Manually Locating Physical and Virtual Reality Objects

Karen B. Chen, **Ryan A. Kimmel**, **Aaron Bartholomew**, **Kevin Ponto**, **Michael L. Gleicher**, and **Robert G. Radwin**, University of Wisconsin–Madison

Objective: In this study, we compared how users locate physical and equivalent three-dimensional images of virtual objects in a cave automatic virtual environment (CAVE) using the hand to examine how human performance (accuracy, time, and approach) is affected by object size, location, and distance.

Background: Virtual reality (VR) offers the promise to flexibly simulate arbitrary environments for studying human performance. Previously, VR researchers primarily considered differences between virtual and physical distance estimation rather than reaching for close-up objects.

Method: Fourteen participants completed manual targeting tasks that involved reaching for corners on equivalent physical and virtual boxes of three different sizes. Predicted errors were calculated from a geometric model based on user interpupillary distance, eye location, distance from the eyes to the projector screen, and object.

Results: Users were 1.64 times less accurate ($p <$ .001) and spent 1.49 times more time ($p =$ .01) targeting virtual versus physical box corners using the hands. Predicted virtual targeting errors were on average 1.53 times ($p <$ .05) greater than the observed errors for farther virtual targets but not significantly different for close-up virtual targets.

Conclusion: Target size, location, and distance, in addition to binocular disparity, affected virtual object targeting inaccuracy. Observed virtual box inaccuracy was less than predicted for farther locations, suggesting possible influence of cues other than binocular vision.

Application: Human physical interaction with objects in VR for simulation, training, and prototyping involving reaching and manually handling virtual objects in a CAVE are more accurate than predicted when locating farther objects.

Keywords: virtual reality, physical interface, simulation, CAVE

Address correspondence to Robert G. Radwin, University of Wisconsin-Madison, 1550 Engineering Drive, Madison, WI 53706-1608, USA; e-mail: radwin@discovery.wisc.edu.

INTRODUCTION

The motivation for this research is the prospect of simulating natural interactions with objects in the physical environment (PE), such as reaching for, acquiring, and handling objects with the hands, by using visually mediated virtual objects in a three-dimensional (3-D) virtual reality (VR) environment. In this study, we examine possible differences between how users physically reach for and locate virtual objects. Observing and comparing user performance in the PE and VR may help better understand how to best use VR simulation technology for prototyping new products and devices and for studying human interactions and behaviors in living environments.

The present study was conducted in a commercially built CAVE (cave automatic virtual environment), which is a projection-based VR first introduced for scientific visualization (Cruz-Neira, Sandin, & DeFanti, 1993; Cruz-Neira, Sandin, DeFanti, Kenyon, & Hart, 1992). In general, projection-based VR is advantageous for studying natural physical interactions with virtual objects because it allows users to see their own hands relative to the virtual objects, and user performance was found to be less awkward than in other VR environments (Havig, McIntire, & Geiselman, 2011; Sander, Roberts, Smith, Otto, & Wolff, 2006; Sutcliffe, Gault, Fernando, & Tan, 2006).

We are ultimately interested in creating more natural VR interactions in which users' hands physically touch, grasp, and handle virtual objects. We wish to first comprehend user accuracy of manually targeting virtual objects in various locations in the 3-D space in accordance with the graphics software and apply those findings to enhance natural interactions with virtual objects so that collisions with visually mediated objects and user movements in the CAVE are accurately detected and displayed. Since 3-D vision in a CAVE creates the perception that the

virtual objects observed are located in three-dimensional space, the coordinates where users perceive and physically locate those images may not perfectly coincide with the coordinates of the object created by the CAVE software. Even small differences between the perceived and actual locations may prevent users from efficiently completing manual tasks. In this study, we investigate the magnitude of such differences and rationalize factors that might affect them.

Authors of human perception research in VR space have previously looked at distance approximation. Generally, researchers have reported distance underestimation in VR (Thompson et al., 2004; Witmer & Kline, 1998) relative to PE (Alexandrova et al., 2010; Willemsen & Gooch, 2002). However, distance estimation through walking in large-screen VR and in PE was similar in adults and children (Plumert, Kearney, Cremer, & Recker, 2005). It was also suggested that distance estimation in different VR simulation technologies had varying degrees of accuracy, and simulations displayed on computer monitors had the lowest accuracy (Lampton, McDonald, Singer, & Bliss, 1995). Although there is a general consensus on distance underestimation in VR, the magnitude of error during target reach and location in a CAVE is still unclear.

Others have tried to understand differences in user performance between VR and PE through target aiming (Liu, van Liere, Nieuwenhuizen, & Martens, 2009) and object grasping (Magdalon, Michaelsen, Quevedo, & Levin, 2011). Task performance and movement time were longer in VR for both head-mounted display and single desktop 20-inch stereo monitor (Liu et al., 2009; Magdalon et al., 2011). Additionally, users who wore pinch gloves to intercept virtual and physical objects in a CAVE spent significantly longer time to complete virtual object movement tasks (Sutcliffe et al., 2006). Generally, there are performance differences between PE and VR for those studied tasks.

Human performance is also affected by various depth perception cues and individual factors as well as by objects and screen distance from the eyes (Bajcsy & Lieberman, 1976; Mather, 1996; O'Shea, Blackburn, & Ono, 1994; Walk & Gibson, 1961). Binocular disparity is one of the cues, and its association with interpupillary distance (IPD; Wann, Rushton, & Mon-Williams, 1995), an individual factor, has been mathematically demonstrated (Ogle, 1953). Accurately perceiving stereoscopic objects has been suggested to be related to binocular disparity, which is associated with IPD (Patterson, 1997). However, binocular disparity alone may be insufficient to provide accurate depth perception (Hibbard & Bradshaw, 2003), and motion parallax was another cue that was identified to aid in the understanding of user depth perception (Bradshaw, Parton, & Glennerster, 2000). Users would strategically utilize different cues under different tasks and constraints, yet seldom were binocular disparity and motion parallax cues incorporated while performing their given tasks (Bradshaw et al., 2000).

