Design of a Practical TV Interface for Teacher-Guided VR Field Trips

Jason W. Woodworth Christoph W. Borst

University of Louisiana at Lafayette

Abstract

We discuss the development of an educational teacher-guided VR environment and address communication problems noticed during formative evaluations of a teacher's interface. We especially describe motivations and problems related to a TV-based virtual mirror interface, and we present a study of 3D pointing in the virtual mirror. Efforts to develop a practical interface for repeated classroom use led to a TV interface that allows the teacher to stand unencumbered in front of a large TV showing their depth-camera-based image in a surrounding virtual environment. The limited field of regard of the TV required adressing several problems. First, to support pointing at virtual objects in an environment where the teacher stood in front of or beside virtual objects, we used the mirror-type view with a wide field of view. Visual pointing cues were added to correct problems observed with teacher pointing related to the indirectness of pointing in a mirror and to a low sense of depth. Additional visual cues allow the teacher to make eye contact with networked immersed students, considering the mirror view does not provide a direct view of student avatars. The development and evaluation of visual pointing cues provides a basis for better understanding and improving 3D pointing with virtual mirrors.

Index Terms: H.1.2 [Models and Principles]: User/Machine Systems—Human Factors I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality

1 Introduction

We discuss the design, development, and testing of an interface made to allow a teacher to guide multiple VR-immersed students through an educational virtual environment. To make the application as immersive as possible for students, we sought to promote teacher presence and teacher-student communication. Practicality and long-term teacher comfort were also important, as repeated teacher sessions are expected. The result is a "virtual mirror" interface, reminiscent of 2D video-based mirror worlds [14], but with a 3D depth camera representation rather than 2D video techniques. The teacher stands unencumbered in front of a large display and is sensed by a depth camera (Kinect), preserving the teacher's classroom oversight. The depth camera data is used to provide a 3D mesh view of the teacher to the students (Figure 1). Networking and rendering aspects of the system were described elsewhere [6].

The teacher's interface was developed through several iterations of informal testing at demonstrations, consisting of real subject-area experts guiding students through a virtual solar plant (modeled after a real solar plant [4]) to describe its fluid loops, devices, and related concepts. An early version immersed the teacher with an HMD along with the students. This made communication through ray-based pointing a simple task. However, because teachers may need to guide multiple sessions of students or maintain classroom oversight, we considered that another display type may be better to ensure teacher comfort and minimize possible effects such as motion sickness. Additionally, we considered that a realistic live view of the teacher, including the teacher's unoccluded face, could improve the student experience. Thus, unencumbered interfaces were investigated.

Using a CAVE system was considered; the wide field of regard would allow for easy pointing and monitoring of all students. However, using such equipment is not feasible in most educational settings, due to device and maintenance costs, and other difficulties of integration into a classroom.

A more practical version had the teacher stand in front of a TV and Kinect with a first-person perspective. This facilitated pointing at objects in front of the teacher, but the single-sided display did not support direct pointing at objects surrounding the teacher, where most objects of interest were found in our test application (virtual field trip of a solar plant). Thus we provided a "virtual mirror" mode for pointing, giving the teacher a viewpoint that matched a default student viewpoint with a wider field of view. The full front-facing mesh of the unencumbered teacher is able to be streamed to the immersed students to enhance their sense of teacher presence.

This interface was more appropriate for simple and comfortable repeated use, but notable communication problems were initially observed. The indirect view created problems related to pointing: teachers tended to use 2D-style "weatherman" pointing, which confused students who experience additional depth cues or varying

