Hand-Based Interaction for Object Manipulation with Augmented Reality Glasses

Zhenyi He*, Xubo Yang†
Shanghai Jiao Tong University

Abstract

An increase number of gesture sensors and AR(augmented reality) displays are emerging these years. There are more and more study working on spatial interaction. In this work, we focus on evaluating the feasibility of hand-based interaction designed for AR glasses to do basic manipulation on 3D virtual objects in AR applications with current interaction technology. We design a user study with 10 volunteers to discuss whether and how to use natural user interfaces like gestures, with the help of 3D menus, to finish the tasks of manipulating AR scenes. The interface including 3D menus presents the advantages of usability and merits of intuition; However, it also exposes some issues like ergonomic discomfort and limitation like imprecise input. So our results indicate the kind of interactive device and technology are valuable in natural interaction without extra knowledge, though they are limited in serious applications now.

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

Keywords: augmented reality, spatial interaction, object creation, glass

1 Introduction

In recent years a more specific topic in mobile AR field became hotter than before. Researchers focus on producing new glasses and pay attention on AR applications as well. The CPU on mobile devices (including phones and tablets) become faster than before, which means complicated computing is available either. The key point about AR device is whether current interaction methods are still useful with AR glass. Interactions in AR application could be split into several styles. Body language is most common and intuitive. Besides, marker[pie] or marker-less pattern are also interaction technologies. Although they sound plausible to interact with virtual objects, limitations are obvious.First, various kinds of cameras could help to capture the real world. The problem is suitableness. Second, whether people wearing an AR glass feel relaxing is a problem of ergonomics. The third issue is how to interact with interfaces and objects. However, despite the limitations, some work has already indicated the potential of interaction research on AR glass[Colaço et al. 2013].

These are the problems we will discuss in this paper. Our research investigates the feasibility and usability of creating virtual objects and do some basic manipulation. Our work contributes to HCI in the following ways.

- Evaluate the feasibility of hand-based interaction on manipulating virtual objects basically when people put on AR glass.
- Evaluate the usability of different styles of 3D menu designs and helpful hint in this environment.

This paper will be organized in 7 sections. Related work is in section 2, based on which we introduce our design in section 3. The whole system description is followed in section 4. User study on feasibility and usability are evaluated in section 5 and 6. In section 7, we provide summary, discussion and future work.

2 Related Work

[Looser et al. 2004] used hand-held eyeglass and hand-held pattern groups to trigger the system. It was more as a pattern tracking method. Some of previous work used pattern to avoid tracking hands, but system becomes artificial when holding some devices. Data glove in FIGI [De Marsico et al. 2012], a system with optical see-through AR glass, data glove and other complicated devices, had a comfort problem. Witt et al. [Witt et al. 2006] had a similar problem. They used the rotation data from data glove to move a cursor and rotate objects both. [Wang and Popović 2009] used a color glove to reconstruct the hand and to control virtual objects through the rigid hand model. The problem is how to grab virtual objects if they are out of range. In this work we used leap motion [Motion 2012] to help tracking the hands and manipulate objects. [Grossman et al. 2004] used a few markers setting on user's several fingers and did click action on its surface. System could get precise input though the setting is complicated. If users changed the marker's position carelessly when in use, it would cause wrong results. Compared to cumbersome devices on hands [Oakley et al. 2008], wrists or fingers [Jia et al. 2007], lightweight setting brings little uncomfort. In this work, wearable devices are almost all avoided to improve user experience.

Comparing with one-camera or multi-camera setting, a depth camera could help to track hand articulations [Oikonomidis et al. 2011]. Given that depth camera is too heavy for users to wear and high power consumption, Mime [Colaço et al. 2013] used 3 TOFs to detect one fingertip by signal processing. Mime applied the interaction technique to applications like Fruit Ninja or Draw Something. But current version is limited with single finger. Brainy hand [Tamaki et al. 2009] had both single-hand and double-hand interaction, and was designed to be lightweight enough to be wear every day. However the system show limited effect in 3D and user interfaces are just like desktop applications. Our work designed both single-hand and double-hand interaction to compare the effects, and created different types of menu to improve the sense of reality. We chose 2 USB cameras as input. This option was not necessarily better, but can provide stereo vision.

