The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments

Brice Pollock, Melissa Burton, Jonathan W. Kelly, Stephen Gilbert, and Eliot Winer

Abstract—Stereoscopic depth cues improve depth perception and increase immersion within virtual environments (VEs). However, improper display of these cues can distort perceived distances and directions. Consider a multi-user VE, where all users view identical stereoscopic images regardless of physical location. In this scenario, cues are typically customized for one “leader” equipped with a head-tracking device. This user stands at the center of projection (CoP) and all other users (“followers”) view the scene from other locations and receive improper depth cues. This paper examines perceived depth distortion when viewing stereoscopic VEs from follower perspectives and the impact of these distortions on collaborative spatial judgments. Pairs of participants made collaborative depth judgments of virtual shapes viewed from the CoP or after displacement forward or backward. Forward and backward displacement caused perceived depth compression and expansion, respectively, with greater compression than expansion. Furthermore, distortion was less than predicted by a ray-intersection model of stereo geometry. Collaboration times were significantly longer when participants stood at different locations compared to the same location, and increased with greater perceived depth discrepancy between the two viewing locations. These findings advance our understanding of spatial distortions in multi-user VEs, and suggest a strategy for reducing distortion.

Index Terms—Perception, stereoscopy, and collaborative interaction.

1 INTRODUCTION
Stereoscopic displays are increasingly prevalent in virtual environments ranging from 3D movies and televisions to surgical training systems and fully immersive virtual reality. Compared to monocular displays, the added depth information conveyed by stereoscopic displays has the potential to improve perception of egocentric (self-to-object) and exocentric (object-to-object) distances. Accurate egocentric distance perception is critical when reaching for a cup, avoiding an upcoming obstacle, or judging the trajectory of a moving vehicle. Accurate exocentric distance perception is critical for proper recognition of objects, since distortions in perceived depth or width of a shape can drastically alter the perceived shape. Exocentric distance perception is also important for remote guidance tasks in which one object must be manipulated relative to another such as guidance of a surgical tool viewed from a laparoscopic camera image. The combination of stereoscopic viewing with immersive virtual reality creates a high fidelity experience in which viewers may treat the synthetic environment as real. This is especially valuable for training applications in which the virtual environment must closely resemble the real environment for appropriate transfer of training.

Despite the potential benefits of stereo depth cues, improper display of those cues can greatly distort perceived distances and directions. These distortions can be experienced when viewing a 3D movie from a seat on the side of the theater or in the front or back row. Perceptual distortions occur in this movie scenario because the stereo cues are generated from a single location, referred to as the center of projection (CoP). Stereoscopic images displayed on a projection screen are appropriate when viewed from the CoP, but displacement from the CoP results in inappropriate stereo cues and distorted perception of the virtual space. Perceptual distortion increases with further displacement from the CoP [1]. This problem plagues virtual environments displayed on projection screens visible to multiple users (Figure 1). In such multi-user virtual environments, one user (the “leader”) receives the projectively correct stereo images, which are updated when he/she moves through the environment (as sensed by head tracking equipment). In this way, the leader is always at the CoP. All other users (the “followers”) receive stereoscopic images intended for the leader. When a follower stands close to the leader, the follower sees a relatively faithful representation of the virtual environment. As the distance between the leader and follower increases, perceived spatial properties of the environment become increasingly distorted for the follower because he/she receives increasingly inappropriate stereoscopic cues.

The goal of this project was to characterize the distortions that occur when viewing stereoscopic virtual environments after displacement from the CoP, and to evaluate the impact of those distortions on collaborative tasks within the virtual environment. If two viewers occupy approximately the same position in the environment (e.g., if they stand side-by-side), then their perceptual experiences of the environment should be similar and they should therefore be able to communicate effectively to one another about spatial properties of that environment. Whether their views of the environment are accurate (as when both users stand near the CoP) or distorted (as when both users stand far from the CoP but close to one another) should not impair their ability to communicate with one another, as long as they both perceive the environment similarly. If two viewers occupy disparate positions in the environment, then they should have greater difficulty communicating to one another about spatial properties of the environment because their perceptual experiences of the environmental layout differ. Such difficulties should occur in communication between a leader and follower who experience accurate and distorted views of the environment, respectively, and also between multiple followers who experience differently distorted views of the environment.