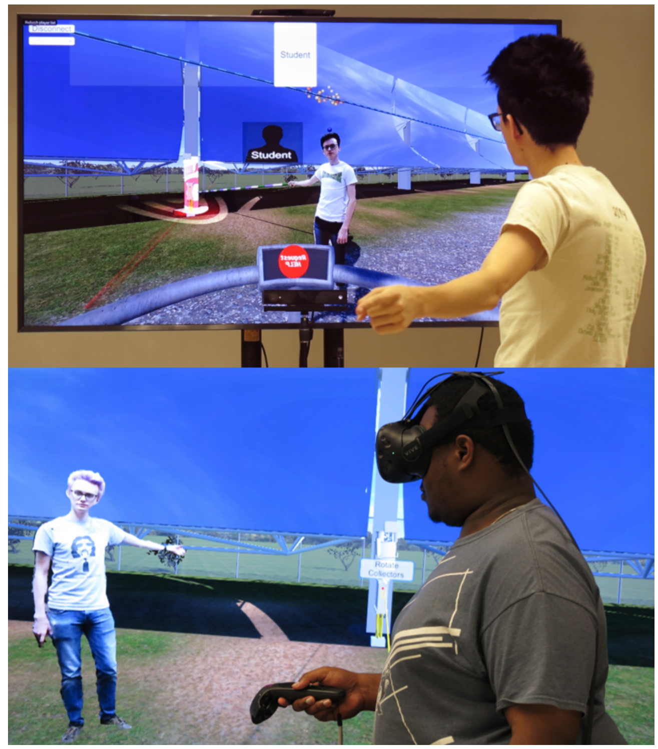

Figure 1: Teacher and Student interfaces. The teacher (top) stands in front of a large TV and Kinect and points out objects during a virtual field trip of a solar plant. The student (bottom) is immersed in an HMD and activates interactable objects according to teacher guidance.

This is an author-formatted version. Original publication: J. W. Woodworth and C .W. Borst, Design of a Practical TV Interface for Teacher-Guided VR Field Trips, Workshop on Everyday Virtual Reality 2017, pp. 29-34; Los Angeles, CA, USA ISBN: 978-1-5386-3881-1

viewpoints. Despite being experienced with the environment and with the 3D models, a teacher did not point to correct depths and had difficulty clarifying pointing to students. The teacher did not understand the pointing problem until trading places with a student, and initally addressed the problem by placing real markers throughout the real environment to act as pointing targets, arranged through trial-and-error placement guided by an immersed student. This is not a suitable long-term approach when there can be many targets, virtual environment changes or flexibility, or changing teacher position. Visual feedback cues integrated into the software were thus investigated to assist and reinforce 3D pointing.

The student-like view also provided the teacher with limited awareness of student positions, and proper eye gaze through direct looking at avatars is unavailable in the mirror view. Formative tests also suggested that this limited personalized communication. This motivated the development of visual indicators the teacher could gaze towards to make eye contact with a student.

To select the best pointing cues, we conducted a more formal study of different possible visual cues, as we did not find established guidelines or solutions for pointing in a virtual mirror interface such as ours. Study participants tuned individual technique parameters, rated techniques, and clicked through target sequences for performance measures. Though not formally tested, we also summarize the solution for our eye gaze indicators.

A subset of our techniques was previously summarized in [6], but some of the top-performing techniques were not included, and no evaluation was previously conducted. Our contributions include:

- The development of a practical TV interface for teacher-guided VR field trips, and resulting knowledge about encountered problems and solutions related to eye gaze and pointing in a virtual mirror view.
- Detailing user-tuned visual feedback techniques for 3D pointing in a virtual mirror.
- Analyzing effects of the pointing cues and unaided stereo and monocopic baselines.

2 Related Works

Collaborative Virtual Environments have long been suggested as a way to enhance education, planning, and meetings [5]. Early collaborative environments, such the Virtual Playground by Roussou et al. [12], allowed multiple users to explore an environment aided by an autonomous guide. Fewer works used heterogeneous displays to allow an expert user to guide other users through the environment. One example that did is the system by Pick et al. [9] which used asynchronous interfaces for factory planning walkthroughs. Due to the importance of pointing for communication in CVEs, works such as this and similar interface setups (e.g. Froehlich et al. [3]) define accurate pointing as an important problem.

Many 3D and 2D pointing techniques have been studied in the past, especially in the context of object selection [2]. Because of the prevalence of the view type, most of the techniques have focused on the first-person perspective. Most often, these techniques use a ray [8] or cone [7] that is extended from the hand or face of the user as a selection tool. While these techniques are appropriate for their perspective, they do not address the problem of indirectness from another point of view.

Another set of tools that more closely resemble our application's problem are those that use an offset or clutching. These techniques lose proper proprioceptive feedback, and thus need to give additional visual feedback. Examples include the Go-Go technique [10] which automatically used a non-linear mapping of the real hand onto the virtual world, and the Virtual Pads [1], which allowed a user to manually decouple the physical and virtual space. However, this technique is mostly used for virtual hand metaphors, and not virtual pointing.

Figure 2: Gaze indicators are drawn on an overlay (left) and overlaid onto the main scene (right) at a position where teacher can look at them to maintain eye gaze with students.

