Virtual vs. Real-World Pointing in Two-User Scenarios

Holger Salzmann*
Carmeq GmbH, Member of Volkswagen Group

Mathias Moehring†
Volkswagen AG Wolfsburg

Bernd Froehlich‡
Bauhaus-Universitaet Weimar

Abstract

We are investigating the utility of a projection-based stereoscopic two-user system for applications in the automotive industry. In this paper we compare real-world pointing to pointing at virtual objects in a two-user scenario. In our study we investigated the following situation: One user points at an object while the second person has to guess the referenced object. Our results indicate that pointing at objects in a virtual world is much less precise than pointing in the real world even for a well calibrated stereoscopic two-user system. Pointing techniques like outlining an object or pointing from a distance produce more errors than the corresponding real-world techniques, but they are less error prone than interactions requiring to touch a virtual object. Our findings are a first step to qualify direct interaction techniques in a multi-user projection-based system.

Keywords: mixed reality, multi-user systems, collaborative virtual environment, human-machine interfaces, co-location.

Index Terms: I.3.1 [Computer Graphics]: Hardware Architecture—Three-dimensional displays; I.3.6 [Computer Graphics]: Methodology and Techniques—Ergonomics; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

1 Introduction

Perspective projection in combination with head tracking is widely used in immersive virtual environments to support users with correct spatial perception of the virtual world. Recently projection-based stereoscopic multi-viewer systems have become available [9], which provide individual stereoscopic images for multiple tracked users. This display technology is well suited for the automotive industry where collaborative 3D-interaction of a group of experts is often desired. Consequently, we set up a two-user stereoscopic display and started to experiment with virtual automotive simulations (Figure 1), where one user often needs to communicate specific information about a particular part of the car to a second person. Our users intuitively used their hands to point at a car part, which worked well in most cases. However, our experts were wondering how accurate is pointing in virtual environments in comparison to real-world pointing?

To answer this question, we designed a study to compare three different pointing techniques: touching an object, drawing an outline around an object and pointing from a distance. These pointing techniques were evaluated for a number of different objects on a real dashboard as well as on an identical virtual dashboard. Our findings indicate that all these techniques are more error prone in virtual environments than they are in the real world. In particular, directly touched virtual objects smaller than two centimeters can not be safely identified, which is trivial in the real world for most situations. The stereoscopic two-user system works quite well for

*e-mail:holger.salzmann@carmeq.com
†e-mail:mathias.moehring@volkswagen.de
‡e-mail:bernd.froehlich@medien.uni-weimar.de

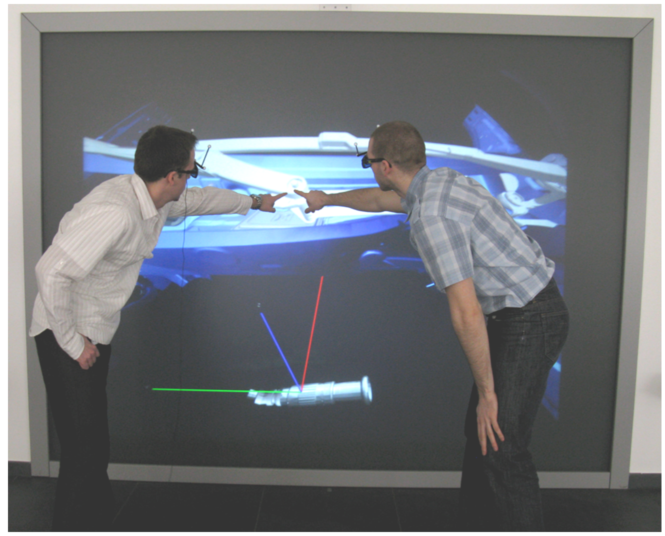

Figure 1: Users are discussing collisions of a windshield wiper joint in a two-user projection system for collaborative assembly simulations.

common assembly scenarios involving larger objects. But, when it comes to interacting with a fully functional virtual navigation device equipped with small buttons, we recommend the use of a virtual pointer representing the user's finger.