Natural user interaction has a great difference from keyboards and mice. It has no obvious switch like click action. Some previous [Looser et al. 2004] work used button. [Zhou et al. 2011] worked on how to select target with help like color ring and markers on the books. [Lee and Woo 2010] used marker cubes to display menus and to click the menus by rotating the cubes. [Hürst and Dekker 2013] used two markers, fixed on user's index finger, to differentiate status by covering one or two markers. Unfortunately, people often forgot to take care of two markers. We could roughly

*e-mail:snowymohehe@gmail.com
†e-mail:xubo.yang@gmail.com

VRCAI 2014, November 30 – December 02, 2014, Shenzhen, China.

http://dx.doi.org/10.1145/2670473.2670505

draw the conclusion that the switch between status must be obvious. In this work, we have designed a color ring to discern the status.

[Hürst and Dekker 2013] enabled users to create complicated models in a tracking-based system. Because field of view was limited and hand was always behind the phone, it aggravated the difficulty of tasks. Some work focuses on eye-free system with sound feedback. DigiTap [Azenkot et al. 2013] implements an eye-free input method with the feedback of sound . The work of Brewster et al. [Brewster et al. 2003] was similar It raised the issue that sound feedback is possible to disturb the communication in the real world. Though smart phones are ubiquitous now, limitations like not hands-free, fingers covering large parts of screen, not easy to control small objects and small field of view still have negative effects on interaction. Last year, Maimone et al. [Maimone and Fuchs 2013] designed a computational optical see-through AR glass with multiple layer strategy. In this work, we decided to construct an AR glass to help users observe the virtual scene and the real world with the same cameras, avoiding the coherence problem.

3 Interaction Design

Figure 1: *AR glass configuration: with the leap motion (front), the USB cameras (middle) and Oculus (rear)*

3.1 Design Principles

- **No cumbersome sensors on hand**

  It would be easier to track user's hand motion if wearing sensors on hands. However, it's uncomfortable and unnatural against our goal. So bare-hand interaction is what we want.

- **Keep few gestures**

  While the larger the amount of gestures are, the more difficulties users would face when using the system. So the amount and whether it is easy to control are things really matter.

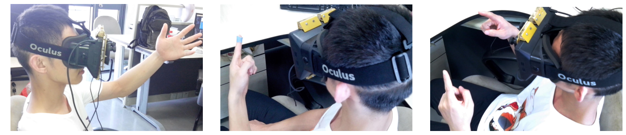
- **Keep obvious status switching**

  As mentioned in section 2, [Hürst and Dekker 2013] lead to some confusing result. So obvious status and easily controlling different status are necessary.

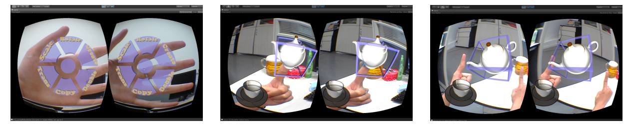
The whole system was based on a video see-through AR glass with Oculus, two USB cameras, two wide-angle lens and leap motion as Figure 1 showed. User could see the image from Oculus, and the real world scenes through two cameras mounted on the Oculus.

3.2 Nearly Bare-hand Interaction and Menu Design

We use leap motion to interact. To avoid introducing other mechanism to differentiate status. We designed a color ring.To help user memorize the specific operation, we design only three gestures. **Menu Action:** the AR menu would show up right in the hand with 5 extended fingers. **Selection Action:** the system would execute click operation with color ring after specific time. **Manipulation Action:** the system is in manipulation mode when users use one or two



(a) menu action in the real world (b) selection action with color ring (c) manipulation action



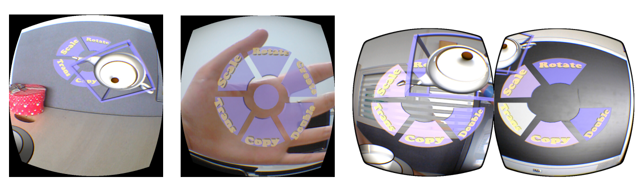
(d) menu action in the ARGlass (e) selection action with color ring (f) manipulation action

Figure 2: *The interaction design*

hands with one extended finger. People could change the selected object's position, size and orientation according to the user's move. That color ring emerges means the system is in selecting mode, and vice versa. In this work, manipulation meant basic operations including translation, scaling and rotation.