Another tool used for pointing is interactive worlds in miniature [13]. Froehlich et al. [3] used this tool for communication via pointing with remote users, where two groups of users would stand in front of a projector that showed the other group in a shared virtual environment. While this application is similar to ours, the view still assumes a first-person perspective. Additionally, using a world-in-miniature tool for pointing would not give the level of detail required by unfamiliar students and could break immersion.

Certain virtual pointing techniques are sometimes classified as "exocentric" [11], but this still refers to a first person perspective in terms of the pointing experience. The third-person point of view provided by our system makes it more similar to a weatherman-style pointing problem, well-known from video-based environments [14]. However, we observed it was necessary to support 3D pointing, unlike 2D video-based worlds, to communicate well with HMD-immersed students. We did not find established methods for addressing 3D pointing with mirror views in the prior work.

3 Eye Gaze Indicators

The gaze indicators are icons drawn on an overlay on the screen (Figure 2). The teacher can look at the indicator to have her mesh make eye contact with a corresponding student. Placement of an indicator requires three pieces of geometric information: the teacher's head position (reported by the Kinect), a student's head position (reported by an HMD tracker), and a virtual representation of the teacher's TV (static in the same coordinate system as the Kinect).

In the 3D virtual environment, we define a ray between the teacher and student heads and find its intersection point with the TV representation. A TV-aligned indicator is then rendered at that point in an overlay view. If the ray does not intersect within the screen bounds, the indicator is positioned at the nearest monitor point to roughly reveal student position.

Note the overlay is rendered with a different projection geometry than the main scene view. This projection for the overlay corresponds to the monitor acting like a window into the virtual world, a conventional behavior for VR displays, rather than a mirror. This makes use of the teacher's reported head position to keep the indicators consistent with movement.

Future work will texture the rectangular icon with a webcam-based image of the student.

4 Visual Pointing Cues

We consider two pointing scenarios. In some cases, the system or developer may be able to predict likely targets based on predefined critical objects or pointing motion analysis. In this case, a visual cue can be designed to focus on a particular target, and a sense of accuracy can be given by the angle between the user's point and the target. In other cases, visual cues are needed to provide a sense of pointing depth more generally, without being tied to particular objects. We investigate both cue types. Examples of all techniques can be seen in Figure 3.