2 Related Work

The research topics touched by our work are multi-viewer projection systems, interaction with three-dimensional user interfaces augmented by real-world objects, correct stereo perception and multi-user scenarios in collaborative virtual environments (CVEs).

Froehlich et al. [9] described some of the fundamental view separation techniques and possibilities to combine these techniques. Our setup is based on a combination of shuttering and polarization to separate the four different images for the two users.

Collaborative virtual environments (CVE) support the interaction of multiple remote or co-located users in a shared virtual environment. Our system has some similarities with a number of systems like the Studierstube project [13], the PIT [2] and the Two-User Responsive Workbench [1]. Two or more users are co-located in the same room and experience the same virtual scene. All those systems provide an individual stereoscopic view for each user, support head tracking, and the perception of the surrounding real environment.

Various studies on supporting social human communication in CVEs cover verbal and nonverbal communication as well as references to objects and the environment [12] [6] [10]. In our experiment on pointing at virtual objects with real hands we artificially excluded verbal communication between subjects [3]. Furthermore our experimental virtual scenario does not make use of any surrounding real environment. We only focus on nonverbal communication and referencing to objects by pointing gestures of the real hand which is in contrast to studies on virtual pointers [5].

Drascic and Milgram [8] report on perceptual issues that should be taken into account when working with stereoscopic displays. They describe several depth cues that influence stereoscopic viewing and perception. The influence of false or uncontrolled depth cues can be compensated by other consistent depth cues in the VR-system. The human visual system is able to adapt very quickly to mis-calibrated systems. However, if the user is forced to switch between real and virtual objects, this may lead to disorientation and misperception. This is also an issue with our experiment on pointing and will be discussed later.

3 Experiments On Pointing

There are different ways how people identify an object by pointing when they try to communicate some information about that particular object.

- **touching:** In reality it is very easy for observers to identify an object which is directly touched by someone. In projection-based display systems one would assume that the stereoscopic perception is affected if the real hand is in proximity to a virtual object due to conflicting occlusion as well as focus and convergence cues. Thus it should be hard to estimate if a real finger is in front of an object, inside or already behind.

- **outlining:** It is slightly harder to identify real objects by drawing their outline in the air with little distance to the corresponding object. It becomes more difficult if more objects of the same shape and size are positioned close to each other but it can be assumed that the same effect occurs in virtual environments.

- **distance pointing:** The most difficult type of pointing in the real world and in virtual reality is pointing at objects from a distance, in particular if objects lie close to each other.

In our study we focus on these three selection techniques while there are certainly other possibilities to reference a particular object.

3.1 Experiment Design

We had to take several aspects into account to find a good scenario for our pointing task. The object had to match our projection setup which supports only a single projection screen. In various scenarios we often faced the problem that virtual objects were clipped at the boundaries of the projection screen due to missing floor and side projections. We identified the dashboard as most suitable because everyone knows most of its elements by name and uses them in every day life. Some of the dashboard items are even labeled. Furthermore the dashboard matches the size of our projection screen, has objects at several depths, and it includes large objects as well as small objects of different shape.

We set up a real dashboard vs. virtual dashboard comparison task. Two participants were asked to stand in front of the dashboard. One person pointing at objects; the other person guessing which object was referred to. This task was performed in front of the real and the virtual model. Each participant had to take over both roles: pointing and guessing. Each pointing participant had to point at three touch objects, three outline objects and three distance objects of the virtual and the real dashboard. The target objects were never repeated during one complete session and they were different for the real and virtual dashboard. That means one pair of participants had a set of eighteen objects from the three different classes. Orders of objects randomly changed between pairs of subjects and half of the groups started with pointing at the virtual dashboard and the other half started with pointing at the real dashboard to avoid learning effects.

During the pointing task subjects were not allowed to move except for arm and small head movements. An instructor showed the pointing participant the target objects he had to point at on a picture of the dashboard. No spoken commands were uttered until the end of the task. Only the observing participant was allowed to say which target object he identified. The instructor recorded the matches and mismatches.

