# Exploring the Effects of Image Persistence in Low Frame Rate Virtual Environments

David J. Zielinski*
Duke immersive Virtual Environment
Duke University

Hrishikesh M. Rao†
Dept. of Biomedical Engineering
Duke University

Marc A. Sommer‡
Dept. of Biomedical Engineering
Duke University

Regis Kopper§
Duke immersive Virtual Environment
Duke University

## ABSTRACT

In virtual reality applications, there is an aim to provide real time graphics which run at high refresh rates. However, there are many situations in which this is not possible due to simulation or rendering issues. When running at low frame rates, several aspects of the user experience are affected. For example, each frame is displayed for an extended period of time, causing a high persistence image artifact. The effect of this artifact is that movement may lose continuity, and the image jumps from one frame to another. In this paper, we discuss our initial exploration of the effects of high persistence frames caused by low refresh rates and compare it to high frame rates and to a technique we developed to mitigate the effects of low frame rates. In this technique, the low frame rate simulation images are displayed with low persistence by blanking out the display during the extra time such image would be displayed. In order to isolate the visual effects, we constructed a simulator for low and high persistence displays that does not affect input latency. A controlled user study comparing the three conditions for the tasks of 3D selection and navigation was conducted. Results indicate that the low persistence display technique may not negatively impact user experience or performance as compared to the high persistence case. Directions for future work on the use of low persistence displays for low frame rate situations are discussed.

**Keywords:** virtual environments, low-persistence, simulator sickness, performance, presence.

**Index Terms:** H.5.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

## 1 INTRODUCTION

When authoring virtual environments (VEs), we often try to make design decisions to ensure the frame rate is at the highest frame rate supported by the display device. However, there are several reasons why low frame rates may be unavoidable. Among them, input devices, computational load, and scene complexity are common reasons why VEs need to run at low frame rates.

For input devices, many VR researchers are used to using high speed low latency tracking systems in a controlled laboratory setting. For example, a current high-end product from ART provides updates at 150 to 300 hertz [3]. However, when outside the confines of the laboratory, with objects tracked in large spaces, lower fidelity tracking systems are the only option. For example, recent work studying UAV's (Unmanned Aerial Vehicle), otherwise known as drones, uses tracking data from GPS systems [30] which do not update at high frame rates. In practice, in this study, the authors had to work with update rates of approximately 9 hertz [30]. Several virtual reality (VR) scenarios have utilized similar low speed GPS

input data: real-time command and control situation rooms displaying GPS tracked devices [27], utilizing Head-Mounted Displays (HMDs) for untethered large space exploration with real walking [15], and finally various augmented reality (AR) applications that may take place over a large physical space [29, 13].

For computational load, we can consider virtual experiences that require large amounts of calculation. For example, when trying to fit (correct) the molecular structures captured via X-ray crystallography or NMR, one would ideally provide feedback to the user on whether the current configuration is physically possible. This can be done by calculating and displaying the all-atoms contacts metric [8]. However, we may not have enough computational resources available to re-calculate this at a high frame rate. This would cause the system to show the structure updating at a slow frame rate.

Finally there are issues with scene complexity. Current GPUs, although faster and faster every year, are still limited in the number of triangles they can display per second. If the system attempts to display a VE with more triangles than its video card can handle, the frame rate will be reduced. The first choice would be to reduce the complexity of the scene by removing objects. However, in training simulations, it has been found that scene complexity reduction may have negative effects. For example, in a visual search training simulation in which the user had to scan for threats in an urban environment, it was found that users were better able to develop proper scanning techniques, and thus result in higher threat detection performance [37]. So, if the scene complexity is simply reduced to achieve higher frame rates, it may actually change the training experience of the user. Motivated from the fact that VEs with low frame rates will exist for the foreseeable future, this paper presents an investigation on the effects of image persistence in such conditions as compared to high frame rate VEs.

## 2 IMAGE PERSISTENCE IN LOW FRAME RATE SYSTEMS

As discussed previously, in many situations VEs have to be displayed at low frame rates. Here, we discuss the concepts of input latency and image persistence in such situations.

### 2.1 Low Frame Rate and Input Latency

It has been previously found that low frame rates lead to several negative outcomes: increased simulator sickness [21], decreased presence [24, 4], and decreased task performance [40, 11]. In fact, a recent project has found all such effects within the same user study [44]. In order to better understand what may be causing these effects, it's important to look closely at two of the components affected by low frame rate: input latency and input update rate.

Input latency (sometimes referred to as transport delay [21]), is the time from the user making an action (e.g., moving their hand) to the time there is visual feedback (e.g., the image on the screen changes color to indicate selection). Previous research has found that increasing the input latency alone leads to lower presence [25] and lower task performance [40, 11]. Input update rate is the rate at which the user actions (e.g., pressed buttons on the controller) are processed and used to change the outcome of the simulation. Often this value correlates with the frame rate, as a feedback loop is standard practice: read the inputs, update the simulation, render the current virtual environment. In this paper, we use an update rate that is matched to the frame rate.

*e-mail:djzielin@duke.edu
†e-mail:hrishikesh.rao@duke.edu
‡e-mail:marc.sommer@duke.edu
§e-mail:regis.kopper@duke.edu

## 2.2 High Persistence

Looking at the component of visual display, there are some issues related to low frame rate displays. When a system is running below its maximum frame rate, it will repeat the previous frame (on the display) while working on generating the next frame (see figure 1). The television industry refers to a particular sub-class of frame repetition as Judder. Judder occurs when upsampling a 24 frames per second (fps) movie to a 60 fps television. In that case some frames are repeated 2 times and other frame are repeated 3 times. The different frame persistencies can cause visible artifacts if there is motion in the scene. More recently in the television field this issue has been refered as FRU (frame rate upsampling), as many new TVs display at 120Hz. In our experience, the visual artifact of low frame rate high persistence that is easily seen is that a vertical line, when moving horizontally, is seen as multiple copies of that line. In order to be exact in how many repeated frames we are showing to the user, we have created a software-based display simulator. This simulator runs at our full system frame rate (55 fps), referred to as the high frame rate(HFR) case, but can be programmed to repeat a single frame as long as necessary. While creating this software we had to make design considerations related to input latency. We decided to use low input latency, in which input values are read and the simulation is updated in the frame before the frame is initially displayed, as if the scene were processed at full fps (figure 1). While it may take away from ecological validity, this design choice was made as it's already known that higher input latency has negative effects [25, 40], and our goal was to look more closely at the issues arising from the visual display motion artifacts. For the remainder of this paper the low frame rate high persistence condition will be referred to as HP.