Recently, a geometric model developed by Ponto, Gleicher, Radwin, and Shin (2013) demonstrated a relationship between user VR binocular perception and a CAVE binocular disparity setting. In their geometric model, participant IPD, eye location, and distances from eye to the projector screen (DS) and from eye to the perceived location of the object projection (DP) are considered for calculating binocular disparity targeting error in a CAVE (Figure 1). The geometry based on the distances and locations of the points is outlined in Equation 1, and these equations are combined and rearranged to solve for DP in Equation 2. The difference between distance to virtual point (DV) and DP is the calculated binocular disparity error, which suggested the importance to investigate the influence of binocular disparity in relation to user performance.

$$\frac{\text{CAVE BD}}{\text{DV}} = \frac{\text{Disp}}{\text{DS} - \text{DV}}$$

$$\frac{\text{IPD}}{\text{DP}} = \frac{\text{Disp}}{\text{DS} - \text{DP}} \tag{1}$$

$$\frac{\text{IPD}}{\text{DP}} = \frac{\text{CAVE BD} \times (\text{DS} - \text{DV})}{\text{DV} \times (\text{DS} - \text{DP})}$$

$$\text{DP} = \frac{\text{IPD} \times \text{DV} \times \text{DS}}{\text{CAVE BD} \times \text{DS} - \text{CAVE BD} \times \text{DV} + \text{IPD} \times \text{DV}} \tag{2}$$
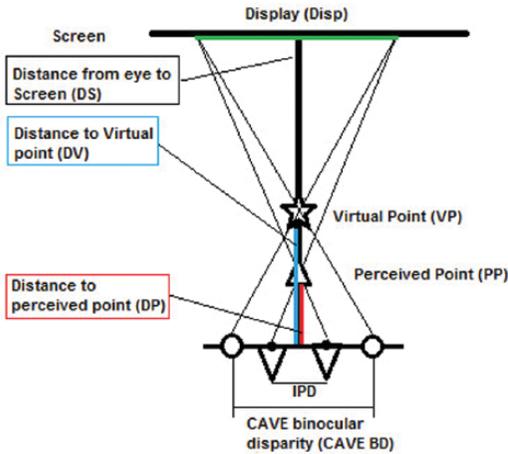
Figure 1. Geometric model of a user's line of sight to a virtual target (indicated as a single point, depicted by the star symbol). CAVE binocular disparity (CAVE BD) is the distance between the CAVE virtual cameras. The location of the projected image by the cameras is labeled as the virtual point (VP), depicted by the star (*), and its distance to the viewer is labeled as DV. Interpupillary distance (IPD) is the distance between the two eyes for each subject. Distance from the eyes to the perceived point (PP) is indicated as DP. Display (Disp) is the distance between the two points of the line of sight from both the cameras and the user that landed on the screen.

In the current study, we compared human reach and localization of visually mediated virtual objects in a CAVE against equivalent physical objects located at arm-length distances. It was hypothesized that human performance in the virtual condition was not equivalent to the physical condition. Participants were asked to reach for the four corners on the upper face of three physical boxes and three equivalent virtual boxes projected in a CAVE. The localization of a box corner represented the perceived location of the box corner. We compare the perceived location of virtual against the physical box corners and then assess the extent of the contribution of external and internal factors to user performance in VR.

METHOD

Participants

Sixteen students were recruited with informed consent from the University of Wisconsin–

TABLE 1: Demographics and Characteristics of the Analyzed Participants

| Gender	3 females, 11 males
Age (*SD*)	22.7 (2.6)
Height in cm (*SD*)	176.1 (7.5)
Interpupillary distance in cm (*SD*)	6.18 (0.26)
Handedness	2 left-handed, 12 right-handed
Vision correction	10 with correction, 4 without correction

Madison campus, and Table 1 lists their self-reported demographics. Inclusion criteria were self-reported normal or corrected-to-normal vision and the ability to stand for at least 20 minutes. Exclusion criteria included reported history of epileptic seizures or blackouts, tendency for motion sickness when experiencing visual motion conflicts, neuromotor impairments, Lasik eye surgery, perception-altering medication, claustrophobia in 3 m × 3 m × 3 m square room, or sensitivity to flashing lights. Two participants were excluded due to Lasik eye surgery or forgotten eyeglasses.

Procedure

Stature and arm length were measured using an anthropometric caliper, and IPD was measured using a digital pupilometer (Digital PD ruler PM-100, Luxvision, www.luxvsion.net). Participants were instructed to stand in the CAVE in their stocking feet (wearing only socks or booties) and performed targeting tasks.

Instrumentation

The VR was created in a 2.93 m × 2.93 m × 2.93 m rear-projected six-faced CAVE consisting of four walls, one ceiling, and one solid acrylic floor. Two 3-D projectors (Titan model 1080p 3D, Digital Projection, Inc., Kennesaw, GA, USA) with maximum brightness of 4,500 lumens per projector, total 1,920 × 1,920 pixels combined, and 70 Hz of update rate per eye projected images onto each surface of the CAVE. Immersive 3-D scenarios were implemented
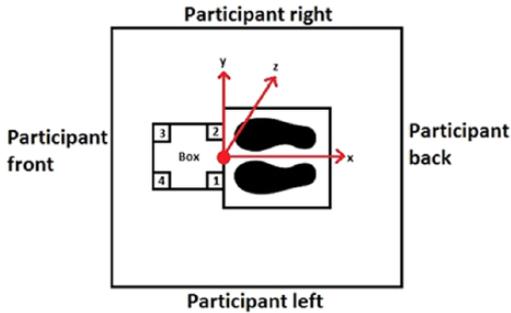
Figure 2. Relative locations of the box, corners, and the participant. Corners of the boxes were labeled 1 through 4; Corners 1 and 2 were the closer left and right corners, and 3 and 4 were the farther right and left corners, respectively. The red dot indicates the midpoint between Corners 1 and 2, which also represents the origin of the coordinate system when measured from the floor. The positive *x*-axis was toward the posterior, the positive *y*-axis was to the right, and the positive *z*-axis was toward the superior direction.

using the VirtualLab software package (Virtual CAVELib API, Mechdyne, Marshalltown, IA, USA), and four workstations (2 × Quad-Core Intel Xeon) generated displayed graphics. Audio was generated by a 5.1 surround sound audio system.

The data acquisition system consisted of an ultrasonic tracker set (VETracker Processor model IS-900, InterSense, Inc., Billerica, MA, USA), including a handheld wand (MicroTrax model 100-91000-EWWD, InterSense, Inc.) and head trackers (MicroTrax model 100-91300-AWHT, InterSense, Inc.) that sampled at 60 Hz. Twelve ultrasonic emitters evenly placed along the upper (two per edge) and vertical (three per edge) edges of the CAVE allowed full 6 degrees of freedom wand and head tracking. Shutter glasses (CrystalEyes 4 model 100103-04, RealD, Beverly Hills, CA, USA) with head trackers mounted on the top rim created stereoscopic images from the user's viewpoint.

Experimental Task

The task was to locate labeled box corners (Figure 2 to 4) in random order, and each corner location action represented one trial. Participants were instructed to stand in the "ready position" (hands pointing down and palms inward while standing erect) over the footprint image projected on the floor in the center of the CAVE throughout the experiment (Figures 2 and 3).

A physical box was situated, or a virtual box was projected, at the same location while participants turned their head to their right side and looked at a bull's-eye target projection on the wall (Figure 3). Participants were asked to face and observe the box for 3 s and then return to ready position before locating any corner. They were then instructed to locate a corner using the protruding midpoint of the wand and to squeeze a trigger when the wand was at the location they perceived was the corner. All wand-triggered click events were recorded. No auditory, visual, or tactile feedback was provided to indicate location accuracy to prevent possible ordering effects in a limited repeated-measures experiment.