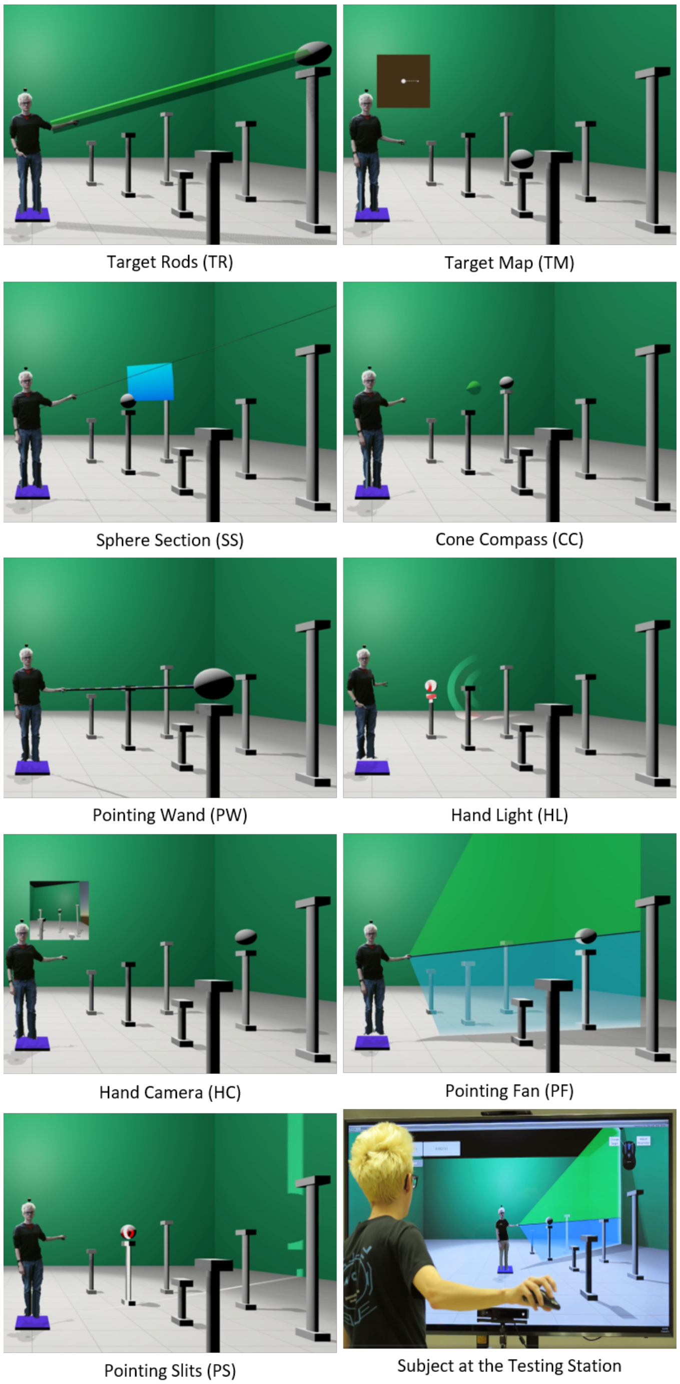

Figure 3: Experiment environment and visual feedback techniques considered in our studies. Techniques are grouped into those with a target defined (top two rows), and without a target defined (other rows). An example subject is shown at the testing station with the 75" TV and Kinect (bottom right).

We use the joint position information given by the Kinect to define pointing for the user. The Kinect reports the positions of all major joints of the user's body in world space. From this, we define a ray from one joint to another to be the user's pointing direction. In our experiments, we define the ray as the vector from the user's elbow to their palm.

4.1 Target Defined

Target Rods (TR): When the user is pointing, a translucent cylinder (rod) is extended along the vector from the target to the virtual elbow position. Depth information is reinforced through the rod's color and the scaling of the rod's ends based on the depth (azimuth angle) between pointing and target. A green rod signifies accurate pointing, and the ends are expanded or shrunk when pointing is in front of or behind an object.

Target Map (TM): A top-down camera high above the user's elbow is used to render known targets to a minimal top-town representation, textured onto a visible quad. A line segment is drawn from its center to represent the current pointing direction, so the line intersects with the target representation when pointing is accurate. The appearance resembles a circular gauge or dial indicator.

Sphere Section (SS): A mostly-invisible sphere is centered on the elbow with radius matching target distance, and a section of its surface is made visible to show the error in the user's current pointing. Specifically, two points on the sphere are defined: the point at which the target lies and the point where the user's pointing direction intersects with the sphere. The section is drawn between these two coordinates. The effect is an angular analog of the rubber band selection tool in most 2D interfaces.

Cone Compass (CC): A small cone glyph is placed either near the hand or the target. The cone acts either as a compass, pointing directly at the target or hand (depending on placement), or as a simple glyph pointing up or down according to pointing depth error. The color also changes to reflect the pointing accuracy, with green signifying accurate pointing.

4.2 No Target Defined

Pointing Wand (PW): A ray-like technique using a thin cylinder with a ringed texture placed on the hand. The wand acts as an extension of the arm, and users can know when they are pointing at an object when they intersect.

Pointing Fan (PF): A translucent vertical fan is attached to an arm joint. Different shapes are considered for the fan. Objects behind the fan are at least partially visually occluded, giving the user a sense of how deep into the scene they are pointing.

Hand Camera (HC): A virtual camera is placed on the user's pointing hand, with the frustum aimed in the pointing direction. The camera's image is textured onto a visible quad. A light reticle is also drawn in the center of the quad to emphasize its center, where targets appear when pointing is accurate.

Hand Light (HL, bullseye reticle): Multiple hand-mounted lights are arranged to project a pattern in the pointing direction. The pattern is a red and white bullseye reticle, which appears on objects pointed to.

Pointing Slits (PS): Thin slits of light are projected out from the user's hand in the pointing direction. A thin vertical slit can be seen along the ground and shows the pointing depth. A horizontal slit or a bullseye reticle is added for pointing elevation.

5 User Study

We conducted a two-phase user study to determine good configurations for each pointing technique and then compare the techniques.

5.1 Subjects

10 subjects (9 Male, 1 Female) participated. Subjects were screened to confirm normal or corrected vision and normal limb function.

Nine were from a local computer science department. Four subjects had prior VR experiences, and 8 had extensive video game experience. All were right handed.

5.2 Apparatus

Subjects stood in front of a 75” Samsung 3D TV and a Microsoft Kinect. As illustrated in Figure 3, the virtual environment contained 7 target spheres on each side of the user. Targets were placed on pedestals of varying heights and at varying depths from where the user was asked to stand.