In this study, pairs of viewers made collaborative judgments of object depth (the distance from the front to the back of the object, an exocentric distance judgment). Participants viewed an object in a virtual environment and came to a mutually agreeable decision about
Fig. 1. Example of a multi-user virtual environment. When using the leader-follower approach, only one user (leader) sees the projectively-correct stereoscopic images of the environment. All other users (followers) view those same images while standing in different locations.

the actual object depth. Collaborative judgments were made while viewers stood side-by-side or in disparate locations. When standing side-by-side, viewers’ judgments of spatial properties of the environment should be very similar to one another because they occupy nearly the same location. When standing in disparate locations, their judgments of the environment should differ from one another. These predictions are based on the assumption that judgments of object depth are based on the individual viewers’ perceptions of object depth, and herein the terms depth perception and depth judgment are used interchangeably. Furthermore, viewers stood at the CoP or were displaced forward or backward from the CoP. Judged object depth and the time required to reach agreement about object depth were recorded.

To preview the findings, depth judgments were distorted as a result of displacement from the CoP, although distortion was less than predicted by a ray-intersection model based on stereo viewing geometry. Collaboration time was fast when viewers stood side-by-side, regardless of whether the viewers were both at the CoP or both displaced from the CoP. Collaboration time increased when viewers stood at different locations from one another, and larger separations between participants led to slower collaboration times. Furthermore, we describe an asymmetry in perceived spatial distortion, whereby forward displacement from the CoP introduced greater distortion than did backward displacement. This led to faster collaboration time when the front-most viewer received the correct stereoscopic view compared to when the back-most viewer received the correct view, and this finding highlights a novel strategy for improving collaboration in multi-user virtual environments by assigning the leader role to the viewer closest to the display screen.

2 BACKGROUND

This section describes the relevant background for evaluating and interpreting visual distortions that occur after displacement from the CoP. This background includes 1) a geometric stereo-vision model for predicting the perceptual experiences after displacement from the CoP, along with past research evaluating this model, 2) existing solutions for displaying virtual environments to multiple users, and 3) previous research on egocentric and exocentric depth perception in virtual environments.

2.1 Modeling and Measuring Perceptual Distortion in Stereoscopic Virtual Environments

By using a ray-intersection approach, geometric properties of the stereo viewing environment can be used to predict the amount of perceptual distortion that should occur when viewing a virtual environment after displacement from the CoP [1][2][3]. For every point in the virtual environment, two corresponding points appear on the projection screen surface. Each point is sent to the appropriate eye via active or passive stereo glasses, which flicker in synchrony with the stereo display or filter out unique light wavelengths or polarities, respectively. A ray is traced from the center of each eye through the corresponding point on the projection screen and out into space. The intersection of those rays is the predicted location of the perceived point in 3D space. Figure 2 shows the predicted distortion (calculated using the ray-intersection model) in egocentric and exocentric distance when viewing a virtual rectangular shape from positions displaced from the CoP. The intended rectangular shape is drawn with solid lines, and the distorted shape is drawn with dashed lines. When viewing from the CoP (Figure 2, center), the perceived shape is identical to the intended shape. However, displacement behind (Figure 2, left) or in front of (Figure 2, right) the CoP leads to distortion in the perceived egocentric distance to the shape and also the perceived depth of the shape. The model predicts that backward displacement relative to the CoP will cause the virtual shape to appear farther away (distortion in perceived egocentric distance) and elongated in depth (distortion in perceived exocentric distance). Furthermore, the model predicts that forward displacement will cause the virtual shape to appear closer and foreshortened in depth.
was confirmed during pilot testing). The model predicts that lateral displacement from the CoP will result in perceived lateral displacement of the shape and distortion in the perceived shape such that rectangles appear to be non-rectangular parallelograms. However, the focus in this project is on distortions that occur after forward and backward displacements from the CoP, and the effect of lateral displacement was not tested.