Pointing at virtual objects should be strongly impacted by focus and convergence mismatch and an error in pointing can be expected as well as an error in perceiving finger positions by the observer. To identify possible sources of errors in pointing at virtual objects we mounted a small tracking target on the index finger of each pointing participant. This allowed us to calculate the distance between real fingertip and a virtual object that was supposed to be touched. This approach is similar to e. g. Corradini and Cohen who conducted an inspiring experiment on the precision of pointing [6]. They let subjects point at certain targets on a wall from different distances with a turned off laser pointer. Then they turned the laser pointer on and measured the distance to the target on the wall.

Twentyfour male subjects aged 25 to 45 volunteered to participate in our study all having different backgrounds like assembly planners, VR-researchers and students with different VR experience. They performed the pointing tasks in pairs of two and filled out a post questionnaire afterwards. The questions aimed at demographic issues, VR-experiences and the pointing task itself.

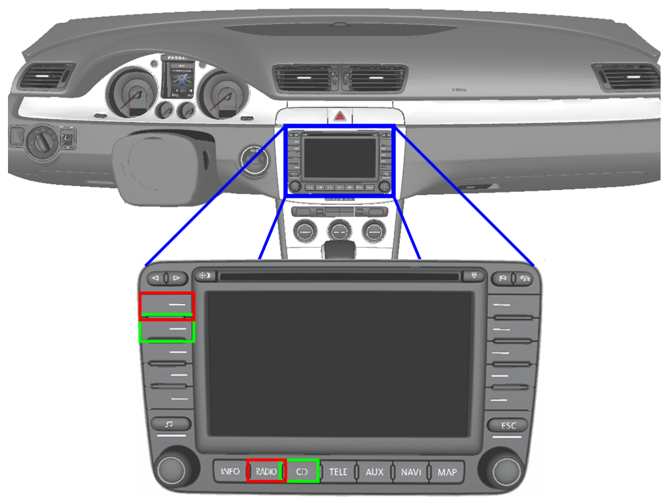

Figure 3: The buttons on the navigation system were the smallest objects in our pointing task. This area of the dashboard was displayed in the middle of the screen, which was also the best calibrated area of our four-projector system. The buttons with red frames are common mismatches for the neighboring buttons marked by green frames.

3.2 Calibration

A high quality of our virtual environment setup was essential for our study. Especially projector matching and tracker calibration had to be done very carefully. Introducing errors in VE calibration would have distorted the results of our evaluation. We defined our virtual environment by carefully measuring the hardware setup. We did not adjust inter-pupillary distance (IPD) individually but relied on an average distance of 65 mm [7]. This is also the case in our daily work, where groups of people use our system and glasses are passed around in such groups. We were interested to find out which accuracy we can expect from such a realistic use of our environment. The projectors were aligned manually, introducing a slight mismatch at the image boundaries due to limited adjustment capabilities. In the middle of the projection screen the projections were perfectly aligned. We placed the smallest and most detailed objects for our pointing task, e. g. the buttons of the navigation device (Figure 3), in the central screen area. We tested our calibration

Figure 2: The right picture shows the pointing experiment in reality. One subject points at a given target object while the other has to decide where he is pointing at. The left picture shows the corresponding virtual pointing condition.

before starting the evaluations by visually comparing both users' projection having the same camera pose and by visually checking registration of a tracked object. There was no eye-to-eye or user-to-user mismatch visually detectable in the screen area used for our evaluation. However, we did not perform a detailed error analysis. The real dashboard scenario was placed on a table right next to the projection. The virtual scenario was set up in a way that it appears at the same height and depth from the participants. The pointing participant took a standing position in front of the left part of the dashboard, with the majority of the objects in reach of the right arm, and the observing participant took a standing position in front of the dashboard's right part (Figure 2). Their distance to the dashboards was 50 cm. Both had a distance of 1.2 m to the wall of the room which corresponds to their distance to the projection wall.