The same box was used for two consecutive trials before it was swapped for the next box. Participants were instructed to locate the corner for both box types, but they were further instructed to reach as close to the corner as possible without actually touching the physical box.

Each participant completed 48 trials, two replications for each of the four corners for three physical and three virtual boxes. The physical boxes sizes were chosen based on convenience and availability (Figure 4), and the images of the virtual boxes were then photographed and constructed to match the dimensions of the physical boxes. No practice was provided.

Variables and Data Analysis

Independent variables were box size (three levels), corner (four levels), and box type (two levels: physical and virtual), and they all varied within subjects. Dependent variables were accuracy (overall error and error along the *x*-axis, *y*-axis, and *z*-axis), approach toward corners (wand rotation angle), and efficiency (task time). Overall error was the calculated Euclidean distance (Equation 3) between the coordinates of the tip of the wand (x_w, y_w, z_w) and the corner (x_C, y_C, z_C). Errors along each orthogonal axis were the absolute difference between the components of that axis (Equation 4). The relationships of errors and the independent variables
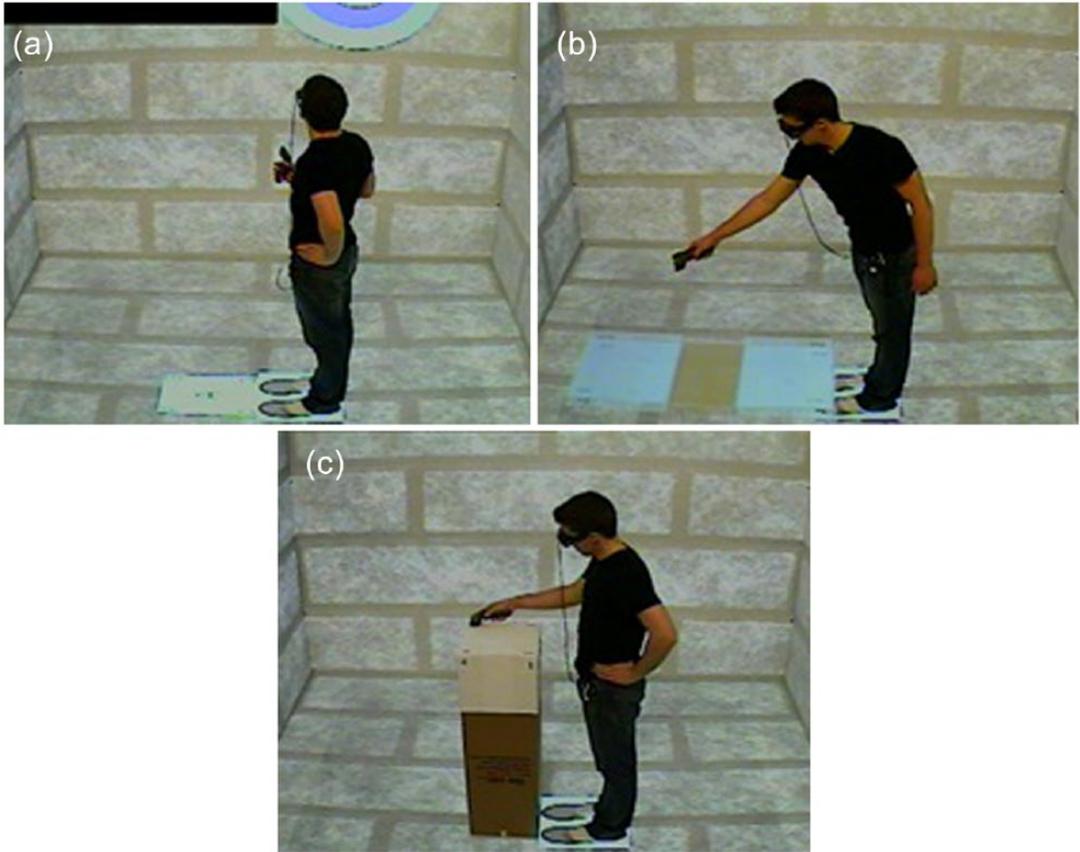
Figure 3. A participant inside the cave automatic virtual environment (CAVE), standing on footprint images projected on the CAVE floor while performing indicated tasks: (a) turning to the right and looking at the projected bull's-eye, (b) pointing to a corner of a virtual box while standing on the projected footprints, and (c) pointing to a corner of a physical box while standing on the projected footprints.

were analyzed with repeated-measures ANOVA with $\alpha = .05$. The designated corner labels and the coordinate system are illustrated in Figure 2. The origin was located along the adjacent edge of the box and the projected footprints, and (0, 0, 0) was defined as midpoint between the participant's toes (Figure 3).

$$\text{Overall error} = \sqrt{(x_W - x_C)^2 + (y_W - y_C)^2 + (z_W - z_C)^2} \quad (3)$$

$$\begin{aligned} x\text{-error} &= |x_W - x_C| \\ y\text{-error} &= |y_W - y_C| \\ z\text{-error} &= |z_W - z_C| \end{aligned} \quad (4)$$

Three distinctive wand rotation angles were measured with respect to the three orthogonal axes. Wand rotation about the x-axis was roll, the y-axis was elevation, and the z-axis was azimuth (Figure 5). Rotation about each axis was indicated by the angle magnitude and a positive or negative sign, in which the positive rotation direction was determined by pointing the right-hand thumb to the positive direction of axis of interest and curling the fingers toward the palm (right-hand rule).