Figure 2(a) and 2(d) presented menu action.We could see *palm-based menus* are right in the hand. Selection action is shown in Figure 2(b) and 2(e), bounding box shows up when being selected. Manipulation action shown in Figure 2(c) and 2(f) indicates that user is rotating the teapot to pour water into the cup. We designed the menus like a palette attached to people's hand for the shape. It seemed like people have a magic palette within their hand, and perform magic on selected objects by choosing different magic in the palette.

Figure 3 shows three menus. One is based on operation itself and the others are based on targets. *Palm-based menus* enable people to select menus before selecting the targets, while *object-tracked menus* require people to choose object at first and then menus will appear for people to select. The third type of 3D menu, screen-fixed menu, which forever stays at the same size and same place right in the center as Figure 3(c) shows.

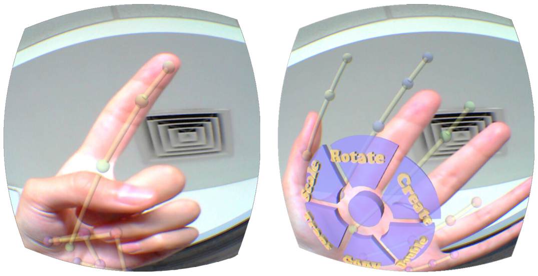


(a) *Palm-based* (b) *Object-tracked* (c) *Screen-fixed menus*

Figure 3: *Three kinds of menus*

3.3 skeleton balls

We added skeleton balls as a visualization of fingers in Figure 4. Same color means same finger. No balls means no hands are detected, probably because they are out of range. Then users could observe the placement to figure out what leap motion captures. In this work, we put skeleton balls where people's hand are.



(a) Manipulation action (b) Menu action

Figure 4: *skeleton balls represent the finger in leap motion.*

4 System Implementation

First, we calibrated cameras respectively and stereo calibrated them later. Then we project images into the ARGlass as background. That's how video see-through glass made. Second, we put two virtual cameras in the scene and the settings are how real ones place. The system put virtual scene in front of background, so users would obtain a mixed reality result. Third, we detect gestures according to color image from two USB cameras and hand model from leap motion. For menu action, if there are more than four valid straight fingers in the model, it is menu action. For selection action, we check two USB color images to see if there's a blue square in the scene. Then we combine skin area and blue square to get a whole contour. If it is a whole contour, the blue square is the color ring wore on user's finger. For manipulation action, if the system has already checked the color ring, the action is definitely a selection action. If there's no color ring, the system use hand model to detect valid fingers. Then the system obtains the position and orientation of fingers to calculate the detail operation. User controls the object to rotate along the trajectory of valid finger and the angle is related to the length of trajectory. Last, after getting knowledge of current action, we just apply the command to the menus or objects.

5 Feasibility Study

We set up an survey to evaluate the feasibility. Five subjects(from 22 to 25, average 23.4) participated. They acquired the knowledge of system and then got training.

Training steps: Use *palm-based menus* to create an object and rotate it. Select the object and use *object-tracked menus* to scale it. Repeat it with different menu types. **Task steps:** Step 1, create a teapot and cup; Step 2, RST the teapot to fake pouring away the water with single-hand and double-hand manipulation; Step 3, copy these objects with two menus. Training time varied from 6 to 11 minutes and two experiments took about 20 minutes. A questionnaire was offered later.

All participant finished the training successfully and one failed in the third task. We found people were poor at manipulating with leap motion since no one used it before. Figure 5 shows the results of feasibility.It's easy for users to call different types of menus and select menus or objects, either. One problem was about the precise manipulation. People's hands and heads were moving slightly all the time, not so stable as a cursor. Another problem is about *object-tracked menus*. People found it impressive to see the menus following the selected object. However, menus would change its position with the objects. It's not easy to control both their heads and hands at one time with new device.