## 2.3 Low Persistence

In addition to looking at high persistence, a new technique is proposed, referred to as low frame rate low persistence display. This term has been recently used in relation to the Oculus Rift HMD [28], which uses the term to indicate a lower duty cycle to reduce motion blur and judder. Our use of the term and technique is similar, although our goal is to explore the technique in low frame rate situations. Another way to describe this low persistence technique would be black frame insertion, which is a technique where black frames are inserted to reduce blur in LCD displays [20]. For the remainder of this paper the low frame rate low persistence condition will be referred to as LP.

Specifically, our use of the low persistence technique involves displaying the original image generated, and then instead of repeating the image (i.e., the high persistence), black is displayed instead. This is achieved by a simple OpenGL call (glClear) after the regular frame is rendered. Another design consideration necessary was in regards to the fps for the low frame rate cases. Our system runs at 110Hz active stereo (thus 55 fps per eye effective frame rate). Dividing this by integer multiples leads to: 27.5, 18.3, 13.75, 11, etc. While all values should eventually be tested, in order to limit the scope of the experiment we decided to test 11 fps versus the normal case of 55 fps. We did this because we found the LP and the HP side effects (strobing effect for LP, and multiple lines for HP) were both very visible at this setting, but the simulation still felt responsive and interactive. Also, 11 fps is a lower limit of interactive frame rates.

## 2.4 Active Stereo

In our CAVE-type system, binocular disparity is achieved through active stereo, running at 110Hz. For purposes of describing the system, and the frame diagram (figure 1), we consider our system as having an effective frame rate of 55Hz, as that is the frame rate at which a stereo pair is presented to the eyes. It's important to note that, in some ways, a 110/55Hz active stereo system even in the HFR condition could be considered a lower persistence display than other types of displays. Specifically, active stereo has a duty
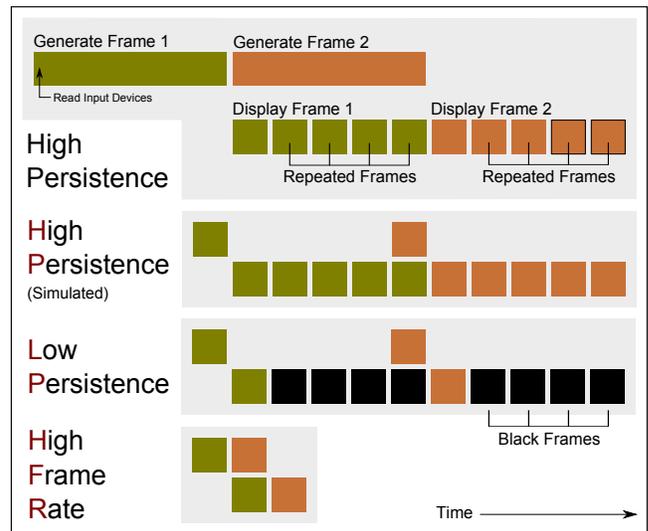


Figure 1: Timeline depicting frame generation and display to user for cases: High Persistence , HP (High Persistence Simulation), LP (Low Persistence), and HFR (high frame rate).

cycle of at most 50% (i.e., the user sees black for 50 percent of the time). As examples of higher persistence displays, we could consider a passive polarized system with 2 projectors running at 55Hz (or more), projected at a single wall or HMDs (excluding the newer Oculus low persistence display). These other displays are worth exploring, however in this study we will focus on the effects of low persistence low frame rate conditions, while inside an active stereo CAVE-type system.

## 3 RELATED WORK

We look at previous research related to interpolation to increase frame rate, the effect of stroboscopic images and on physiological effects of visual persistence.

### 3.1 Interpolation

Previous work has dealt with low frame rate issues by using interpolation between frames, which is a common technique in television displays. Many new TVs are capable of displaying at 120Hz, yet may receive input from movies (24 fps) or regular tv (60 fps). The simple average between two frames as the interpolated frame is not an acceptable solution, and researchers have worked on algorithms to improve it. For example, previous work developed advanced techniques often referred to as motion compensated frame rate up sampling [39]. These interpolation techniques often rely on knowing what the next frame is, which is not feasible in realtime interactive VEs. One alternative would be to delay the stream by one or more frames, which is typically not a problem for TVs playing back non-interactive content. However, with interactive content, adding an additional frame of delay may cause more problems then the benefits from the interpolation. Other researchers have relied upon prediction for interpolation [36]. In this case, there were issues with the interpolation technique relating to occlusion. Also, prediction may work for head data [1], perhaps because users don't instantly start or stop head motions. However, predicting motion of objects in the scene may not be so easy, as objects may very well start or stop instantly. So, in certain situations, interpolation with prediction may be a useful tool to reduce the effects from low frame rates (effectively making them look like the HFR case). This, however, is an application-specific solution which may not be generalizable to a wide range of VEs.

Another related technique, known as "time warp", is utilized in the new Oculus Rift DK2 [28]. With this technique, the tracking

system is sampled after the image is rendered, but before it is displayed to the user. The system warps the rendered image using the most recent tracking data, changing the viewpoint to match the latest tracking info, and reducing latency with head movements. Because the image is warped between two tracking points, such technique only works on high frame rate systems, as the re-projection process would show artifacts in low frame rate situations.

## 3.2 Stroboscopic Glasses

As the low persistence technique we propose to test has the effect of a strobe light, it is important to look at some of the previous research related to stroboscopic glasses. Researchers have interestingly found that stroboscopic glasses (at the 8Hz level) are effective in reducing nausea [31]. The exact cause is unclear, but the prevention of retinal slip is cited. Another interesting research study showed that, by utilizing stroboscopic glasses during american football training, central visual field motion sensitivity and transient attention abilities were increased [2]. These results provide evidence that displays with stroboscopic properties, present in the proposed LP technique, may be useful for training, and may have some nausea reduction properties.