Binocular disparity error was predicted using the geometric model demonstrated by Ponto et al. (2013). Figure 6 shows an example of the relationship between the calculated binocular disparity errors using the geometric model, based on IPD values and the mean population standing eye height (156.9 cm) from the U.S. anthropometric survey (Gordon et al., 1989). No
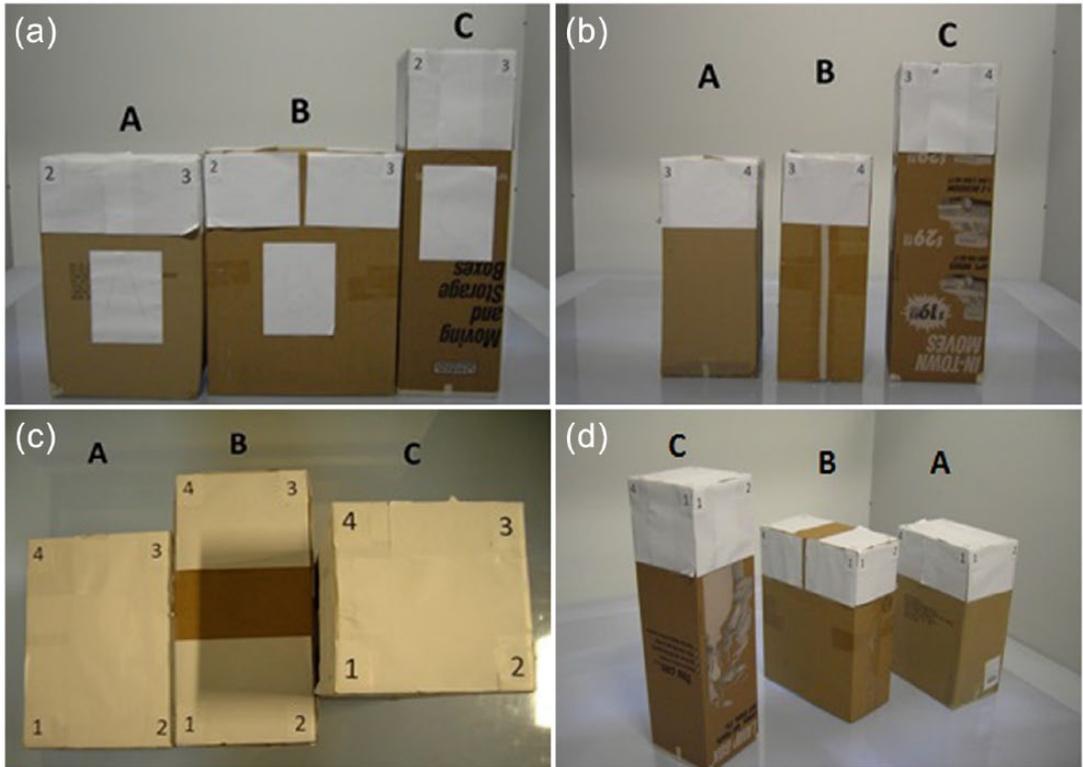
Figure 4. Views of three physical boxes from different planes and angles showing their relative sizes. Images a, b, and c display the boxes in the order of Boxes A, B, and C (from left to right). Box A's dimensions were (length × width × height, in centimeters) 47 × 32 × 72, Box B's were 59 × 28 × 74, and Box C's were 31 × 34 × 102. Image d displays the boxes in the order of C, B, A (from left to right). The illustrations show the (a) side view looking at the *x-z* plane, (b) side view looking at the *y-z* plane, (c) top view looking at the *y-x* plane, and (d) side view looking at boxes from an angle.
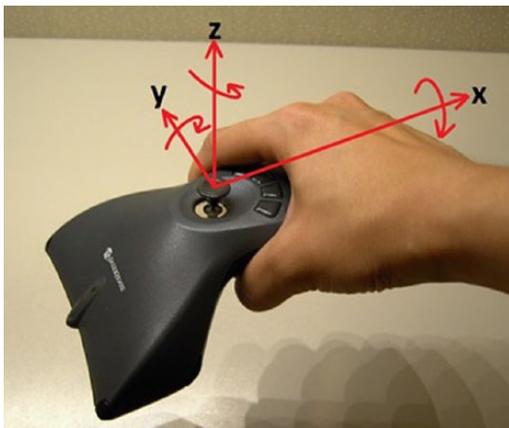

Figure 5. Three rotation directions of the wand about the three orthogonal axes.

systematic relationship was observed between binocular disparity error and distance from the eyes to the target corner.

RESULTS

Mean overall error for the VR cases was plotted against trial number, and a log-log regression curve was fitted to examine potential practice effect over the 24 trials, $F(1, 22) = 11.96$, $p = .002$. Mean error of the first VR trial was the average error of all participants' first virtual box trial, and then subsequent data points were calculated similarly to obtain the mean errors of remaining VR trials. The difference in overall error between VR Trials 1 and 4 was 1.36 cm, whereas the difference between VR Trials 4 and 24 was 0.3 cm.
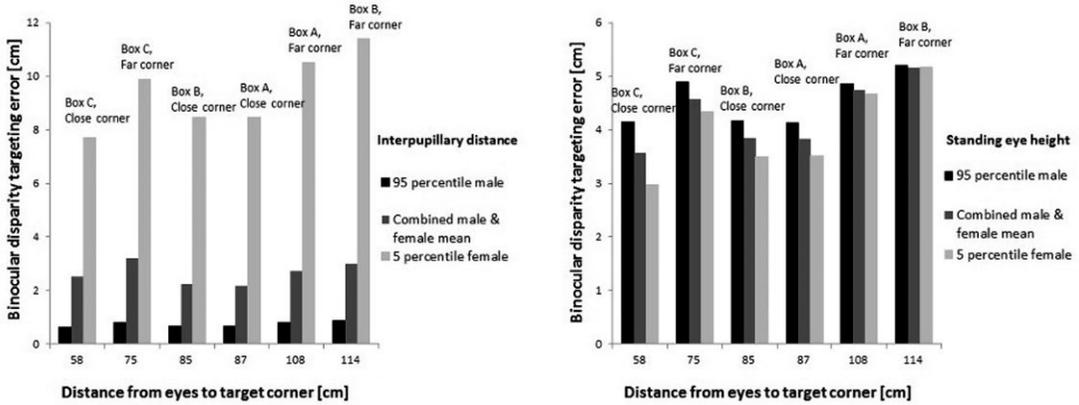
Figure 6. The left graph is the binocular disparity targeting error calculated from the geometric model against increasing distance to virtual point (DV) while holding user standing eye height at the mean of the U.S. Army Personnel Anthropometric Survey (156.86 cm) with varying interpupillary distance (IPD) from the U.S. Army Personnel Anthropometric Survey (Gordon et al., 1989). The right graph is the binocular disparity targeting error calculated from the geometric model against increasing DV while holding user IPD at the mean of the U.S. Army Personnel Survey (6.35 cm) with varying standing eye height from the U.S. Army Personnel Anthropometric Survey (Gordon et al., 1989). Binocular disparity error in all cases increased with increasing IPD or increasing standing eye height. Closer corners were Corners 1 and 2, and far corners were Corners 3 and 4 (Figures 2 and 4).

The first replicate of each corner for all three virtual and three physical boxes was excluded from data analysis to remove practice effects and also to maintain a full factorial experiment. To test if the practice effect was removed, we regressed the errors over order for the remaining VR trials. No statistically significant effect of time on error was observed, $F(1, 10) = 2.51$, $p = .144$.

Though the experimenter did not observe any physical box touches, potential touch instances were still analytically verified. Wand trajectories were sampled at 60 Hz, and the velocity and acceleration profiles were calculated using numerical differentiation for each participant to assess potential physical box contacts. High-frequency noise of the profiles was filtered out using a Gaussian smoothing algorithm. It was anticipated that the wand would have zero velocity and rapid deceleration if a touch occurred. Assuming that it would take at least 90 ms to react to a touch, at a 60 Hz sampling rate, we would expect at least five data points that were at or near zero velocity while rapid deceleration occurred. Trajectory data for 10 s before and

after every click response of all physical box trials were examined, and it was determined that no profiles matched the criteria listed. It was concluded that the participants did not touch the physical boxes as instructed.