5.3 Methods

5.3.1 Parameter Adjustment Phase

The first experiment phase had users freely adjust certain per-technique parameters while pointing at targets in the environment. It consisted of 9 rounds, one per technique, in random order and with random initial parameters. The subject was given a wireless mouse to hold in the pointing hand to use as their interaction device.

One target was shown at a time, and the subject could freely change to another random target at any time by left-clicking the mouse. To focus on pointing with the hand closest to the target, only targets on the subject’s dominant side were used.

The subject was able to modify one parameter at a time by scrolling the mouse wheel. Once the subject reached the preferred value for that parameter, they notified the proctor and were moved to the next parameter. Once all parameters were tuned, the subject could move to the next round or readjust any other parameters at will. The configurations chosen during this phase were used for the second phase of the experiment.

Allowing subjects to use different parameters in the next phase ensured that subjects were using a preferred setting and would not misuse a technique or rank it lower for differing from their preference. It also resembles real-world interfaces in which users adjust techniques as they find best, and thus is a legitimate random variable. No parameters were designed to change the fundamental effect of a technique. Low variances in accuracy (azimuth shown later) suggest that the differing parameters did not cause problems with respect to our interpretations of results.

5.3.2 Performance and Ranking Phase

During the second phase, the subject was presented with each technique again, in a random order. Two baseline techniques were added: stereo 3D and a monoscopic (no-cue) baseline. Stereo 3D was not included with other techniques, to focus specifically on the sense of depth from our visual cues. And, in our application, we prefer to avoid occluding a student’s view of a teacher’s eye gaze with 3D glasses.

The subject went through a set of steps similar to a Fitts’ Law experiment in which they pointed at and selected (left click) targets as quickly as they can while still maintaining accuracy. Accuracy was not enforced, as it would not be enforced in our application, so the time to click was up to a subject’s judgment. A single round consisted of the subject selecting 14 targets (each target appeared twice in random order) 3 times, with a short rest between the three repetitions. Click timing and error data were recorded.

At the end of each condition, the subject scored the technique’s helpfulness from 1 to 10. Subjects were also asked to explain anything they liked or disliked about that technique.

6 RESULTS

We present the user preferences from the adjustment phase and compare results of our techniques and baselines from the ranking phase. We especially want to identify which techniques consistently perform the best and are most preferred, in order to do further investigation. Thus, we use accuracy, time, and preference results to eliminate candidates for further study. For time and accuracy measures, we performed Student’s paired t-tests for interesting cases, using an alpha of 0.05. For subjects’ subjective rankings, we compare interesting cases with a Wilcoxon signed-rank test. We also compare the two groups of pointing techniques (target defined and not defined) and infer important components and traits of effective techniques.

6.1 Adjustment Phase Preferences

Target Rods (TR): Subjects adjusted the length and positioning of the rods, as well as whether the end facing the user expands or shrinks. The rods can be either extended the full length from the elbow to the target, extended from the elbow to the vector’s midpoint, or extended from the vector’s midpoint to the target. While there was no general consensus on preferred scaling behavior, the majority (6) of subjects preferred the rod to be extended from the elbow to the target, as it enabled them to see when their arm was fully eclipsed while still seeing the correct target.

Target Map (TM): Subjects adjusted the position of the quad in the environment, as well as the size of the quad. The quad can be positioned to hover near the user’s head, their shoulder, or their moving pointing hand. Participants preferred a more static position, but there was no preference between the shoulder and head. Preferred sizes spanned from .5 meter to 1 meter in width and height, with the mean preferred size being 0.77 meters.

Sphere Section (SS): Subjects adjusted the shape of the visible section. The shape can be circle-like, where the two relevant coordinates are opposite points on the circle, or rectangle-like, where the two coordinates are opposite corners. A majority (6) of subjects preferred the rectangle, due to it showing both axes of error.

Cone Compass (CC): Subjects adjusted the position of the cone and whether it acted as a compass or a simple glyph. Subjects preferred the compass mode while positioned by the hand.

Pointing Wand (PW): Subjects adjusted rod transparency and switched between the rod having a static length or changing length to exactly touch the closest object pointed to (ray’s first hit). While there was no strong consensus on transparency, every participant preferred varying length, as it more obviously showed intersections.

Pointing Fan (PF): Subjects modified the shape of the fan, its length (how far out from the hand it extends), and its material, with one option being a standard transparent material and the other turning white, per pixel, in proportion to the depth of nearby pixels behind it. While there was no consensus on shape, a majority (9) preferred the material that indicated depth, as it showed when they were close to an object. For length, subjects always chose lengths that were as far as the furthest target in the scene or higher.