## 3.3 Neuroscience Perspectives on Visual Persistence

Previous research from the neuroscience community investigated what happens when images are presented in rapid succession. It was found that these discrete images appear to the subject as a single continuous movie. This phenomenon is termed flicker fusion [34]. Subjective experience as well as neural evidence suggest that stimuli presented less than 40ms apart (thresholds generally 25-50 Hz) will appear as a smooth continuum of images, and that this perceptual fusion involves neural systems beyond the retina [5], likely involving frontal and parietal brain areas [17]. Naturalistic, large field movies achieve adequate fusion with frame rates of 24 fps.

Movements of gaze can interact with flickering stimuli to create stimulus-dissociation artifacts. For example, in a simple psychophysical experiment, naive subjects were asked to make a rapid eye movement (saccade) across a flashing light [14]. Observers subjectively experienced that, as the eye was in flight, there was a phantom row of lights appearing in the opposite direction of the motion of the eye. While that is an example of eye movement causing the phantom array illusion, in our study, the same illusion is obtained when the eye is stationary but the visual scene moves. In both cases, there is motion of the image across the retina. We hypothesize that this could cause a break-down of the persistence of vision in sub flicker-fusion threshold (sub 17 fps) conditions. Perhaps the visual system is attempting to interpolate between images, and in the HP condition the repeated frames provide input inconsistent with (what should be) a smoothly moving line.

Interestingly, in the LP case, we have not observed any of the phantom lines, and the motion of the line feels smooth. In visual neuroscience, a mask is neutral high contrast stimulus that is interleaved within successive frames of the target stimulus (or the movie of interest) [19]. In our case, instead of high contrast stimuli, black frames are inserted between target frames. It is suggested that masking interrupts the ongoing processing of visual information and resets the feedforward & feedback mechanisms that help us perceive what we see [23].

## 4 USER STUDY

We evaluated 3 conditions: HFR (55 fps), HP (11 fps), and LP (11 fps), for the tasks of selection and navigation. The goal was to explore the differences among these conditions in terms of presence, simulator sickness, task performance, and user preference.

## 4.1 Apparatus

We used a six-sided CAVE-type [7] system to perform the experiment. Since the experiment involved the user always looking towards the front screen, the door was left open and the rear projector turned off, effectively making the system a 5-sided CAVE-type system for the experiment. Leaving the door open allowed us to watch the participant to ensure adherence to the experiment protocol. Head tracking with 6 degress of freedom (DOF) was provided by an Intersense IS-900 tracker for both tasks. For the selection tasks, participants used an Intersense IS-900 6 DOF wand controller, and for the navigation task, participants used a wireless Microsoft Xbox gamepad. Crystal Eyes CE3 shutter glasses synced to projectors running at 110hz provided 3D stereoscopic graphics. A cluster of computers (1 computer per projector) plus one computer as a master computer powered the system. All computers had NVIDIA Quadro FX 5600G graphics cards. The application for the different display cases (HFR, HP, LP) and tasks (selection,navigation) was written in Syzygy [32]. We verified the frame correctness of the application by utilizing an Andor high speed camera, at 480 frames per second.

## 4.2 Design

We used a fully crossed within-subjects design with two independent variables for the tasks of selection and navigation.

For selection, the independent variables were: display method (HFR, HP, LP) and target size (angular widths 1°, 2°, 3°). Time to successfully complete the task and number of errors were selection-specific metrics. An error happened when the user clicked with the wand outside the target, but did not invalidate the trial. Thus, a single trial could have multiple errors.

The navigation task had display method (HFR, HP, LP) and path width (45.7cm, 30.5cm, 15.2cm) as the independent variables. The metrics for the navigation task were: time to traverse the path, number of times the user got off the track and time spent off track.

After each display method was performed, participants filled out surveys on presence [35] and simulator sickness [18] . At the end of the study, participants filled out a display method preference survey on selection, navigation, and overall preference. The order of the display method was counterbalanced. We conducted the study over three days for each participant. Each day we selected the display method based on the counterbalancing schedule, and utilized that method for the two tasks (selection first, then navigation) for that day. We decided to not counterbalance the ordering of the selection task and navigation task, as the tasks were not utilized as independent variables. Each task was run three times, each with a different target or path size. We started with the easiest (largest) targets, then medium, and finally hardest. We decided that counterbalancing the target size would not have been possible due to the limited number of participants. We also felt that by going from easy to hard allowed the user additional training on the system, so they were better prepared for the hard case. From observation, we found that having the color of the scene closer to the blanking color (in our case black), caused less visual fatigue for the LP condition. Thus, the scenes were constructed to have a dark backdrop (figures 2 and 4).

### 4.2.1 Selection Task

For the selection task, we used a ray-casting selection [26] based on the ISO9241-411 standard [16] (figure 3). In this task, a series of circles (only one visible at a time) appear in front of the user. We chose the location of the circles to appear on the front wall of the system (1.45 meters in front of the user). In this task the user saw a green circle target, and she could then move the ray controlled by the wand to intersect with the circle. When contact was made, the circle changed color to red. The user would then click the trigger button on the IS-900 wand, and the current target would disappear, and the next target would appear. Figure 3 shows the sequence of targets. This selection task was used because it is a standard that has been utilized in previous studies [43, 9]. We decided to test three target widths to reflect 1°, 2° and 3° of visual angle. The circle of targets (figure 3) was centered on the user hand (calibrated at the beginning of the session) and each target was 32.67° of visual angle apart. Such angular measures resulted in selection tasks that were easy ($ID_{DP} = 1.31$), moderate ($ID_{DP} = 5.50$) and difficult ($ID_{DP} = 25.74$) [22]. There were 15 targets per cycle (see figure 3)
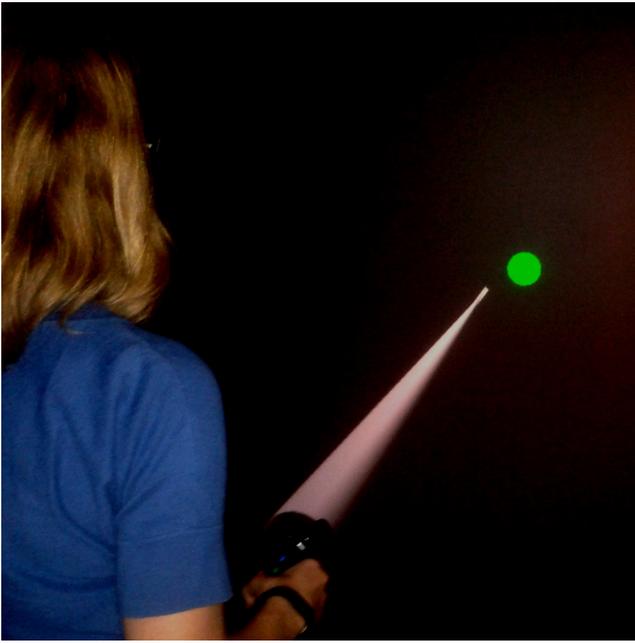
Figure 2: Example of a user engaged in the selection task.