Overall Error

Overall error was significantly affected by box type, $F(1, 12) = 31.2$, $p < .001$, and corner, $F(3, 11) = 3.70$, $p = .046$. Post hoc analysis revealed that average virtual boxes error was 1.64 times greater than physical box error ($p < .05$). In general, farther corners (Corner 4) from the participant resulted in greater error (1.1 times greater) than closer corners (Corner 2), statistically controlling for box type and box (A, B, and C). Two-way interactions were observed for Box Type × Box, $F(2, 12) = 69.2$, $p < .001$; Box Type × Corner, $F(3, 11) = 5.92$, $p = .012$; and Box × Corner, $F(6, 8) = 9.49$, $p = .003$. Errors of the three boxes (collapsed across corners) for the different box types are graphically compared in Figure 7.
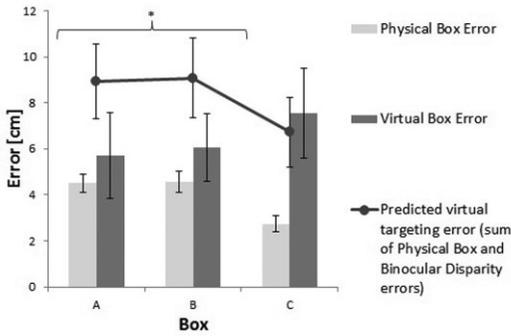
Figure 7. The bars represent mean overall physical box and virtual box errors collapsed across corners for different boxes (±1 *SD*). The predicted virtual box error was significantly greater than the physical box error for all boxes (*p* < .05). The circles represent the summed physical box and predicted errors. The bracket with a star (*) sign indicates the statistical significant difference between the summed predicted errors and virtual box errors (*p* < .05) for Boxes A and B, yet the summed predicted error was not statistically significantly different from the virtual error of Box C.

Errors in the Forward, Lateral, and Vertical Directions

Error in the forward direction was represented by the *x*-component of the coordinate, and it was significantly affected by box type, $F(1, 13) = 95.0$, $p < .001$, and box length along the forward direction, $F(3, 11) = 19.2$, $p < .001$. On average, virtual box error was 5.54 times greater than the physical box error in the forward direction (Figure 8). Post hoc analysis indicated that error in the forward direction was significantly greater for the farther targets (*p* < .05). However, there were no statistical significant differences between the average forward errors of the targets that were 47 cm and 59 cm distance from 0.0 cm. There was a two-way interaction for Box Type × Box Length, $F(3, 11) = 150.7$, $p < .001$. Average virtual target error at 59 cm in the forward direction was 1.9 times greater than the virtual target at 0.0 cm, yet the average physical target errors were not different (Figure 8).

Error in the lateral direction, represented by the *y*-component of the coordinate, was not significantly affected by box type or box width (length in the *y*-axis). Error in the vertical direction,

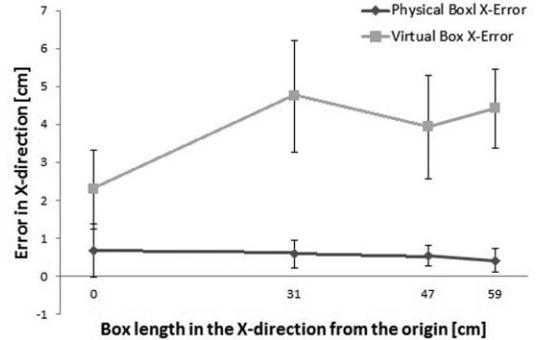*Figure 8.* Error in the *x*-axis with respect to change in box length in *x*-axis (box length in the *x*-axis presented in the order of origin: Box C, Box B, and Box A). Error bars are ±1 *SD*.
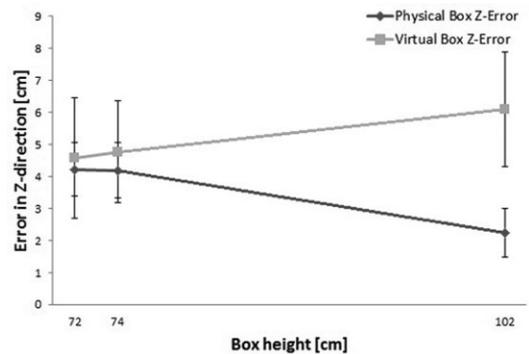

Figure 9. Error in the *z*-direction with respect to change in box height (in increasing height order: Box A, Box B, and Box C). Error bars are ±1 *SD*.

represented by the *z*-component of the coordinate, was significantly affected by box type, $F(1, 13) = 14.6$, $p = .002$, and corners, $F(3, 11) = 5.31$, $p = .017$. Average virtual box error in the vertical direction was 1.45 times greater than that for the physical boxes (*p* = .002). Regardless of box type and box height, error at farther corner (Corner 3) from the participant was 1.1 times greater than at the closer corner (Corner 1) to the participants. Significant two-way interactions between box type and box height, $F(2, 12) = 66.9$, $p < .001$, and box height and corners, $F(6, 8) = 9.82$, $p = .003$, were observed. Error for the tallest virtual box was 1.33 times greater than for the shortest virtual box, yet error for the tallest physical box was 1.89 times less than for the shortest physical box (Figure 9).
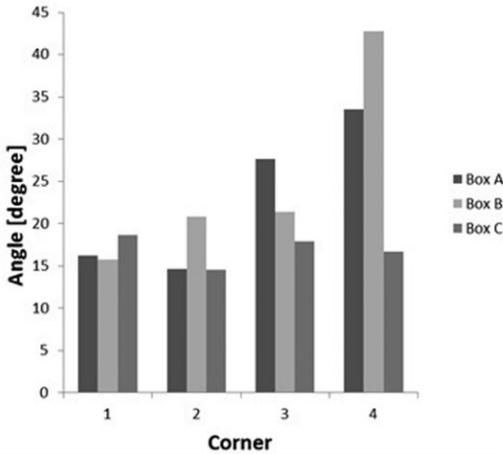
Figure 10. The differences of angles between the unit vector when approaching a physical target and the unit vector when approaching a virtual target.

Wand Rotation Angles

Three distinctive wand angles of each trial were converted into individual unit vectors in polar coordinate (Equation 5). Each unit vector represented the direction where the wand was approaching the corner from at the time of trigger for that specific trial.

$$X = \cos(\text{elevation angle}) \times \cos(\text{azimuth angle})$$
$$Y = \cos(\text{elevation angle}) \times \sin(\text{azimuth angle})$$
$$Z = \sin(\text{elevation angle}) \qquad (5)$$

Dot product of the physical and virtual unit vectors of the same box and corner was calculated to determine the angle between the physical and virtual unit vectors, which would suggest variation in approaching directions. The approaching directions varied the greatest for the farthest corners. The greatest difference in corner approach directions between the physical and VR trials was Box B, which had the longest length in the x-direction (forward), followed by Box A and then C (Figure 10).