Whether perceptual distortions adhere to the predictions of the geometric ray-intersection model has received surprisingly limited attention. To our knowledge, only two studies have evaluated perceptual distortions when viewing stereoscopic environments after displacement from the CoP [4]. In one study [1], participants adjusted a virtual hinge formed by two planes connected at one edge. The hinge appeared in front of the viewer, with the open end oriented toward the viewer (similar to the spatial arrangement of surfaces when reading a book). The virtual environment was otherwise empty. Participants controlled the hinge angle remotely and adjusted the hinge until it appeared to form a 90° angle. The hinge was viewed in stereo, and participants adjusted the hinge angle while viewing the display from the CoP or from a position displaced to the left or right of the CoP. Responses made from the CoP were quite accurate, but viewer displacement away from the CoP resulted in errors which increased with increasing displacement. Moreover, errors made after displacement were nearly identical to the predictions of the ray-intersection model. These results suggest that perceived spatial distortions after displacement from the CoP can be fully accounted for by the stereo viewing geometry.

The finding that perceptual distortions can be fully accounted for by stereo viewing geometry [1] is a significant departure from similar studies using monocular displays. When viewing monocular displays from locations displaced from the CoP, spatial judgments remain quite accurate, especially when there is a visible frame around the display. Under such conditions, judgments of perceived shape are biased toward the predictions of the geometric model, but only by 10%-50% (depending on the amount of displacement) of the expected magnitude [1][5]. This relative accuracy when viewing monocular displays after displacement from the CoP may be due to awareness of both 2D and 3D properties of the picture. Awareness of the 2D picture plane may help to reduce or correct for perceptual distortions [6][7], but may also detract from the perception of three-dimensionality within the virtual scene [8][9].

Recent work in our lab [4] indicates that virtual environments containing both stereoscopic and monoscopic depth cues result in distortions of judged depth that are in between those found when using environments with only stereoscopic [1] or only monoscopic [5] cues. In that study, judgments of object depth and angle were biased in the direction predicted by the ray-intersection model, but the magnitude of the bias was significantly less than predicted. For example, participants made verbal judgments of the depth of a rectangle placed on a textured ground surface in a virtual environment. Depth judgments were made while participants stood at the CoP (center viewing position), 2.5° in front of the CoP (front viewing position), or 2.5° behind the CoP (back viewing position). Depth judgments are shown in Figure 3. Judgments made from the center position were only 65% of the intended object depth, consistent with past work showing depth underestimation in virtual environments (see section 2.3 for more details on depth underestimation in virtual environments). To account for this foreshortening even at the CoP, model predictions were scaled by 65%. Consistent with model predictions, depth judgments made from the front position were foreshortened relative to judgments made from the center position, but the magnitude of the distortion was considerably less than predicted by the model. Furthermore, the distortion experienced after forward displacement was larger than that experienced after backward displacement, inconsistent with the model’s predictions. These results indicate that the distortions introduced by displacement from the CoP are smaller than predicted by the ray-intersection model, and that forward and backward displacement produces asymmetric distortion.

2.2 Techniques to Reduce Distortion

Multiple alternatives to the leader-follower approach have been proposed to remedy the perceptual distortions experienced by followers. Image blending [10] involves rendering a novel viewpoint for each of multiple head-tracked users. Using a compositing-based system, a blend zone is created in which each head-tracked user’s view overlaps. View clustering [10] is another technique that can be used to supplement image blending. View clustering is appropriate when multiple users look at the same portion of the screen. Gaze-intersection points are calculated and grouped together, resulting in equally distorted views for all users.

Image blending and view clustering are both promising techniques for displaying virtual environments to multiple users, but implementation is challenging and the necessary user studies evaluating the perceptual effects of these techniques are lacking. An alternative for displays with high frame rates is to present each user with his/her actual view by dividing the frame rate among multiple users. However, this approach is only possible with a small number of users, and larger groups of users result in unacceptably low frame rates and low image brightness.

2.3 Distance Perception in Virtual Environments

Egocentric and exocentric distances in virtual environments are commonly under-perceived. Depth judgments in virtual environments are often only 50-80% of the intended distance, even when the viewer receives stereo images from the correct perspective (i.e., even when standing at the CoP). Distance underestimation occurs whether virtual environments are displayed via head-mounted displays [11][12][13][14][15][16][17][18] or projection screens [19][20][21]. In contrast to the underestimation of perceived distances in virtual environments, real-world distance judgments are
typically quite accurate [9]. The exact cause of distance compression in virtual environments is unclear, and is an ongoing topic of research.