Hand Camera (HC): Subjects adjusted the size and position of the camera quad, as well as the field of view (zoom) of the camera. A majority (6) of subjects preferred shoulder placement. Preferred sizes ranged from 0.61 meters to 1 meter, with the mean being .88 meters. Preferred fields of view ranged from 34 to 61 degrees, with the mean being 48 degrees.

Hand Light (HL): Subjects adjusted the angle (size) of the bullseye and light intensity. The majority of subjects preferred highest possible light intensity. The preferred light angles ranged from 11 to 41 degrees, with a mean of 19 degrees, indicating preference for a smaller, more precise bullseye.

Pointing Slits (PS): Subjects adjusted the width of the slits and reticle, the intensity of the lights, and the configuration of lights, choosing between an added horizontal slit or an added bullseye reticle. While there was no strong consensus on the exact configuration, a majority of subjects preferred the added reticle. A majority of participants also preferred the light intensity to be as high as possible. The width, determined by the light’s field of view, ranged from 57 to 151 degrees, with a mean of 115.

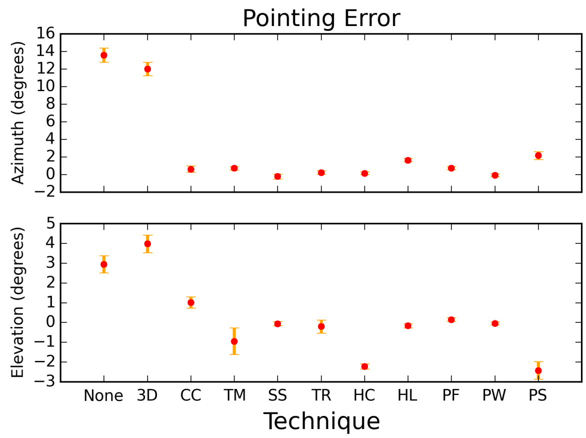

Figure 4: Mean azimuth and elevation errors at time of target selection, with standard error bars. Azimuth error represents an error in the depth of pointing, with positive error indicating pointing in front of the target. Elevation error represents height error, with positive error above the target.

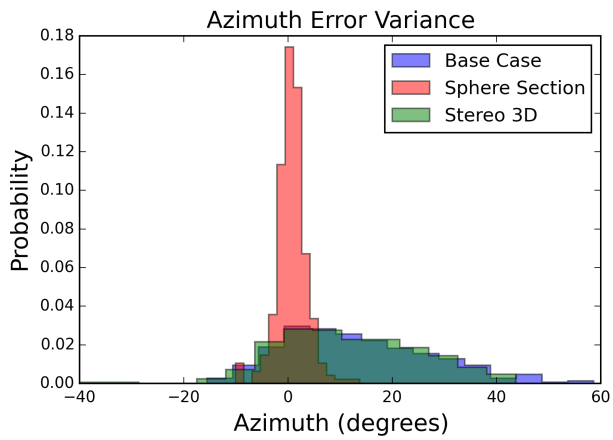

Figure 5: Histogram of azimuth error for the entire population for the two baselines and the sphere section technique.

6.2 Accuracy

The main purpose of the visual cues was to improve the users' pointing depth and to discourage the problematic weatherman-type pointing behavior. Angle error results are seen in figure 4. For azimuth error, every technique was shown to be statistically significantly different ($p < 0.05$) from both baselines. However, no significant difference was found between the two baselines ($p = 0.077$). For elevation error, every technique was significantly different ($p < 0.05$) from both baselines. No significant difference was found between the two baselines ($p = 0.052$). When comparing error between target-defined and no-target-defined groups, we find statistically significant differences in both azimuth and elevation, indicating that all techniques produce better pointing accuracy, though those designed for predefined targets are overall more accurate.

More importantly, the baselines have large variances, reflecting low precision and consistency of pointing. A high average error may reflect differences between our estimation of direction (elbow to hand) and natural human pointing posture, but higher variances indicate problems regardless of mean error estimate. Figure 5 shows a wide variance among baselines when compared to a more precise technique (sphere section). Note that subjects already experienced pointing techniques prior to this stage, and naive subjects could be expected to have even more trouble with 3D pointing. Overall, results are consistent with users only performing coarse pointing for the baselines.

Notably, while the target map is accurate and precise for depth, it is inaccurate and imprecise in elevation, caused by the lack of elevation feedback. This, timing data, and subject feedback (detailed next) are consistent with users quickly finding the correct depth and not being concerned with elevation when there is no related feedback. Similarly, pointing slits suffer a similar lack of precision in elevation due to minimal or confusing elevation feedback.