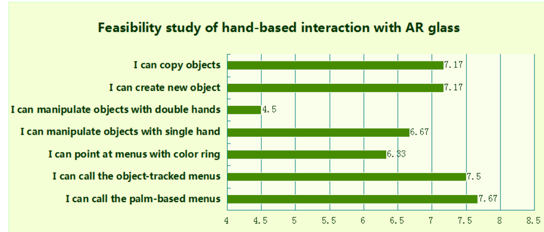


Figure 5: *Questionnaire results of the first version about menus actions, selection actions and manipulation actions*

6 Usability Evaluation

We conducted a formal user study with 10 subjects, including only 2 females because they may have different experiences. In this experiment, we need to figure out our hypotheses. What kind of menus do participants prefer most? Whether the preferences vary from person to person or from task to task. Whether skeleton balls help users to finish their tasks. What people is better at in AR applications, single-hand or double-hand interaction.

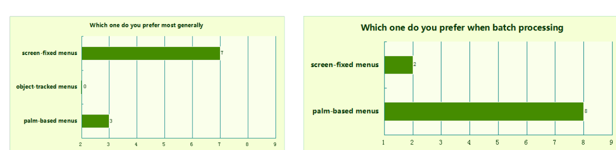
Training process is split into 3 steps. Step 1.warm up with leap motion. Step 2, call three types of menus to create an object and manipulate it respectively. Step 3, create a new object with skeleton balls. After training, we started the formal tasks.

6.1 Tasks Introduction

Task one: Participants create two virtual teapots and manipulate by single hand either with skeleton balls showing or not. Then participants decide whether to enable the skeleton balls. **Task two:** Participants are asked to create one virtual object and scale it with *object-tracked menus* and *screen-fixed menus*. After this task, participants need to choose one style of these two menus. **Task three:** Participants need to copy three objects and rotate each to face right. **Task four:** There are three virtual objects A, B, C in the AR scene. A is used as a reference. Participants are asked to use single-hand manipulation to manipulate B to be similar as A in the orientation, and use double-hand manipulation to do the same on C.

6.2 Results

Almost all users mentioned that the whole device, 522g, is too heavy to wear. Since there were only three different gestures, no one got confused. We asked whether they still remembered all gestures several days later, and every one do.



(a) general situation (b) Menu action

Figure 6: *Choices among three styles of menu.*

Skeleton balls made the whole scene vivid.The result of question whether users want to keep the visualization shows the usability of skeleton balls. 6 out of 10 chose yes, although some felt a little strange when they saw the skeleton ball, they still found it helpful to manipulate the objects. And average time of every task with skeleton ball is 44 seconds which is shorter than 70 seconds without skeleton ball. One participant who does't need the skeleton ball said that the emerging of skeleton ball disturb the whole scene. Also other participants advised that pick another place to put the skeleton ball or it's enough just tell the user whether their hands are now in range of leap motion or not.

One objective of this work is to figure out which menu people prefer and would they change their ideas in different contexts. Figure 6(a) presents that no one preferred the object-tracked menu. When in task three, all people choose the latter one between *object-tracked menus* and *screen-fixed menus*. Because of slight motion of head and hands, virtual objects with *object-tracked menus* hardly stayed static in people's view. People said though the menu group which track the object was a meaningful design, when the object was far away from the center of view, they needed to move their heads to follow the menu group. That's why they preferred the *screen-fixed menus*. The number of choices between *palm-based menus* and *screen-fixed menus* were almost the same in task four. Some participants determined the target first so they prefer *screen-fixed menus*, while others did the opposite. When they met batch processing tasks in task three, most people chose *palm-based menus* due to fewer steps. Figure 6(b) indicated that more people prefer *palm-based menus* than *screen-fixed menus* which contradicts previous result. So in this scenario, *palm-based menus* are more suitable.