In this project we assumed that distortion resulting from displacement will occur in addition to the underestimation that is found in nearly all virtual environments, similar to previous work in our lab [4]. In other words, we expected underestimation of distance to occur even when standing at the CoP, and that displacement from the CoP would cause over- or underestimation of distance relative to judgments made from the CoP.

3 Experiment Overview and Hypotheses

The goal of this study was to evaluate the magnitude of perceived depth distortion caused by forward and backward displacement from the CoP, and whether such perceptual distortions affect collaboration when using the leader-follower approach to rendering virtual environments for multiple viewers (i.e., when all viewers receive the same stereoscopic images of the virtual environment rendered from the leader’s perspective). Pairs of participants viewed a virtual “test” object within a stereoscopic virtual environment and adjusted the depth (i.e., the front-to-back dimension of the shape) of an undistorted virtual “match” object so that it appeared to match the depth of the test object. The test object was a rectangle rendered on the ground plane 16 ft in front of participants, beyond the distance of the front projection surface (as depicted in Figures 2 and 4), whereas the match object was fixed on the floor projection surface. Because objects rendered at the exact same distance as the projection surface have zero stereo disparity (i.e., there is no difference between left- and right-eye images) they appear undistorted regardless of viewer position. This ensured that participants accurately perceived the shape of the match object. The experimenter remotely controlled the depth dimension of the match object, and participant pairs directed the experimenter to increase or decrease the depth of the match object until it appeared to be the same depth as the test object. Adjustment continued until the participants settled on a mutually agreeable response.

Participant pairs made judgments of object depth while standing side-by-side or in different locations separated by 3 feet or 6 feet (Figure 4 shows two participants standing at different locations). Side-by-side viewing allowed both participants to see the same view of the environment from approximately the same position, and we expected this would lead to similar perceptual experiences for the two participants. Viewing position could be in front of (front viewing position), in back of (back viewing position), or at the CoP (center viewing position). This allowed for six possible pair locations: front-front, front-center, center-back, front-center, center-back, or front-back. Regardless of relative positioning, participants were instructed to come to a mutually agreeable decision about the depth of the test object.

The ray-intersection model predicts that forward and backward displacement from the CoP should result in perceived depth compression and expansion, respectively, and that distortion in perceived depth should be of similar magnitude regardless of displacement direction (forward or backward). Previous work from our lab [4] suggests that perceptual distortions will be in the predicted direction but of smaller magnitude than the model predictions. We expected to replicate those findings when both participants stood side-by-side at the same distance from the CoP (front-front, center-center, and back-back conditions). Furthermore, when participants stood at different locations (i.e., at different distances from the CoP) we expected depth judgments to reflect an average of the perceptual experiences at the two occupied locations. For example, collaborative depth judgments made in the front-center condition should be in between those made in the front-front and center-center conditions, thereby reflecting an average of the perceptual experiences of the two participants in their respective locations.

4 Method

4.1 Participants

Participants were 36 students enrolled at Iowa State University. Participants received either research credit or $10 for their participation in the study. Participants were grouped into gender-matched pairs, and included 5 female pairs and 13 male pairs.

4.2 Stimuli and Design

The virtual environment was displayed within the C6 (depicted in Figures 1 and 4), a six-sided cube configuration of projection screens measuring 10 x 10 x 10 ft. Each screen was rear-projected with 4K stereo resolution per screen. The rear retractive screen was not used because it was directly behind participants, and therefore outside of their field of view. Participants wore active shutter glasses synchronized with the projectors in order to receive stereoscopic depth cues. The virtual environment was constructed using OpenSceneGraph and VR Juggler, a networking and hardware API commonly used for clustered graphics applications [27].

The virtual environment (depicted in Figure 4) contained two green rectangles—the test and match objects—on a ground plane covered with an irregular texture. The ground plane was at the same height as the CoP. The test object was on the ground plane, 16 ft. forward from the center of the C6. The match object was on the ground plane at the center of the C6.