3.3 Results

The pointing accuracy measurements for the interactions requiring to touch an object are a good indication how well a pointing participant could localize his finger in relation to a virtual object. Based on the statements in the questionnaire regarding VR-experience, taken from the scaled questions mentioned above, we created three groups among the participants with respect to their use of virtual reality technology: "often", "sometimes" or "seldom". We calculated the mean pointing accuracy for each group by computing the distance between each virtual objects' face-center closest to the participant and his index-fingertip. Virtual object size and finger size should have been taken into account but we expected the users to tend to point always in the middle of the object. The results are shown in Table 1.

Table 1: Means for the accuracy of matches between real index fingertip and virtual target object among the three subject groups.

vr usage	pointing accuracy in cm
often	2.36 ($\sigma = 1.68$)
sometimes	3.31 ($\sigma = 1.06$)
seldom	3.64 ($\sigma = 1.59$)

Table 1 indicates that regular users of VR-systems are more accurately pointing than those participants with less VR-experiences. However, the results are not statistically significant.

For each participant we calculated the mean pointing accuracy in the pointing task. These values served us as a control measure to decide if the pointing person did not point at the right object or if the observer did not see the correct object. In other words the mean pointing accuracy helped us to classify how accurate the pointing person really was. We also counted the mismatches the observing subjects made when trying to figure out where the pointing person is pointing at for both conditions, real and virtual, over the three pointing object classes. The results are shown in Table 2.

Table 2: Mismatches in percent over the three pointing object classes for virtual and real dashboard.

object class	real dashboard	virtual dashboard
touch	0.0	30.0
outline	1.6	11.6
distance	13.3	25.0

As expected it was very simple for the observer to decide where the pointing person is pointing at when directly touching the object on the real dashboard with his real finger (0.0% of mismatches). It was also simple for the observers to identify real objects for which the outline was drawn in the air (1.6% of mismatches). For the objects pointed at from a distance the subjects had 13.3% of mismatches. However, one has to consider that no verbal communication was allowed.

For the virtual condition the observers generated 30.0% mismatches for directly touched objects. This is not surprising since we expected that it would be hard for the pointing participants to judge the position of their real finger in relationship to the virtual objects due to focus and convergence mismatch and occlusion effects. It is difficult for the pointing users to estimate the distance of their finger to the object since the finger occludes the object that should be touched. This effect can be seen by the values measured for pointing accuracy. It is also described in [8] that if people see a displayed object in context with their own body their perception changes. If real hand and virtual object are in close proximity the disparity depth cue will always deliver the information that the object is in front of the hand. At the same time the accommodation depth cue tells the user that hand and object are at different depths.

Our measurements indicate that the observers' mismatches are mostly due to inaccurate pointing of the pointing users. It must be emphasized that all of the mismatches leading to the 30% error rate for directly touched objects had a deviation of maximal one centimeter from the object's center like the example in Figure 3 showing small buttons. In these cases the mean pointing error was already larger than the object size. There were also mismatches that repeated between subjects. Surprisingly, ten of the subjects made no mistakes when guessing directly touched virtual objects. This is a significant fraction of our 24 participants. This observation needs further investigation.

The outlined virtual objects were identified with 11.6% of mismatches and the objects, which are pointed at from a distance, with 25.0% of mismatches. These error rates indicate that it is easier to identify objects that are outlined from a distance than objects that are pointed at from a distance without drawing an outline, no matter if the pointing happens in real life or in virtual reality. The participants also had to rate how well they were able to estimate the position of their real index fingertip when pointing at virtual objects. They answered on a 1 (very good) to 5 (very bad) likert-scale and the mean answer was 2.65 ($\sigma = 0.98$) suggesting that this is not an easy task. Overall participants liked to point at virtual objects with their real hand: the mean answer was 2.25 ($\sigma = 0.85$).

4 Conclusions And Future Work

We presented first results of an initial pointing experiment comparing real world pointing to pointing in a virtual environment. In this experiment, participants were not allowed to verbally describe the object in addition to pointing at it, which is a somewhat artificial situation. The results indicate that outline pointing was the most secure way to help the observer to identify the correct object. We need to further investigate interactions that include directly touching virtual objects with the real hand. In our experiments these touch-based selections resulted in the largest number of mismatches, but there were also 40% of the users who could correctly identify the targets in this situation. Distance pointing is prone to error in reality as well, but it performs even worse in virtual scenarios. During informal tests of our two-user system, users pointed at objects and verbally described them as well. In these situations we did not see any pointing inaccuracies. The famous "put-that-there" [4] approach seems to work quite well even for smaller objects. However, this needs to be further investigated as well as the exact thresholds at which the different pointing techniques start to fail.