Figure 3: Example of ISO9241-411 selection task. User moves to select target 2 (in green). Sequence of targets as text labels.

and participants went through 5 cycles per size, such that each size had 75 total targets to be selected. Thus, for each experimental session, participants selected 75 large targets, followed by 75 medium targets, and finally 75 small targets.

### 4.2.2 Navigation Task

For the navigation task, we used a locomotion task consistent with the work by Zhai and Woltjer [42], where users navigated down a hallway (figure 4). In such a task, by narrowing the path width, users are expected to move slower and to move off the path more often. The path was 127.9m long and the overall width of the hallway was 137.2cm. A path inside the hallway had different widths for each trial: 45.7cm, 30.5cm and 15.2cm. In order to travel forward the user could variably squeeze the right trigger on the Xbox gamepad. While on the path, the user could travel at a maximum speed of 213.4cm/s. When the user went off the path, the floor was highlighted in green and the maximum speed was reduced to 64 cm/s. This effectively provided an incentive for users to stay on the path, as in early pilots we observed that some users ignored such instructions. The user was able to steer by moving the left joystick on the Xbox gamepad, which was sensitive to the amount the user pressed.

### 4.3 Participants

Twenty-two participants were recruited from the University community. Of those, 18 participants (6 female) completed all 3 days of the study and were compensated with 20 USD. Ages ranged from 18 to 51 years (M=25.77; SD=9.06). Thirteen participants self identified as right-handed, four as left-handed, and one as ambidextrous. For the actual experiment all participants used their preferred hand (16 right and 2 left) for the selection task. The navigation task with the Xbox controlled was two-handed and used the same controls (left joystick steering, right trigger speed) for all participants.

In terms of experience, 16 participants (89%) had previous experience with video games with natural/gestural interactions, (e.g., the Wii Remote, the Microsoft Kinect, or the PlayStation Move). All participants had experience with video games with gamepad interactions (e.g., Nitendo NES, Xbox, PlayStation). 9 participants (50%) had been inside a CAVE-type VR system before, and 8 (44%) had experience with an HMD system before.
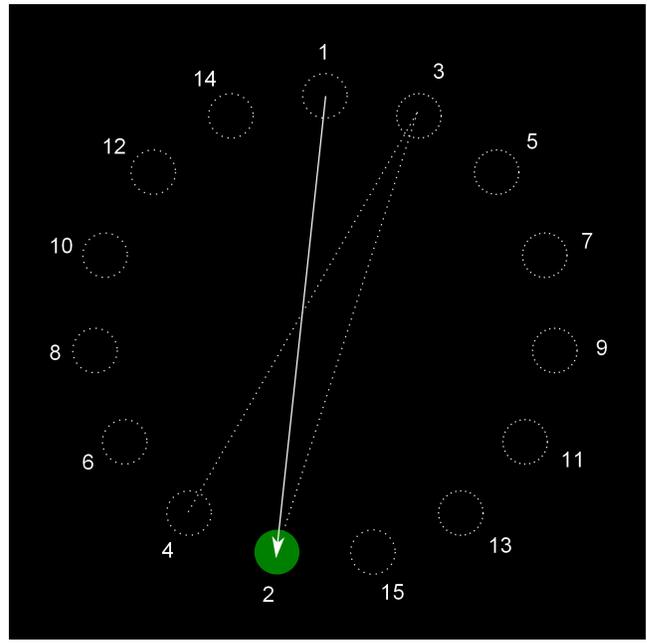
### 4.4 Procedure

Figure 5 shows the experimental procedure for one ordering of display method. The experiment took place over three sessions, with each session presenting a different display method (HFR, HP, LP). Each session was on a different day, and all participants completed the experiment (all 3 days) within 1 week of starting the experiment. The experiment was split across three days to allow participants to recover from any simulation sickness they might have experienced while participating in the experiments.

On the first day participants reviewed and signed an informed consent form. Participants were then assigned the respective condition for the day based on the counterbalancing schedule (HFR,HP,LP). After that, they were introduced to the tracked shutter glasses and the wand pointing device. At the beginning of each selection task there was a calibration phase where the user would comfortably hold the wand, and point at the front wall. They would click the trigger to register the position. This position was used to verify that the participant was in fact near the center of the system (required to be within 5 cm), so that no participants had differences by being too close or too far from the front screen where the selection targets appeared. The second purpose of the calibration phase was to capture the height off the ground of the neutral position. We utilized this number when calculating the center of circle for the display of the selection targets. Thus the location (in the y axis) of the targets was set for each participant to account for differences in height. We then described the task to the users, and reminded them to select the targets as fast as they could, without making mistakes. Clicking the button while not making contact with the target was labeled as a mistake and logged for analysis. The users were allowed to practice for up to one minute with large targets. After practice, the program was restarted and participants performed 75 large, 75 medium, and 75 small selection tasks.