Each participant stood at one of three possible distances from the front projection screen: 1.5, 4.5, or 7.5 ft., referred to as front, center, and back viewing positions, respectively. The CoP was rendered from the center viewing position. All three viewing positions were
shifted 3 ft to the left of center, so that the front participant did not occlude the back participant’s view of the test object (Figure 4 shows participants standing at the center and back viewing positions). Pairwise combination of the three viewing positions resulted in six viewing conditions: front-front, center-center, back-back, front-center, center-back, and front-back. Participants stood side-by-side when occupying the same viewing position (e.g., in the front-front viewing condition). Viewing condition was blocked and manipulated within-participants. Viewing condition block order was counterbalanced using a balanced Latin square design with six unique condition orders in order to prevent carryover and practice effects from biasing the data.

Within each viewing condition block, participants viewed a random sequence of nine test objects. The nine test objects comprised factorial combinations of object width and object depth, each of which could be 2, 4 or 6 ft. Only object depth was tested, but object width varied in order to prevent participants from relying on an aspect ratio strategy. Pilot testing indicated that perceived object width was unaffected by displacement in front of and behind the object. The match object was manipulated within-participants. The match object was adjustable in depth but fixed at the same width as the test object. The experimenter controlled the depth of the match object using a remote device. On each trial, participants verbally directed the experimenter to adjust the depth of the match object until it appeared to match the depth of the test object. In this way, both participants controlled the depth of the match object, and the participant pair attempted to come to a mutually agreeable response. Dependent measures were the adjusted depth of the match object and response time.

4.3 Procedure

After providing informed consent and completing a demographic questionnaire, participants donned shutter glasses and were led into the C6. The head tracking system was locked at the center viewing position, facing the front screen so that the rendered view did not update with participant head rotations or translations. In this way, the CoP was fixed at the center viewing position.

The experimenter told participants where to stand prior to each viewing condition block. On each trial, participants were shown a test object displayed on the front screen. The initial depth of the match object was set to a random value within a range of approximately 1-8 ft. The experimenter controlled the depth of the object using a joystick capable of adjusting the match object in 3-inch increments. Participants directed the experimenter if they wanted to increase or decrease the depth of the match object, and thereby adjusted the depth of the match object until it appeared identical to the depth of the test object. These verbal directions were typically brief statements, such as “A little bigger,” or “I think it needs to be one foot smaller.” By restricting manipulation and trial advancement capability to the experimenter, each participant was given an equal chance to convey his or her opinion and an agreement could be checked between participants before the response was entered and the next trial began. Response time began at the moment the test object appeared and ended when both participants were satisfied with the depth of the test object. Before entering a response, the experimenter requested verbal confirmation from each participant that the response was acceptable.

5 Results

Depth judgments and response times were analyzed in separate repeated-measures ANOVAs, and specific hypotheses were evaluated using planned contrasts.

5.1 Judged Depth

Depth judgments made in each viewing condition and for each test object depth are shown in Figure 5. Regardless of viewing condition, larger depth judgments accompanied increased object depth [F(2,34)=689.95, p<.001, $\eta^2=.98$]. Furthermore, viewing condition also had a significant effect on judged depth [F(5,85)=62.28, p<.001, $\eta^2=.79$]. The ray-intersection model only makes clear predictions about depth judgments made when both participants stood side-by-side at the same distance from the projection screen (front-front, center-center, and back-back conditions), and so we focus on those conditions first. Depth judgments made in the center-center viewing condition (i.e., judgments made while both participants stood side-by-side at the CoP) were significantly compressed, averaging 86.3% of actual object depth ($t(17)=4.33, p<0.001$). This is consistent with previous reports of perceived depth compression in virtual environments [9].

Based on the depth compression when participants stood side-by-side at the CoP, depth judgment predictions generated by the ray-intersection model were scaled accordingly [A]. Specifically, the model-predicted depth for each viewing condition and object depth was scaled by 86.3% (scaled predictions for forward and backward displacements are labeled as “front prediction” and “back prediction” in Figure 5. Depth judgments made in the front-front condition were 26.8% smaller than judgments made in the center-center condition. This compression was significantly less than the scaled predictions of the model, in which depth judgments in the front-front condition were predicted to be 41.7% smaller than judgments made in the center-center condition ($t(17)=4.01, p=.001$). Depth judgments made in the back-back condition were 14.9% larger than judgments made in the center-center condition. This expansion was significantly less than the scaled model predictions, in which judgments in the back-back condition were expected to be 41.7% larger than judgments made in the center-center condition ($t(17)=6.408, p<0.001$). Furthermore, absolute distortion relative to judgments made in the center-center condition was larger after forward displacement compared to backward displacement ($t(17)=3.93, p=.001$). This indicates an asymmetric bias toward larger distortion after forward displacement compared to backward displacement. These findings are quite consistent with those of a prior study from our lab [4] in which individual participants viewed and judged object depth (see Figure 3).