The hand camera is also shown to be inaccurate in elevation compared to a more accurate technique (e.g. pointing fan, $p < 0.05$). Based on verbal subject feedback, we infer the cause to be a subject aiming to just get the target within the camera's frustum rather than specifically in the center of it, despite the target reticle. Because of this, we consider these three techniques inappropriate for use.

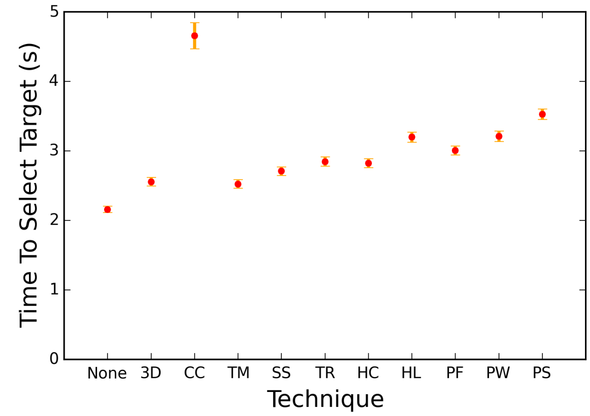

Figure 6: Mean time-to-click (seconds) between targets per technique, with standard error bars.

6.3 Time

Time-to-click can be seen in figure 6. Every technique was statistically significantly different ($p < 0.05$) from the no-cue baseline. Compared to stereo 3D, however, there was no statistically significant finding for the sphere section ($p = 0.078$) and target map ($p = 0.614$) ($p < 0.05$ for all others). The two baselines were found to differ significantly from each other ($p < 0.05$). When comparing times between defined-target and no-defined target groups, we find no statistically significant difference ($p = 0.633$). However, we do find a significant difference ($p < 0.05$) if the Cone Compass technique is excluded as an outlier. In that case, lower times tend to be found in target-defined techniques.

Participants reported that they felt they were able to click most quickly when completely unencumbered (no technique), but also reported little to no confidence in their accuracy. Responses were similar for stereo 3D. Again, it seems an interface without pointing cues encourages coarse pointing behaviors, even among subjects who previously experienced 3D cues reinforcing better behavior.

The cone compass technique is shown to take significantly longer than all other techniques. Subjects reported difficulty knowing which direction to move to correct their pointing due to its small size (along the hand-to-target vector) and conical shape. Because of this, we consider this technique inappropriate for use.

6.4 Subjective Technique Ratings

Subjective ratings are seen in figure 7. All techniques are shown to have statistically significantly higher ratings than the no-cue baseline, except for the cone compass and hand light ($p = 0.766$ and p

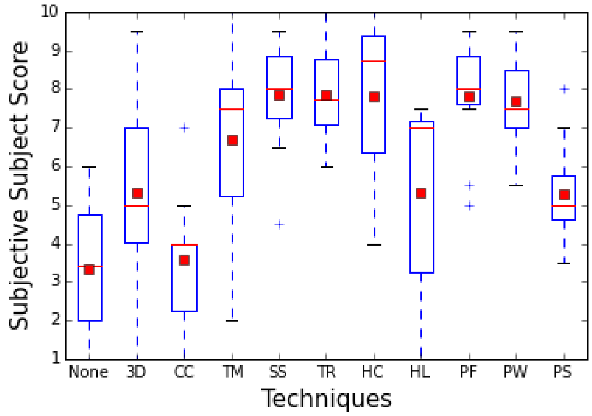

Figure 7: Rating scores given to each technique by subjects. Squares denote means, pluses denote outliers, and the horizontal line denotes medians.

= 0.059). Five of the techniques are statistically significantly better than the stereo 3D baseline, but the target map, cone compass, pointing slits, and hand light techniques were not found statistically different ($p = 0.443$, $p = 0.125$, $p = 1.0$, and $p = 0.766$, respectively).

For practical and frequently used applications, we would want to use techniques that not only have high ratings, but have few low ratings and little variance. For these reasons, we consider the hand light technique to be inappropriate for this use.

7 Discussion

Considering the good accuracy and positive subject ratings, we chose the pointing fan as the default pointing technique for our application. While the sphere section and target rod techniques did provide better accuracy and time, they require known targets, and are therefore suitable only in certain applications. Using them in scenes with several simultaneous candidate targets could create visual clutter and confusion. Thus we consider the slight additional time for the pointing fan to be a worthwhile tradeoff.