After completing the selection tasks, participants took a break, during which time the navigation task was described. At that point, the wand device was switched for the Xbox gamepad and the control system and the task were described. They were allowed to practice on the narrow path for up to one minute, when we made sure that the participants understood the concepts of being on and off the
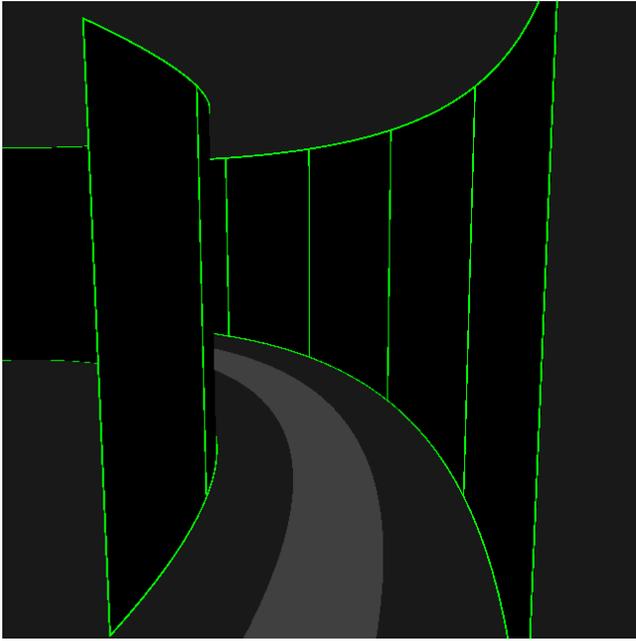
Figure 4: Screenshot of navigation task showing the hallway and the path on the ground for the user to follow.

path. We then restarted the program and they performed a run on the large path, medium path, small path. The path was always the same for all conditions.

After completing the tasks, participants filled out presence [35], and simulator sickness [18] questionnaires. On day two participants filled out a background survey, to indicate their experience with videogames and virtual reality. (This was done in day two to balance the time participants spent filling out paperwork on each day). At the end of the final session, participants were asked to determine their preferred technique and to provide additional feedback.

## 5 RESULTS

### 5.1 Selection task

For the selection task, the first target selection of each target size was not considered, as it did not start from an opposite target (figure 3). Thus, $14 + (4 \times 15) = 74$ selections per participant per target size per display method were considered in the analysis. The total selection trials per participant was, then, $3 \times 3 \times 74 = 666$.

#### 5.1.1 Outliers

One participant was dropped from the data analysis as he did not follow the instructions for the selection task. The experimenter instructed participants to be as fast as possible while not making errors. This participant made errors in 495 of the 666 trials (74.3%), making as many as 23 errors in a single trial. We believe this participant took a "machine gun" approach to the selection task, which is not consistent with the intended task. Thus, 17 participants were considered in the analysis.

We considered an outlier any trial that took more time than 2 standard deviations from its condition (display method × target size) mean, which is consistent with the literature [12, 22]. From this criteria, 408 data points (3.6%) were discarded as outliers. A further 45 trials (0.42%), which contained 3 errors or more but were not discarded as time outliers, were also discarded as error outliers. Again, we discarded trials with more than 2 errors on the basis that participants were instructed to not make errors.



Figure 5: Example of study procedure for a participant with ordering HFR-LP-HP.

#### 5.1.2 Data Analysis

IBM SPSS 22 was used for analyzing the experimental data. For the selection task, data was aggregated for all the repetitions in each selection task (4 for the first selection, 5 for the remaining 14, minus outliers). Previous work has used data aggregation for repeated measures analysis [22], as multiple readings of a single trial provide a truer mean for the motor behavior task of selection by pointing.

A 3-way factorial analysis of variance (ANOVA) with repeated measures and one between-subjects factor was used on the mean time to select a target and on the mean number of errors. Along with the display method and target size, we included the movement direction as a factor in the analysis. Movement direction is defined as the direction of the line from the initial pointing position to the target (figure 3), and had 15 levels. Ordering was the between-subjects factor. The analysis, thus, considered the following design: $(3 \times 3 \times 15)$ WTH $\times 6$ BTW. Data was tested for sphericity, and whenever the assumption of equal variances was violated, the Greenhouse-Geisser correction on the degrees of freedom was used.

#### 5.1.3 Movement time

An ordering effect was found on movement time ($f_{5,11} = 5.18, p < .05$). A posthoc Tukey HSD test show that the ordering HP-LP-HFR was significantly faster than LP-HP-HFR ($p < .05$) and LP-HP-HFR ($p < .05$). It's important to note that there were only 3 data points for each ordering (2 for HFR-LP-HP due to the dropped participant), so any statistical results may not be conclusive.

There was a significant main effect of movement time for display method ($f_{2,22} = 60.03, p < .0001$). Pairwise comparisons show that HFR was significantly faster ($p < .0001$) than HP and LP, and no significant difference was found between HP and LP. There was also a significant effect of movement time for target size ($f_{1.35,14.82} = 801.1, p < .0001$). Pairwise comparisons showed the typical speed/accuracy tradeoff, with larger targets being significantly faster ($p < .0001$) than smaller ones. A significant main effect of movement direction on movement time was found as well ($f_{14,154} = 3.53, p < .0001$). Pairwise comparisons showed that the only significant difference in movement time was between movement directions 4 and 9 (figure 3), with 9 being significantly faster than 4 ($p < .05$). This single significant difference across all 15 movement directions suggests that it could be a result from an artifact of the experiment, rather than a true effect.

Significant interactions were also found for movement time. There was a significant interaction between ordering and display method ($f_{10,22} = 5.26, p < .001$). Looking at pairwise comparisons, we see that the only significant differences happened when HP was performed first, LP second and HFR last (HP-LP-HFR). In this case, HP was significantly slower than when it was performed second ($p < .05$) or last (with HFR being performed first) ($p < .05$). When HP was performed last and HFR was performed second, HP was marginally significantly faster than LP ($p = .07$). Another