In sum, distortion caused by displacement of both participants in the same direction depended on the direction of displacement: relative to judgments made when standing side-by-side at the CoP,
forward displacement caused compression of depth judgments whereas backward displacement caused expansion of depth judgments. Although distortion of depth judgments was in the direction predicted by the ray-intersection model, distortion was smaller in magnitude than the model’s predictions. Also contrary to the model’s predictions, distortion was asymmetric, whereby forward displacement resulted in greater distortion than did backward displacement.

When participants stood at disparate locations (front-center, center-back, and front-back conditions) depth judgments were expected to fall in between those made when both participants occupied the same location, reflecting a compromise between the two perceptual experiences at participants’ disparate locations.

Evidence supporting this prediction can be seen in Figure 5. Depth judgments in the front-center condition are between (and significantly different than) those made in the front-front [F(1,17)=21.08, p<.001, η²=.55] and center-center [F(1,17)=20.52, p<.001, η²=.55] conditions. Similarly, judgments in the center-back condition are between (and significantly different than) those made in the center-center [F(1,17)=11.30, p=.004, η²=.40] and back-back [F(1,17)=9.84, p=.006, η²=.37] conditions. Lastly, judgments in the front-back condition are in between (and significantly different than) those made in the front-front [F(1,17)=75.10, p<.001, η²=.82] and back-back [F(1,17)=56.68, p<.001, η²=.77] conditions, and nearly identical to those made in the center-center condition [F(1,17)=2.25, ns].

5.2 Response Time

Time required to reach collaborative agreement on depth judgments was predicted to depend on the relative standing locations of the two viewers. We hypothesized that when viewers stood side-by-side at the same location (three red bars on the left of Figure 6), any perceptual distortion experienced should be similar for both viewers, and so agreements should be reached relatively quickly. In contrast, when viewers occupied disparate locations (three blue bars on the right of Figure 6) perceived depth should differ between the two viewers and the time required to reach agreement should be longer as a result of the perceptual discrepancy between viewers. Furthermore, we expected larger discrepancies in perceived depth would result in longer times to reach agreement.

Response latency across condition was evaluated using the predicted pattern testing procedure [28][29]. This procedure evaluated the fit of planned contrasts that coded the pattern of performance predicted by our hypotheses. Based on judged depth in the front-front, center-center, and back-back conditions, response times were expected to be ordered as follows (from fastest to slowest): front-front = center-center = back-back < center-back < front-center < front-back. Using the condition ordering from left to right shown in Figure 6, the contrast weights are -1, -1, -1, 1, 0, 2. This predicted pattern is based on the amount of discrepancy in perceived object depth between the two participants at their respective locations (derived from data shown in Figure 5). There is no discrepancy in perceived depth when participants stand side-by-side in the same location, and so response times should be shortest for those three conditions. The discrepancy in perceived depth when standing at the center versus back location is relatively small (see Figure 5; depth judgments in the back-back condition were only 15% larger than in the center-center condition), and this should result in a small increase in response time. The discrepancy in perceived depth when standing at the center versus front location is larger (see Figure 5; depth judgments in the front-front condition were 27% smaller than in the center-center condition), and should result in a greater increase in response time. Finally, the discrepancy in perceived depth when standing at the front versus back location is largest, and should result in the longest response times.

Response latencies were analyzed in a one-way ANOVA to evaluate the effect of viewing condition on time required to reach agreement. The effect of viewing condition was significant [F(5,85)=5.35, p<.001, η²=.24]. The contrast testing the pattern predicted by our hypothesis significantly fit the data [F(1,17)=31.45, η²=.65], accounting for 97.9% of the variance associated with position and leaving a non-significant amount of variance unaccounted for [p=0.94]. In sum, change in response latency across condition was well-described by our hypothesis that larger discrepancies in perceived depth between the two participants would result in longer times to reach agreement.