However, we still consider it useful to identify other techniques that performed well and discuss them. Thus, based on the techniques providing minimal selection time and error while showing strong user preference, we can identify the sphere section, target rods, pointing fan, and pointing wand as ones that could be improved upon and evaluated further with a larger study. Identifying common traits among these, we can see that the pointing wand, rods, and fan all involve an object that is stretched from part of the virtual body to the target. The sphere section and target rods both give a clear indication of error while keeping the visual part of the technique near the target.

Techniques that performed a 2D projection of an image onto the environment (hand light and pointing slits) were disliked because projection warped user perception of where they were pointing unless they were already correct, leading to a high time-to-click. Techniques that encouraged looking at something other than the target (hand camera and target map) were also disliked and inaccurate. We can thus infer from this that the most important features of this type of visual feedback is that the indication of correctness be viewable near the target and, to a lesser extent, the feedback mechanism include some 3-dimensional component along the pointing vector.

8 Conclusion and Future Works

We presented the design of an interface made for an unencumbered teacher guiding networked immersed students through a virtual environment. Testing and development led to a "virtual mirror" view which necessitated the design of two major communication components. We addressed the problem of eye gaze with visual indicators that the user can look at to maintain gaze with a remote user. We addressed the problem of inaccurate pointing with a two-phase initial study that showed reasonable parameter configurations for various visual feedback techniques, and then evaluated those techniques. Findings showed that we can achieve the desired pointing accuracy with just a small increase in targeting time, which we consider worthwhile due to the benefits of improved communication. These techniques were important in the development of a practical interface that avoids having to immerse or otherwise restrict the user.

Acknowledgment

This work was supported by the National Science Foundation under Grant Number 1451833 and by the Louisiana Board of Regents Support Fund under contract LEQSF (2015-16)-ENH-TR-30. We thank Kenneth A. Ritter and Terrence L. Chambers for prior work.

References

[1] C. Andujar and F. Argelaguet. Virtual pads: Decoupling motor space and visual space for flexible manipulation of 2d windows within ves. In *Proceedings of the 2007 IEEE 3DUI*, March 2007.

[2] Ferran Argelaguet and Carlos Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121 – 136, 2013.

[3] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE TVCG*, 19(4):616–625, April 2013.

[4] Terrence L. Chambers, J. R. Raush, and B. Russo. Installation and operation of parabolic trough organic rankine cycle solar thermal power plant in south louisiana. *Energy Procedia*, 49:1107 – 1116, 2014.

[5] E. F. Churchill and D. Snowdon. Collaborative virtual environments: An introductory review of issues and systems. *Virtual Reality*, 3(1):3–15, 1998.

[6] Sam Ekong, Christoph W. Borst, Jason Woodworth, and Terrence L. Chambers. Teacher-student vr telepresence with networked depth camera mesh and heterogeneous displays. In *Advances in Visual Computing: 12th International Symposium*, pages 246–258, 2016.

[7] Jiandong Liang and Mark Green. Jdcad: A highly interactive 3d modeling system. *Computers & Graphics*, 18(4):499 – 506, 1994.

[8] Mark R Mine. Virtual environment interaction techniques. *UNC Chapel Hill computer science technical report TR95-018*, 1995.

[9] S. Pick, S. Gebhardt, B. Weyers, B. Hentschel, and T. Kuhlen. A 3d collaborative virtual environment to integrate immersive virtual reality into factory planning processes. In *Proceedings of the 2014 3DCVE*, pages 1–6, March 2014.

[10] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. The go-go interaction technique: Non-linear mapping for direct manipulation in vr. In *Proceedings of the 1996 ACM UIST*, UIST, pages 79–80, New York, NY, USA, 1996. ACM.

[11] Ivan Poupyrev and Tadao Ichikawa. Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages & Computing*, 10(1):19 – 35, 1999.

[12] Maria Roussou, Martin Oliver, and Mel Slater. The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning. *Virtual Reality*, 10(3):227–240, 2006.

[13] Richard Stoakley, Matthew J. Conway, and Randy Pausch. Virtual reality on a wim: Interactive worlds in miniature. In *Proceedings of the 1995 SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pages 265–272, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[14] Vincent John Vincent. The mandala virtual reality system: The vivid group. In *Proceedings of the 3rd Annual Virtual Reality Conference and Exhibition on VR Becomes a Business*, pages 167–170, 1993.