Task Time

Task time was significantly affected by box type, $F(1, 13) = 8.7$, $p = .011$, and box size, $F(2, 12) = 6.32$, $p = .013$. On average, task time for the virtual boxes (4.57 s) was 1.49 times longer than for the physical boxes (3.06 s) ($p = .011$).

Average task time decreased as box height increased. Users on average spent 1.43 times more time to complete the trials involving the shortest box (4.62 s) than they did the tallest box (3.22 s).

Predicted Binocular Disparity Error and Observed Error

The observed physical box error was generally less than observed virtual box error ($p < .05$) (Figure 7). Further analysis was conducted to examine the role of binocular disparity on user performance in terms of observed error and to understand the contribution of binocular perception in VR. We first confirmed that IPD and stature distribution among our participants were not biased from the U.S. anthropometric survey (Gordon et al., 1989) and then calculated the binocular disparity error using the geometric model and our collected data (Ponto et al., 2013). A calculated variable, the predicted virtual targeting error, was the sum of the binocular disparity errors and the physical box errors. The predicted virtual target error represented the expected virtual box error for this task. Post hoc pairwise comparisons were conducted between the predicted virtual targeting error and the virtual box error for all four corners of all three boxes, with the Holm-Bonferroni adjusted alpha.

Predicted virtual targeting errors were always significantly greater than virtual box error for all corners of Boxes A and B ($p < .001$), and the predicted virtual targeting error was 1.13 cm greater than the virtual box error for Corner 1 (close-up corner) of Box C, $F(1, 13) = 6.77$, $p = .02$, and they were not statistically significantly different for the other three corners ($p > .05$). Collapsed across the corners, predicted virtual targeting errors were 3.22 cm and 3.00 cm greater than virtual box errors for Box A, $F(1, 13) = 79.1$, $p < .001$, and Box B, $F(1, 13) = 87.5$, $p < .001$, respectively. The predicted virtual targeting errors were on average 1.53 times greater than the virtual box errors for Boxes A and B. Overall virtual box errors were 0.83 cm greater than the predicted virtual targeting errors for Box C, which was not significantly different, $F(1, 13) = 2.87$, $p = .11$.

DISCUSSION

In the present study, we investigated targeting of virtual and physical objects in a CAVE. The

results indicated that user performance based on accuracy and time involving virtual objects was significantly poorer than with physical objects, and this finding was consistent with previous literature. More importantly, user performance was related to the location of the target as other depth cues compensated at different distances.

Prior to any statistical data analysis, we examined the mean overall error of the VR trials and observed a 1.36-cm decrease in error after the fourth VR trial and then no obvious improvement in mean overall user error in the subsequent trials. This finding suggested a practice effect, and therefore the first trial of the two replications of corner targeting was removed. We regressed the remaining VR trials against time, and there was not a statistically significant difference of error over time. The participants did not receive any visual feedback from the VR system or verbal feedback from the experimenter regarding their performance, as we would anticipate greater improvement in performance if the participants adjusted their final wand location by considering the feedback provided. Relatively smaller error observed in the physical box trials confirmed that even if a practice effect existed for the physical boxes, the effect was much smaller than for the virtual boxes, and therefore their effects were not considered.

A limitation of the experiment was the participants were not instructed to look away at the bull's-eye between two consecutive touches that involved the same box. However, the presentation of boxes was randomized, and there were no observable systematic trends among consecutive trials.

User Accuracy (Error)

Users generally had significantly greater accuracy when reaching for corners of a physical box versus a virtual box. Averaged across all three boxes of the same type (physical or virtual), participants triggered the wand at 6.6 cm away from the virtual corners, and the error was 4 cm for the physical corners. It was hypothesized that human performance in VR and PE are different, and this hypothesis is supported by our data. Similar trends were previously reported for traversed and verbal estimated distances (Alexandrova et al.,

2010; Witmer & Kline, 1998). Interestingly, there was on average 4 cm of physical box error, which may be explained by participant instructions not to actually make contact with the physical box surface. It is expected that the difference between the virtual and physical box errors would increase if participants were permitted to touch the physical corners.

Errors were further analyzed with distinctive components along the x-, y-, and z-axes. Results indicated that errors in the forward and vertical directions (x-axis and z-axis, respectively) were influenced by corner location (i.e., target distance) but not in the lateral direction (y-axis). Virtual box error in the forward direction (along the x-axis) was significantly greater for the farther corners than for the closer corners, yet the physical box errors were not statistically significantly different among the corners of the same direction. Moreover, users had greater error when they aimed at far virtual targets in the horizontal direction (4.4 cm error for the far corners compared to 2.3 cm error for the closer corners). This result suggested that the accuracy of aiming at a virtual target is related to its location, and it is consistent with the literature (Interrante, Ries, & Anderson, 2006), but this relationship was not seen in physical targets. It is plausible that virtual box errors were related to the reaching posture since they increased as the target was farther in the forward direction. Images taken by the CAVE video camera were also reviewed, and we observed that users reached for virtual box corners with relatively different bending postures compared to reaches for physical box corners. There was more bending and wrist turning for far virtual box corners. On the basis of this finding, we would suggest that the source of the physical box error was not due to target location or reaching posture instability since the physical box error did not change significantly in relation to the increase in target distance in the forward direction (Figure 8).

The magnitude of virtual box error increased with increasing box height, but physical box error decreased. In this study, taller boxes represented closer targets to the participants since all boxes were shorter than all participants' standing eye height. As a result, virtual box errors were the greatest when the target was the closest to the participants' eyes along the z-axis. On the

other hand, greater accuracy was observed for closer physical targets that varied in vertical distances (Figure 9). It is anticipated that physical targeting error is affected by whole-body posture such that when fewer degrees of freedom and fewer body joints are involved in pointing at a target, the error should decrease. Targeting the 102-cm-height box often did not involve torso movement, and therefore less error was anticipated. We hypothesize that close-up virtual box errors may be due to visual perception, and farther virtual box errors observed in the forward direction may be due to physical reach and postural constraints. In order for users to interact with virtual objects based on their perceived location of the object, targets may need to be adjusted for activation boundaries.

Another plausible explanation for the observed effects was the use of a physical wand to aim at virtual targets, such that users may have viewed the physical wand while aiming for a virtual box corner, which may have resulted in an accommodation mismatch (Drascic & Milgram, 1996). In this condition, accommodation and the one vergence point for the physical wand and physical box are the same. However, the accommodation of the virtual target was on the CAVE screen, whereas the vergence point for the wand was on the virtual box corner. We suggest that the virtual box errors were most likely perceptual, which relates back to the earlier discussion. The possibility that an accommodation mismatch was responsible for the errors observed should be considered in future studies.