6 Discussion

The primary purpose of this project was to evaluate the distortions that occur when viewing stereoscopic virtual environments after displacement from the CoP, and the impact of those distortions on collaborative judgments within the virtual environment. Participant pairs made collaborative judgments of object depth within a virtual environment, which they viewed while standing side-by-side in approximately the same location or in different locations. Viewing location was at the CoP or displaced forward or backward.

When participants stood side-by-side at the CoP, judgments of object depth were compressed relative to actual object depth, consistent with past studies reporting under-perception of distance and depth in virtual environments even when stereo cues are appropriate for the viewing location [9]. When participants stood side-by-side after forward and backward displacement, depth judgments were compressed and expanded, respectively, relative to judgments made when standing side-by-side at the CoP. However, the distortion that resulted from forward and backward displacement was significantly less than the predictions of the ray-intersection model based on stereo viewing geometry [2][3]. Past studies on perceptual distortion after displacement from the CoP reported that perceived distortion could be fully described by the ray-intersection model [1]. However, the divergent findings of the current study and past research might be due to differences in available depth cues. In the study by Banks and colleagues [1], participants viewed a stereoscopic virtual hinge floating at eye-level, and the only monocular depth cue was provided by a grid-like texture gradient on the hinged surfaces. In contrast, participants in the current study viewed a virtual shape on the ground plane, which provided monocular depth cues defined by the texture gradient of the ground [30] and also the angle of declination from the eyes to the object on
the ground [31]. The availability of these additional monocular depth cues might have allowed participants to partially correct for the distorted stereo cues by combining multiple depth cues. This seems especially likely in light of research showing that spatial judgments when viewing monocular displays are relatively unaffected by displacement from the CoP [1][5].

Forward and backward displacement relative to the CoP in the current study is functionally similar to image magnification and minification, respectively. Image magnification has been shown to cause an increase in egocentric distance judgments, whereas image minification causes a decrease in egocentric distance judgments [13][32][33]. The direction of the effects produced by magnification and minification are consistent with the effects produced by forward and backward displacement in the current study, although the relative magnitudes of these effects have not been directly compared. The distortion in depth judgments caused by forward and backward displacement from the CoP was asymmetric, such that forward displacement resulted in larger distortion than did backward displacement. This finding of asymmetric distortion was not predicted by the ray-intersection model. The cause of the asymmetry is unclear, but previous work from our lab testing individual participants (rather than pairs of participants) have produced this same finding [4]. Those data are shown in Figure 3. One possible explanation for the asymmetric distortion is that the asymmetry arises from distortion of monocular depth cues, but further research is required to evaluate this hypothesis. Spatial judgments are known to rely on combined estimates from stereoscopic and monocular depth cues [34][35]. If distortions in depth judgments are asymmetric when viewing monocular displays after displacement forward and backward from the CoP, then this would provide a potential explanation for the asymmetry found in the current studies and could explain why an asymmetry would occur even though the ray-intersection model predicts symmetric distortion. However, no existing studies have evaluated distortions in spatial judgments when viewing monocular displays after forward and backward displacements, and this remains an open topic for future investigation.

Asymmetric distortion caused by forward and backward displacements from the CoP points to a novel method for displaying multi-user virtual environments. In the leader-follower approach, the leader receives the projectively correct stereo images and followers receive stereo cues intended for the leader. The leader role does not typically change, regardless of the relative positions of users in the environment. In light of the asymmetry in perceived depth distortions, a better approach would be to assign the leader role to the user nearest the projection screen, and we refer to this as the “role-switching” method. Backward displacement from the CoP causes less distortion than does forward displacement, so the average distortion experienced by multiple users will be reduced if the frontmost user is assigned the leader role. As users walk around the environment, one of the followers will eventually stand closer to the projection screen than the leader, and their roles will switch. This approach is especially relevant in virtual reality systems with a clear primary display, but also in six-sided virtual reality systems. Multiple users commonly look in the same direction as one another (see Figure 1 for an example). When the group’s focus of attention shifts to another screen, the user nearest that screen is assigned the leader role. Additional research is needed to determine the appropriate assignment of the leader role when users are looking at different projection screen surfaces.