significant interaction was observed between ordering and target size($f_{10,22} = 5.81, p < .0001$). Pairwise comparisons show that orderings were only different among each other for the small target size, where the ordering HP-LP-HFR was significantly slower than the orderings LP-HP-HFR ($p < .05$) and both orderings starting with HFR ($p < .05$). The last significant interaction effect found for movement time was between display method and target size ($f_{1.94,21.36} = 8.61, p < .005$). Looking at pairwise comparisons, we see, for each target size, the same effects of display mode as in the main effect. Thus, although a significant interaction effect occurred between target size and display method, all display methods maintained their differences when blocking by each target size. The interaction was expressed due to the variance in the amplitude of the differences between display methods for each target size, as can be seen in figure 6.
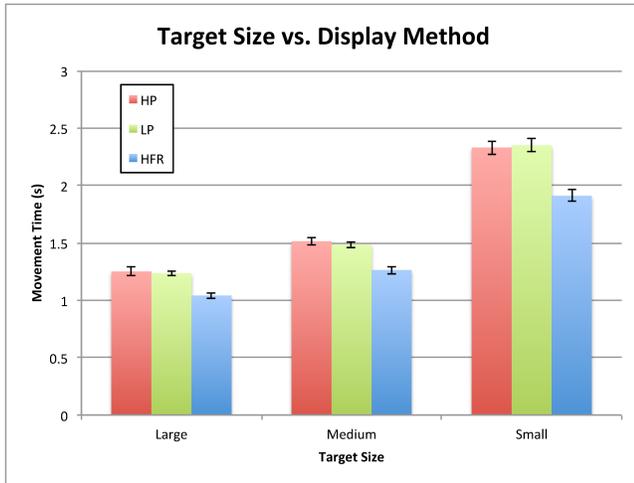


Figure 6: Interaction effect between target size and display method. Error bars represent standard error.

### 5.1.4 Errors

The analysis of the mean number of errors for the selection task found few significant effects. As expected, target size had a significant main effect on the mean number of errors ($f_{1.18,13.02} = 35.23, p < .0001$). Pairwise comparisons showed that the smaller the target, significantly more errors were made ($p < .0001$), which is consistent with the expected speed accuracy tradeoff of this kind of pointing task. Movement direction also had a significant main effect ($f_{14,154} = 2.32, p < .01$). However, pairwise comparisons did not show any significant difference among movement directions. This indicates that, although the overall model was significant for movement direction, the effect may not have been large enough to show differences when looking at each direction pair individually. There was no significant main effect of display method on the mean number of errors ($f_{2,22} = 1.77, p = .19$).

No significant interactions were found on the mean number of errors for selection.

## 5.2 Navigation Task

The navigation task involved one single path traversal per participant per condition. Thus, each participant performed a total of 9 navigation tasks.

### 5.2.1 Data Analysis

IBM SPSS 22 was used for analyzing data. A 2-way factorial analysis of variance (ANOVA) with repeated measures and one between-subjects factor was used on the following metrics: total time to traverse the path, total number of deviations from the path and total time spent off the path. The analysis, thus, had a $(3 \times 3)$ WTH×6 BTW design. Data was tested for sphericity, and whenever the assumption of equal variances was violated, the Greenhouse-Geisser correction on the degrees of freedom was used.

### 5.2.2 Traversal Time

No ordering effect was found for total path traversal time ($f_{5,11} = 2.33, p- = .11$).

There was a significant main effect of total path traversal time on display method ($f_{1.38,15.17} = 15.24, p < .001$). Pairwise comparisons showed that HFR was significantly faster than LP ($p < .05$) and than HP ($p < .005$). As expected, there was a strong significant main effect of path width on traversal time ($f_{1.38,15.14} = 46.64, p < .0001$). Pairwise comparisons showed that the narrowest path took significantly longer than both other paths ($p < .0001$), while there was a marginal difference showing that the medium path was slower than the widest one ($p = .085$).

There was a significant interaction between display method and ordering on traversal time ($f_{10,22} = 3.58, p < .01$). Pairwise comparisons showed that the only significant difference among orderings occurred when HP was used between the HP-LP-HFR and LP-HFR-HP, with HP being slower when it was performed first ($p < .05$). When fixing the display method for the same interaction effect, there was no significant difference across orderings. However, blocking each ordering, the differences across display methods varied. We could identify two ordering groups, which had similar results: HP-LP-HFR, LP-HP-HFR, LP-HFR-HP & HP-HFR-LP, HFR-HP-LP, HFR-LP-HP. In the first group, there were significant differences among the display methods, consistent with the main effect differences for display method, while in the second group, there were no significant differences at all. Due to the limited number of participants per ordering (3), it is risky to derive conclusions from these results, as these could have been caused by a participant selection artifact.

A significant 3-way interaction effect was found among display method, path width and ordering ($f_{20,44} = 2.09, p < .05$).

### 5.2.3 Path Deviations

No ordering effect was found for the total number of path deviations ($f_{5,11} = 1.87, p = .18$).

No significant main effect of total number of path deviations was found on display method ($f_{2,22} = 2.69, p = .09$). There was a significant main effect of path width on path deviations ($f_{2,22} = 192.12, p < .0001$). Naturally, pairwise comparisons showed that the the narrower the path, the more deviations were committed ($p < .05$).

As with traversal time, there was a significant interaction between display method and ordering on the total number of path deviations ($f_{10,22} = 4.81, p < .001$). Similar results from pairwise comparisons were also found as compared to traversal time.

### 5.2.4 Time Spent Off Path

No ordering effect was found for the total time spent off path ($f_{5,11} = 1.32, p- = .32$).

There was a significant main effect of total time spent off path on display method ($f_{2,22} = 3.82, p < .05$). Pairwise comparisons showed that HFR had significantly less time off path than HP ($p < .05$). There was also a significant main effect of path width on time off path ($f_{2,22} = 56.97, p < .0001$). Pairwise comparisons showed that the narrower paths had significantly more time off path than wider paths ($p < .001$).

As with both other measures, there was a significant interaction between display method and ordering on the total number of path deviations ($f_{10,22} = 3.58, p < .01$). Similar results from pairwise comparisons were also found.

## 5.3 Survey Analysis

Survey data on the display methods was analyzed using the non-parametric test Friedman's ANOVA, and questionnaires were tested for internal reliability using Cronbach's Alpha. All surveys were internally reliable with $\alpha > .75$.

For the presence score, we found a significant main effect ($\chi^2_f(2) = 10.53, p < .01$). Pairwise comparisons showed that HFR was rated significantly higher ($p < .05$) than both HP and LP display methods, which were not rated significantly different among each other ($p = 1.0$).

In terms of simulator sickness, we analyzed both the SSQ subscores for nausea, oculomotor fatigue and disorientation as well as the overall SSQ score. There were significant results for oculomotor fatigue ($\chi^2_f(2) = 8.87, p < .05$), with pairwise comparisons showing that LP was rated significantly higher than HFR. Overall SSQ yielded signifcant results ($\chi^2_f(2) = 6.53, p < .05$), but pairwise comparisons failed to yield significant differences. Looking at the individual nausea question HFR was rated marginally higher than HP (p=.084). Looking at individual scores, the only two participants who ranked their nausea as severe were in the HFR condition. Also, two participants who dropped out of the study from sickness, both had the HFR condition.