Though it is possible that the shutter glasses may have interfered with the users' ability to perceive the location of the physical box corners, the shutter glasses allowed the users to see through them, and therefore we do not anticipate interference. Assuming the physical box errors were not attributed to wearing the shutter glasses, then the physical box errors would be the smallest possible error in this type of task and setting. If this effect exists, it is representative of the conditions that would arise in a CAVE simulation involving people interacting with virtual objects, and therefore the findings of this study reflect that experience.

The present study outcome was consistent with other studies that resulted in greater error for virtual targets and less error for the physical targets, although we primarily investigated user interaction with virtual and physical objects within arm-length distances with boxes as opposed to longer distance estimation. Errors in the x-axis and z-axis were also significantly affected by box type. Both x-axis and z-axis errors in VR were greater than for the PE.

Task Time

Task time was also significantly affected by virtual and physical box types. Our participants spent 1.49 times longer to complete virtual box trials than physical box trials, and Liu et al. (2009) also reported less efficient aimed movements in VR. The result suggested that users moved more cautiously when aiming at virtual corners, and this finding is important to consider when analyzing movement time in VR. Since users were slower in aiming virtual targets, simulated tasks may not be completely transferrable to the physical equivalent.

Approach Angles

Participants used the wand to approach different corners from various angles. Wand approach angles were significantly different among the four corners regardless of box type, implying that users approached the same corner number similarly in both physical and virtual box types. Although overall wand approach angles were not significantly different between the virtual and physical boxes, an increase in the difference between the virtual and physical approach angles was observed for the farthest corner. This finding could be explained by potential obstruction of the projection to form virtual objects by the participant's hand when reaching for the farthest corner. It would be relatively more difficult to aim at a corner when parts of the image were blocked. This possibility reinforces the importance of visualizing complete objects to perform tasks that require accuracy in VR. The participants may have approached the corners from an angle that would minimize blocking the images, which could explain smaller errors observed for closer virtual corners in the forward direction. On the other hand, users may have taken a more natural route when approaching the physical

corners by moving across the box since they did not have to be concerned about blocking visual information because none of the physical box components relied on projector image creation.

Predicted Targeting Error and Observed Error

We found no significant difference between the predicted virtual targeting error and the virtual error for Box C but found that participants tended to perform better than anticipated for Boxes A and B (Figure 7). We propose that the difference in performance was due to the difference in target locations (distances) to the eyes. Box C (102 cm in height) was the tallest box, and therefore the target locations were closer to the eyes of the participants than were the targets on Boxes A and B (72 cm and 74 cm in height, respectively). Moreover, the virtual box error of the closer corner of Box C was statistically significantly greater than its predicted virtual targeting error, supporting the suggestion that participant targeting accuracy decreased for closer targets. Specifically, the mismatch of visual accommodation and vergence increases for virtual objects that are closer to the eyes in the CAVE, which may be related to the decreased targeting accuracy. These results mirror those found by Pollock, Burton, Kelly, Gilbert, and Winer (2012) when studying difference in perception of nontracked users in VR and by Woods, Docherty, and Koch (1993) when studying image distortions in stereoscopic video systems.

A recent study conducted by Renner, Velichkovsky, Helmert, and Stelzer (2013) also pointed out that accounting just for IPD in the CAVE setting was not sufficient to completely reduce the error of the users, in which IPD was related to binocular disparity. We further suspect that users utilized depth cues other than binocular disparity when viewing farther virtual targets, as authors of some studies have suggested that depth information in VR does not solely rely on binocular disparity cues, and the type of visual cues utilized is dependent on the type of task (Bradshaw et al., 2000; Hibbard & Bradshaw, 2003). It is likely possible those visual cues not based on binocular disparity assisted in overcoming the incorrect perceived position of the target location, specifically for the corners of

Boxes A and B (relative to Box C). Kelly et al. (2013) recently reported that leftward and rightward displacement from the center of projection caused less distortion than predicted by predictions of models based on monocular and binocular viewing geometry.

Results of this study revealed that there are varying levels of user aiming accuracy and approach angle due to target location, which provides some insights for future designs of visually mediated objects in VR. It is important to account for user aiming inaccuracy by modifying the activation boundary of virtual objects in order to provide users better experience when studying natural gestures and manipulation of virtual objects. Moreover, user accuracy was lower for targets in the far horizontal location as well as for close-up corners. Consequently, users performed better when they aimed at corners that were not too far or too close. This result suggests that there may be a range that users would best operate in for a target aiming task.

Any virtually projected object on a screen should appear to be at that location, as there is no virtual disparity (Ponto et al., 2013). As the virtual object moves away from the screen, the artifacts of VR become more pronounced (Woods et al., 1993). For instance, as the object comes farther out from the screen, depth compression becomes a greater factor. Additionally, motion parallax will increase as the object is closer to the user, resulting in increased problems from any incorrect head tracking/positioning (Cruz-Neira et al., 1993). Adjustment of the CAVE binocular disparity setting relative to user IPD to account for differences in user binocular disparity individually might help reduce some error, but the results of this study indicated that binocular disparity may be just one source that contributed to user accuracy. Furthermore, individual IPD adjustments may not be feasible when multiple observers are participating in a CAVE simulation.

CONCLUSIONS

Human performance in VR was less accurate (greater error) and less efficient than in the PE, as error was greater for both close and far virtual objects. Users approached the virtual and physical targets (within 1 m) from similar angles. It

is important to consider how users approach and acquire virtual objects within a distance range, which will help researchers better understand the activation boundaries of virtual objects and then manipulate VR to allow more natural interactions.

Our data indicated that physical target error was not due to the target location or the postural instability of the user reaching for the target because the magnitude of error was not affected by distance to the target. We anticipate that farther virtual target error was associated with the awkward reaching posture, and the closer virtual target error was associated with binocular disparity because closer targets involved less awkward reaching posture. The evaluation of user perception of virtual and physical objects provides some insights to user performance, which may contribute to future studies involving natural interactions of hands and virtual objects in VR.

Finally, we anticipate that other depth cues in addition to binocular disparity may be involved with targeting farther virtual objects, because users performed better than our prediction using a geometric model. The geometric model accounted for the binocular disparity but did not take into account all factors related to performance, as it is a gross model. For instance, other means of determining depth besides binocular disparity could be used to determine corner position. We also suspect that mismatch of visual accommodation and vergence played a role in the increased targeting inaccuracy in aiming for closer virtual targets. In future studies, we will aim to better understand these factors.