The potential benefit of assigning the leader role to the frontmost user is evidenced by the collaboration times in the current study. Time to reach agreement on perceived object depth increased monotonically with increased discrepancy between the perceptual experiences of the two participants. This finding parallels past research on common ground, in which the success of collaborative tasks such as repairing a bicycle depends on shared knowledge and shared awareness of the environment [25]. Discrepancies in perceptual experiences are likely to reduce common ground between users, and thereby impair performance on collaborative tasks performed in the virtual environment.

Compared to existing techniques for displaying virtual environments to multiple users, the role-switching method holds multiple advantages. Image blending and view clustering techniques [6] have the potential to provide projectively correct images to all users, especially when they are looking in different directions. However, when the gaze directions of multiple users overlap (as in Figure 1), the rendered images are a blend between multiple perspectives, and this blending is likely to introduce additional distortion. One advantage of the role-switching method is that the images are rendered from a single user’s perspective, which prevents any artifacts caused by image blending. Furthermore, image blending and view clustering methods present greater computational challenges than the role-switching method, which is relatively simple to implement. Frame-rate reduction is another approach to displaying virtual environments to multiple users, whereby the frame rate is split evenly among users. For example, a display capable of refreshing at 120 Hz can be used to present half of the images to one user and the other half to a second user. In this way, each user receives the projectively correct stereo images of the virtual environment at 60 Hz (30 Hz per eye). However, large numbers of users result in very low frame rates and low image brightness, whereas the role-switching and view blending methods can be used with large numbers of users and no reduction in frame rate or image brightness.

It is unclear what impact the role-switching method will have on the way users perceive and interact with virtual environments. It is possible that switching the leader and follower roles will be disruptive to collaboration between users, or that it will disrupt lower-level tasks such as navigation. Additional research is required to evaluate the role-switching method in a more typical collaborative context. Furthermore, the role-switching method might apply best to scenarios in which both users are looking at the same region of space (i.e., when their fields of view overlap). When multiple users view substantially different parts of the virtual scene, then other techniques, such as image blending [10], might be better. Additionally, future work should address whether the relative heights of the users will affect their perceptions of spatial layout and their responses to the role-switching method. Relative participant height was neither measured nor controlled in the current study.

When participants in this study stood in disparate locations, depth judgments reflected a combination of the perceptual experiences at each individual location. For example, judgments made in the front-center condition were larger than those made in the front-front condition but smaller than those made in the center-center condition. Similar response patterns occurred in the center-back and front-center conditions. Anecdotally, participants who stood at different distances typically attempted to strike a compromise between the experiences of each individual participant. Participants frequently commented that they thought the match object should be a little bigger (or smaller), but they were willing to accept a slightly smaller (or larger) response in order to accommodate their partner’s preference. This suggests that participants were aware that their depth judgments were biased by their standing location, and that they would need to compromise in order to come to a mutually agreeable response. Participants were only instructed to reach a mutually agreeable response, and were not explicitly instructed to consider that their viewing position might result in a different perceptual experience than that of their partner, but they were probably able to infer this after viewing the environment from multiple locations over the course of the experiment. Different instruction sets might influence participant strategies, such that an emphasis on determining actual object depth could result in a failure to reach agreement or reliance on only one participant’s input.

In summary, the current study indicates that distortions in judged depth caused by viewing virtual environments after displacement from the CoP are smaller than predicted by models of stereo viewing geometry, and that distortion is greater after forward displacement.
than after backward displacement. More research is needed to understand the causes of these distortions, and why they diverge from model predictions. Regardless of the displacement direction, larger discrepancies between the perceptual experiences of multiple users led to greater difficulty in making collaborative judgments about the environment. We propose a novel role-switching method for reducing perceptual discrepancies by assigning the leader role to the front-most user in multi-user virtual environments, which capitalizes on the asymmetric distortion after forward and backward displacement.

FOOTNOTES

[A] This scaling was done under the assumption that the foreshortening of judged depth that occurs when viewing virtual environments from the CoP also occurs when viewing from locations displaced from the CoP.

ACKNOWLEDGMENTS

This work was supported by a grant from the United States Army RDECOM – STTC.

REFERENCES