The topic usability between single-hand and double-hand interaction was interesting. The design of double-hand interaction was like natural interaction in the real world. People felt hard to control it mainly because they were not used to making gestures in the air.9 out of 10 favored single-hand manipulation. They felt that it was just like doing operation on a touch screen, and it was cool they don't need to hold the screen since AR glass is hands-free. The obvious defect in single-hand interaction was command changes. In task 4, there are different difficulties in two among user feedback. Single-hand interaction was not intuitive enough, but double-hand one was not friendly enough because of the unfamiliar sensor. In addition, no matter using single hand or double hands, people feel not easy to align one object to the reference one.

7 Conclusion and Future Work

Briefly, there are two contributions in this work. Evaluate the feasibility of hand-based interaction with AR glass and discuss how menus and interaction styles impact user experience. Hand-based interaction is feasible in AR system in section 5. The investigation between single-hand and double-hand interaction showed that the former is similar to interaction on touch screen.The later imitates a lot from natural world. How to balance the merits of intuition and practicality is next step. In order to keep manipulation both intuitive and valuable, we could add a smart align help. In future, we will make efforts on methods to counteract the movement of the head, or locate the hands' relative position to increase the precision of manipulation. We will design more different menus and applications to see which is fit for various kinds of application.

Acknowledgements

To Guosheng He, he helped a lot creating the AR glass. To DALAB, people took part in the experiments and give us helpful advice.

References

AZENKOT, S., BENNETT, C. L., AND LADNER, R. E. 2013. Digitaps: eyes-free number entry on touchscreens with minimal audio feedback. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ACM, 85–90.

BREWSTER, S., LUMSDEN, J., BELL, M., HALL, M., AND TASKER, S. 2003. Multimodal'eyes-free'interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 473–480.

COLAÇO, A., KIRMANI, A., YANG, H. S., GONG, N.-W., SCHMANDT, C., AND GOYAL, V. K. 2013. Mime: compact, low power 3d gesture sensing for interaction with head mounted displays. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ACM, 227–236.

DE MARSICO, M., LEVIALDI, S., NAPPI, M., AND RICCIARDI, S. 2012. Figi: floating interface for gesture-based interaction. *Journal of Ambient Intelligence and Humanized Computing*, 1–14.

GROSSMAN, T., WIGDOR, D., AND BALAKRISHNAN, R. 2004. Multi-finger gestural interaction with 3d volumetric displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, 61–70.

HÜRST, W., AND DEKKER, J. 2013. Tracking-based interaction for object creation in mobile augmented reality. In *Proceedings of the 21st ACM international conference on Multimedia*, ACM, 93–102.

JIA, Y., LI, S., AND LIU, Y. 2007. Tracking pointing gesture in 3d space for wearable visual interfaces. In *Proceedings of the international workshop on Human-centered multimedia*, 23–30.

LEE, H., AND WOO, W. 2010. Tangible spin cube for 3d ring menu in real space. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 4147–4152.

LOOSER, J., BILLINGHURST, M., AND COCKBURN, A. 2004. Through the looking glass: the use of lenses as an interface tool for augmented reality interfaces. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, 204–211.

MAIMONE, A., AND FUCHS, H. 2013. Computational augmented reality eyeglasses. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, 29–38.

MOTION, L. 2012. Leap. *URL: https://www. leapmotion. com.*

OAKLEY, I., SUNWOO, J., AND CHO, I.-Y. 2008. Pointing with fingers, hands and arms for wearable computing. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 3255–3260.

OIKONOMIDIS, I., KYRIAZIS, N., AND ARGYROS, A. A. 2011. Efficient model-based 3d tracking of hand articulations using kinect. In *BMVC*, 1–11.

TAMAKI, E., MIYAKI, T., AND REKIMOTO, J. 2009. Brainy hand: an ear-worn hand gesture interaction device. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, 4255–4260.

WANG, R. Y., AND POPOVIĆ, J. 2009. Real-time hand-tracking with a color glove. In *ACM Transactions on Graphics (TOG)*, vol. 28, ACM, 63.

WITT, H., NICOLAI, T., AND KENN, H. 2006. Designing a wearable user interface for hands-free interaction in maintenance applications. In *PerCom Workshops*, 652–655.

ZHOU, Y., DAVID, B., AND CHALON, R. 2011. Innovative user interfaces for wearable computers in real augmented environment. In *Human-Computer Interaction. Interaction Techniques and Environments*. 500–509.