ACKNOWLEDGEMENT

KEY POINTS

- Virtual box errors were generally greater than physical box errors.
- Participants approached the physical and virtual box corners from similar angles, but the variation increased as the distance from the user to the corner increased.
- Inaccuracy of the nearer virtual targets is associated with binocular disparity, and the inaccuracy of the farther virtual targets is associated with user reaching posture.
- Task time was longer with the virtual trials.

REFERENCES

Alexandrova, I. V., Teneva, P. T., de la Rosa, S., Kloos, U., Bulthoff, H. H., & Mohler, B. J. (2010). Egocentric distance judgments in a large screen display immersive virtual environment. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization* (pp. 57–60). New York, NY: ACM.

Bajcsy, R., & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing*, *5*, 52–67.

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, *40*, 3725–3734.

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 135–142). New York, NY: ACM.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, *35*(6), 64–72.

Drascic, D., & Milgram, P. (1996, January). *Perceptual issues in augmented reality*. Paper presented at the Electronic Imaging: Science & Technology, San Jose, CA.

Gordon, C. C., Churchill, T., Clauser, C. E., McConville, J. T., Tbbetts, I., & Walker, R. A. (1989). *1988 anthropometric survey of U.S. army personnel: Summary statistics interim report*. Natick, MA: U.S. Army Natick Research, Development and Engineering Center.

Havig, P., McIntire, J., & Geiselman, E. (2011). Virtual reality in a cave: Limitations and the need for HMDs? In *SPIE Proceedings: Vol. 8041. Head- and Helmet-Mounted Displays XVI: Design and Applications* (n.p.). Bellingham, WA: SPIE.

Hibbard, P. B., & Bradshaw, M. F. (2003). Reaching for virtual objects: Binocular disparity and the control of prehension. *Experimental Brain Research*, *148*, 196–201.

Interrante, V., Ries, B., & Anderson, L. (2006). Distance perception in immersive virtual environments, revisited. In *Proceedings of IEEE Virtual Reality* (pp. 3–10). New York, NY: IEEE.

Kelly, J. W., Burton, M., Pollock, B., Rubio, E., Curtis, M., De La Cruz, J., Gilbert, S., & Winer, E. (2013). Space perception in virtual environments: Displacement from the center of projection causes less distortion than predicted by cue-based models. *ACM Transactions on Applied Perception*, *10*(4), Article 18.

Lampton, D. R., McDonald, D. P., Singer, M., & Bliss, J. P. (1995). Distance estimation in virtual environments. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1268–1272). Santa Monica, CA: Human Factors and Ergonomics Society.

Liu, L., van Liere, R., Nieuwenhuizen, C., & Martens, J.-B. (2009). Comparing aimed movements in the real world and in virtual reality. In *Proceedings of IEEE Virtual Reality Conference* (pp. 219–222). New York, NY: IEEE.

Magdalon, E. C., Michaelsen, S. M., Quevedo, A. A., & Levin, M. F. (2011). Comparison of grasping movements made by

healthy subjects in a 3-dimensional immersive virtual versus physical environment. *Acta Psychologica*, *138*, 126–134.

Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *263*, 169–172.

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America*, *43*, 906–913.

O'Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research*, *34*, 1595–1604.

Patterson, R. (1997). Visual processing of depth information in stereoscopic displays. *Displays*, *17*(2), 69–74.

Plumert, J. M., Kearney, J. K., Cremer, J. F., & Recker, K. (2005). Distance perception in real and virtual environments. *ACM Transactions on Applied Perception*, *2*, 216–233.

Pollock, B., Burton, M., Kelly, J. W., Gilbert, S., & Winer, E. (2012). The right view from the wrong location: Depth perception in stereoscopic multi-user virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, *18*, 581–588.

Ponto, K., Gleicher, M., Radwin, R. G., & Shin, H. J. (2013). Perceptual calibration for immersive display environments. *IEEE Transactions on Visualization and Computer Graphics*, *19*, 691–700.

Renner, R. S., Velichkovsky, B. M., Helmert, J. R., & Stelzer, R. H. (2013). Measuring interpupillary distance might not be enough. In *Proceedings of the ACM Symposium on Applied Perception* (p. 130). New York, NY: ACM.

Sander, I., Roberts, D. J., Smith, C., Otto, O., & Wolff, R. (2006). Investigating the impact of method of immersion on the naturalness of balance and reach activities. In *Int Conf Series Disabil Virtual Reality Assoc Technol* (ICDVRAT) (pp. 317–324). Reading, UK: ICDVRAT/University of Reading.

Sutcliffe, A., Gault, B., Fernando, T., & Tan, K. (2006). Investigating interaction in CAVE virtual environments. *ACM Transactions on Computer-Human Interaction*, *13*, 235–267.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, *13*, 560–571.

Walk, R. D., & Gibson, E. J. (1961). A comparative and analytical study of visual depth perception. *Psychological Monographs: General and Applied*, *75*(15), 1.

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, *35*(19), 2731–2736.

Willemsen, P., & Gooch, A. A. (2002). Perceived egocentric distances in real, image-based, and traditional virtual environments. In *Proceedings of Virtual Reality, 2002* (pp. 275–276). New York, NY: IEEE.

Witmer, B. G., & Kline, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence: Teleoperators and Virtual Environments*, *7*, 144–167.

Woods, A. J., Docherty, T., & Koch, R. (1993, January). *Image distortions in stereoscopic video systems*. Paper presented at the IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology, San Jose, CA.

Karen B. Chen is a PhD candidate at the University of Wisconsin–Madison in biomedical engineering. She has a BS degree (2009) and an MS degree (2010) from the University of Wisconsin–Madison.

Ryan A. Kimmel has BS (2010) and MS (2011) degrees in biomedical engineering from the University of Wisconsin–Madison. He currently works for Abbott Laboratories in Lake County, Illinois.

Aaron Bartholomew has BS degrees (2012) in computer science and mathematics, a BBA degree (2012) in business management and human resources, and an MS degree in mechanical engineering (2013) from the University of Wisconsin–Madison. He currently works for NVIDIA.

Kevin Ponto is an assistant professor at the University of Wisconsin–Madison in the Design Studies Department and at the Wisconsin Institutes for Discovery. He has a BS degree (2004) from the University of Wisconsin–Madison, an MS degree from the University of California–Irvine (2006), and a PhD from the University of California–San Diego (2010).

Michael L. Gleicher is a professor at the University of Wisconsin–Madison in the Department of Computer Sciences. He earned a BSE degree (1988) from Duke University and MS (1991) and PhD (1994) degrees from Carnegie Mellon University in Pittsburgh, Pennsylvania.

Robert G. Radwin is a professor at the University of Wisconsin–Madison in biomedical engineering, industrial and systems engineering, and orthopedics and rehabilitation. He has a BS degree (1975) from New York University Polytechnic Institute, and he earned an MS degree (1979) and a PhD (1986) from the University of Michigan in Ann Arbor.