Participants rated their preference in terms of selection ($\chi^2_f(2) = 13.44, p < .005$), navigation ($\chi^2_f(2) = 10.78, p < .005$) and overall ($\chi^2_f(2) = 11.44, p < .005$). In all three cases, HFR was rated significantly highest, and there was no significant difference between HP and LP.

In terms of individual rankings, no participants ranked LP as their first choice. 4 participants (22%) ranked HFR last. 3 participants (17%) ranked LP ahead of HP.

## 6 DISCUSSION

Results from the study provided several interesting insights.

### 6.1 Task Performance

The task performance of the HFR case was expected and agrees with previous research [40]. It is interesting that there was no significant difference in task performance in HP vs LP. The selection task is a motor-based task, and thus, perhaps the feedback update (i.e., the input and display update rates) is the primary driver of performance. Also, during the selection task, the only object that was affected by HP artifacts (i.e., doubling of the lines) was the selection ray. We hypothesize that we may see more performance differences when looking at moving targets, that are always affected by HP artifacts. Future work should address different types of VR tasks, to find out if there are tasks that benefit from the LP condition.

In the navigation task, the following of the path on the ground was potentially problematic. We had some participants report that they looked more towards the ground then straight ahead. We feel that LP may help with stereo fusion, as when looking at distant objects in HP case, they become a jumble of lines. Future work should investigate navigation tasks where users were required to look forward. The PenguFly navigation technique assessment [38] used a ring gathering task, which may be suitable for this, as it keeps the user looking at objects ahead.

Another consideration is that brightness was not controlled for. HP and HFR cases were much brighter than the LP case (due to the black frames). While ecologically valid (i.e., decreased brightness is what happens in a real LP system), it would be interesting to see if maintaining equal brightness in all cases would change the results. We propose this as an additional area of inquiry for future work. Additionally, LP VEs could use a different color on the inserted frames as to match the overall brightness of HP and HFR cases.

### 6.2 Nausea

We found only marginally significant results when analysing the individual nausea question on the SSQ questionnaire (HFR marginaly

higher than HP (p=.084)). When we look at the actual scores, the only two participants who ranked their nausea as severe were in the HFR case. Previous researchers have hypothesized that increased optical flow [41] may be responsible for increased sickness. In another experiment, it was found that by increasing the FOV, both presence and simulator sickness increased [33]. In our case presence was indeed higher in the HFR case. While we were unable to discover true significance, this points toward an interesting picture that the higher fidelity, more immersive case, may actually promote more sickness in some individuals.

Two things are interesting about the nausea results. First, that HFR, which we would think of as a more desirable condition, was not the clear winner in terms of being lowest in simulator sickness. We expected the low frame rate HP and LP conditions to have larger simulator sickness than the HFR case. However, we must remember that we are utilizing a simulated low frame rate, that has less input latency then a real world low frame rate application. It could be that the increased input latency in real low frame rate virtual environments is a large factor in simulator sickness. Second, we didn't see any of the nausea reduction properties from the LP case as had previously been observed [31]. This may be related to the fact that in real world shutter testing, even during the short open periods of the glasses (often fixed at 100ms), the image was changing. That is, in our simulated case, the exposure is a fixed image, where as the real world shutter glasses in the open condition can see movement.

Another thing to consider is the research done on independent visual backgrounds (IVB) [10] also referred to recently as "rest frames" [6]. In these studies it was shown that providing a view of the real world (often a view of some part of the laboratory) or a graphical grid structure inside the VE, would give the user a stable frame of eference and reduce nausea. It is unclear in our experiment, if during the dark portion of the frame in the LP case, the user perceived the structure of the CAVE-type system itself as a reference frame. This would be interesting to explore in future work.

### 6.3 Oculomotor

We took care to construct a virtual environment that would best facilitate the LP condition. As previously discussed we utilized dark colors to more closely match the color of the black LP frames (figures 2 and 4). However, even with that care, users experienced eye strain and other oculomotor symptoms as reported on the SSQ questionnaires. So there is a cost to utilizing LP techniques at such low frame rates. In a real world application, where the scene is not so carefully constructed, there may be even more strain and fatigue.

### 6.4 Overall Preference

We find it interesting that 22 percent of subjects ranked HFR as their last choice. We conjecture that this may be due to several factors. First, it may be that the users were unable to tell the different between the HP and the HFR cases. The LP case (with its strobing visuals) is obviously different. Our other idea, is that it could be because of the increased nausea. Looking at the 4 subjects that ranked normal last, all of them had the worst nausea on the HFR condition.

Finally, we find it interesting that 3 participants (16.66%) rated LP above HP in the overall rankings. While small, this is still interesting, in that there is a subset of participants who prefer LP.

## 7 CONCLUSION AND FUTURE WORK

This paper presented an initial exploration on the effects of image persistence in low frame rate VEs. Several interesting insights surfaced from the presented study. The LP condition did not offer any performance advantages in the tasks of selection and navigation. Nor did it provide a statistically significant reduction in nausea. However, interesting trends were found that pointed towards HFR actually causing more simulator sickness. We also found that a small number of users did prefer to utilize the strobing LP condition over the HP condition. Further work should be conducted to

determine if and what VR tasks could benefit from low frame rate LP conditions.

## 8 ACKNOWLEDGEMENTS

## 9 DOWNLOADS

Our source code for the user study tasks can be downloaded from: http://github.com/djzielin/persistence-sim.

## REFERENCES

[1] F. Ababsa, M. Mallem, and D. Roussel. Comparison between particle filter approach and kalman filter-based technique for head tracking in augmented reality systems. In *Proc. ICRA 2004*, pp. 1021–1026. IEEE, 2004.

[2] L. Appelbaum, M. Cain, J. Schroeder, E. Darling, and S. Mitroff. Improving visual cognition through stroboscopic training. *J Vis*, 13(9):603–603, 2013.