The possibility of pointing with the real hand at virtual objects enables the most natural and intuitive interaction metaphor which is highly appreciated in our collaborative virtual scenarios. The two-user stereo display ensures that both participants of a VR-simulation discuss about the same objects or problems at any time. Nevertheless our findings show that very small objects that are directly touched can not be identified without the chance of a mismatch. Scenarios involving objects larger than a few centimeters seem to be a good choice for our projection-based multi-user system. These findings are a first guideline for implementing multi-user interactions. However, much further research is needed to identify the factors that influence the accurate perception of such multi-user interactions in stereoscopic multi-viewer displays.

References

[1] M. Agrawala, A. C. Beers, B. Froehlich, P. Hanrahan, I. McDowall, and M. Bolas. The Two-User Responsive Workbench: Support for Collaboration Through Individual Views of a Shared Space. In *Proceedings of Computer Graphics 1997 (SIGGRAPH '97)*, pages 327–332, 1997.

[2] K. Arthur, T. Preston, R. M. Taylor, F. P. Brooks, M. C. Whitton, and W. V. Wright. Designing and Building the PIT: a Head-Tracked Stereo Workspace for Two Users. Technical report, University of North Carolina at Chapel Hill, 1998.

[3] A. Bangerter and D. M. Oppenheimer. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture*, 6(1):85–102, 2006.

[4] R. A. Bolt. Put-That-There: Voice and gesture at the graphics interface. In *Computer Graphics and Interactive Techniques*, volume 14, pages 262–270, 1980.

[5] J. Chastine, K. Nagel, Y. Zhu, and M. Hudachek-Buswell. Studies on the Effectiveness of Virtual Pointers in Collaborative Augmented Reality. In *Proceedings of IEEE Symposium on 3D User Interfaces 2008*, pages 117–124, 2008.

[6] A. Corradini and P. R. Cohen. Multimodal Speech-Gesture Interface for Handfree Painting on a Virtual Paper Using Partial Recurrent Neural Networks as Gesture Recognizer. In *Proceedings of the International Joint Conference on Neural Networks*, pages 2293–2298, 2002.

[7] N. A. Dodgson. Variation and extrema of human interpupillary distance. In *Proceedings SPIE*, volume 5291 of *Stereoscopic Displays and Virtual Reality Systems*, pages 36–46, 2004.

[8] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In *Proceedings SPIE*, volume 2653 of *Stereoscopic Displays and Virtual Reality Systems*, pages 123–134, 1996.

[9] B. Froehlich, R. Blach, O. Stefani, J. Hochstrate, J. Hoffmann, K. Klueger, and M. Bues. Implementing Multi-Viewer Stereo Displays. In *WSCG (Full Papers)*, pages 139–146, 2005.

[10] O. Otto and D. Roberts. Importance of Communication Influences on a Highly Collaborative Task. In *Proceedings of the 7th IEEE International Symposium on Distributed Simulation and Real Time Applications*, pages 195–201, 2003.

[11] K. Riege, T. Holtkamper, G. Wesche, and B. Froehlich. The Bent Pick Ray: An Extended Pointing Technique for Multi-User Interaction. In *3DUI '06: Proceedings of the 3D User Interfaces*, pages 62–65. IEEE Computer Society, 2006.

[12] D. Roberts, R. Wolff, O. Otto, D. Kranzlmueller, C. Anthes, and A. Steed. Supporting Social Human Communication between Distributed Walk-in Displays. In *Proceedings of VRST'04*, pages 81–88, 2004.

[13] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavari, L. M. Encarnacao, M. Gervautz, and W. Purgathofer. The Studierstube Augmented Reality Project. In *Presence Teleoperators and Virtual Environments*, volume 11, pages 33–54, 2002.