ar headset shines despite rough edges [resources_tools and toys]. *IEEE Spectrum*, 53(6):21–21, 2016.

[7] J. Jankowski and M. Hachet. Advances in interaction with 3d environments. In *Computer Graphics Forum*, vol. 34, pp. 152–190. Wiley Online Library, 2015.

[8] P.-W. Lee, H.-Y. Wang, Y.-C. Tung, J.-W. Lin, and A. Valstar. Transection: hand-based interaction for playing a game within a virtual reality game. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 73–76. ACM, 2015.

[9] S. Lin, H. F. Cheng, W. Li, Z. Huang, P. Hui, and C. Peylo. Ubii: Physical world interaction through augmented reality. *IEEE Transactions on Mobile Computing*, 16(3):872–885, 2017.

[10] K. R. Moser and J. E. Swan. Evaluation of hand and stylus based calibration for optical see-through head-mounted displays using leap motion. In *Virtual Reality (VR), 2016 IEEE*, pp. 233–234. IEEE, 2016.

[11] K. R. Moser and J. E. Swan. Evaluation of user-centric optical see-through head-mounted display calibration using a leap motion controller. In *3D User Interfaces (3DUI), 2016 IEEE Symposium on*, pp. 159–167. IEEE, 2016.

[12] T. Ni, D. A. Bowman, C. North, and R. P. McMahan. Design and evaluation of freehand menu selection interfaces using tilt and pinch gestures. *International Journal of Human-Computer Studies*, 69(9):551–562, 2011.

[13] B. Petry and J. Huber. Towards effective interaction with omnidirectional videos using immersive virtual reality headsets. In *Proceedings of the 6th Augmented Human International Conference*, pp. 217–218. ACM, 2015.

[14] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, G. Lee, and M. Billinghurst. Grasp-shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pp. 73–82. IEEE, 2014.

[15] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pp. 955–960. ACM, 2013.

[16] G. Ren and E. O'Neill. 3d selection with freehand gesture. *Computers & Graphics*, 37(3):101–120, 2013.

[17] B. S. Santos, J. Cardoso, B. Q. Ferreira, C. Ferreira, and P. Dias. Developing 3d freehand gesture-based interaction methods for virtual walkthroughs: Using an iterative approach. In *Handbook of Research on Human-Computer Interfaces, Developments, and Applications*, pp. 52–72. IGI Global, 2016.

[18] M. Sato, S. Suzuki, D. Ebihara, S. Kato, and S. Ishigaki. Pseudo-softness evaluation in grasping a virtual object with a bare hand. In *SIGGRAPH Posters*, pp. 40–1, 2016.

[19] S. Suzuki, H. Suzuki, and M. Sato. Grasping a virtual object with a bare hand. In *ACM SIGGRAPH 2014 Posters*, p. 51. ACM, 2014.

[20] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Communications of the ACM*, 54(2):60–71, 2011.

[21] C. Weichel, M. Lau, D. Kim, N. Villar, and H. W. Gellersen. Mixfab: a mixed-reality environment for personal fabrication. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 3855–3864. ACM, 2014.

[22] M. L. Wittorf and M. R. Jakobsen. Eliciting mid-air gestures for wall-display interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, p. 3. ACM, 2016.

[23] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1083–1092. ACM, 2009.

[24] Y. Yao, P.-T. Chiu, and W.-T. Fu. A gestural interface for practicing children's spatial skills. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion*, pp. 43–47. ACM, 2017.

[25] D. Yim, G. N. Loison, F. H. Fard, E. Chan, A. McAllister, and F. Maurer. Gesture-driven interactions on a virtual hologram in mixed reality. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces*, pp. 55–61. ACM, 2016.