[3] ARTTRACK5 Tracking System Technical Specifications. http://www.ar-tracking.com/products/tracking-systems/arttrack-system/arttrack5/. Accessed: 2014-09-08.

[4] W. Barfield and C. Hendrix. Presence as a function of update rate within virtual environments. Technical Report 951569, SAE Technical Paper, 1995.

[5] D. Carmel, N. Lavie, and G. Rees. Conscious awareness of flicker in humans involves frontal and parietal cortex. *Curr Biol*, 16:907–911, 2006.

[6] E. Chang, I. Hwang, H. Jeon, Y. Chun, H. T. Kim, and C. Park. Effects of rest frames on cybersickness and oscillatory brain activity. In *Proc. BCI 2013*, pp. 62–64, Feb 2013.

[7] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proc. SIGGRAPH 1993*, pp. 135–142. ACM, 1993.

[8] I. W. Davis, L. W. Murray, J. S. Richardson, and D. C. Richardson. Molprobity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res*, 32(suppl. 2):W615–W619, 2004.

[9] S. A. Douglas, A. E. Kirkpatrick, and I. S. MacKenzie. Testing pointing device performance and user assessment with the iso 9241, part 9 standard. In *Proc. CHI '99*, pp. 215–222, 1999. ACM.

[10] H. B.-L. Duh, D. E. Parker, and T. A. Furness. An independent visual background reduced simulator sickness in a driving simulator. *Presence-Teleop Virt*, 13(5):578–588, 2004.

[11] S. R. Ellis, F. Breant, B. Manges, R. Jacoby, and B. D. Adelstein. Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. In *Proc. VR 1997,*, pp. 138–145. IEEE, 1997.

[12] T. Grossman and R. Balakrishnan. A probabilistic approach to modeling two-dimensional pointing. *ACM Trans Comput-Hum Interact*, 12(3):435–459, 2005.

[13] W. Guan, S. You, and U. Neumann. Gps-aided recognition-based user tracking system with augmented reality in extreme large-scale areas. In *Proc. MMSys '11*, pp. 1–10, New York, NY, USA, 2011. ACM.

[14] W. Hershberger and J. Jordan. The phantom array: A perisaccadic illusion of visual direction. *Psychol Rec*, 48:21–32, 1998.

[15] E. Hodgson, E. Bachmann, D. Waller, A. Bair, and A. Oberlin. Virtual reality in the wild: A self-contained and wearable simulation system. In *Proc. VR 2012*, pp. 157–158, IEEE, 2012.

[16] ISO/TS 9241-411, Ergonomics of human-system interaction - Part 411: Evaluation methods for the design of physical input devices, 2012.

[17] D. Kelly, R. Boynton, and W. Baron. Primate flicker sensitivity: Psychophysics and electrophysiology. *Science*, 194(4269):1077–1079, 1976.

[18] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int J Aviat Psychol*, 3(3):203–220, 1993.

[19] C.-Y. Kim and R. Blake. Psychophysical magic: Rendering the visible 'invisible'. *Trends in Cog Science*, 9(8):381–388, 2005.

[20] T. Kim, B. Park, B. Shin, B. H. Berkeley, and S. S. Kim. 60.1: Response time compensation for black frame insertion. In *Proc. SID 2006*, volume 37, pp. 1793–1796. Wiley, 2006.

[21] E. M. Kolasinski. Simulator sickness in virtual environments. Tech. Report ARI-TR-1027, DTIC Document, ALEXANDRIA, VA, 1995.

[22] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *Int J of Hum-Comput St*, 68(10):603–615, 2010.

[23] S. L. Macknik and M. S. Livingstone. Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neurosci*, 1(2):144–149, 1998.

[24] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr. Physiological measures of presence in stressful virtual environments. *ACM Trans Graph*, 21(3):645–652, July 2002.

[25] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks, Jr. Effect of latency on presence in stressful virtual environments. In *Proc. VR '03*, pp. 141–, 2003. IEEE.

[26] M. R. Mine. Virtual environment interaction techniques. Technical report, UNC Chapel Hill, Chapel Hill, NC, USA, 1995.

[27] T. Ni, G. S. Schmidt, O. G. Staadt, M. A. Livingston, R. Ball, and R. May. A survey of large high-resolution display technologies, techniques, and applications. In *Proc. VR 2006*, pp. 223–236. IEEE, 2006.

[28] Oculus Rift DK2 Specifications. http://www.oculusvr.com/dk2/. Accessed: 2014-09-08.

[29] T. Oskiper, S. Samarasekera, and R. Kumar. Multi-sensor navigation algorithm using monocular camera, imu and gps for large scale augmented reality. In *Pro. ISMAR, 2012*, pp. 71–80, IEEE, 2012.

[30] D. Pitman and M. L. Cummings. Collaborative exploration with a micro aerial vehicle: a novel interaction method for controlling a mav with a hand-held device. *Adv Hum-Compu Int*, 2012:18. Hindawi, 2012.

[31] M. F. Reschke, J. T. Somers, and G. Ford. Stroboscopic vision as a treatment for motion sickness: strobe lighting vs. shutter glasses. *Aviat Space Env Med*, 77(1):2–7, 2006.

[32] B. Schaeffer and C. Goudeseune. Syzygy: native pc cluster vr. In *Proc. VR 2003*, pp. 15–22. IEEE, 2003.

[33] A. F. Seay, D. M. Krum, L. Hodges, and W. Ribarsky. Simulator sickness and presence in a high fov virtual environment. In *Proc. VR 2001*, pp. 299–300. IEEE, 2001.

[34] E. Simonson and J. Brozek. Flicker fusion frequency. *Physiol Review*, 32:349–378, 1952.

[35] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence-Teleop Virt*, 3(2):130–144, 1994.

[36] F. Smit, R. Van Liere, and B. Froehlich. A programmable display layer for virtual reality system architectures. *IEEE Trans Vis Comput Graphics*, 16(1):28–42, Jan 2010.

[37] C. Stinson, R. Kopper, S. Scerbo, E. Ragan, , and D. Bowman. The effects of visual realism on training transfer in immersive virtual environments. In *Human Systems Integration Symposium*, 2011.

[38] A. von Kapri, T. Rick, and S. Feiner. Comparing steering-based travel techniques for search tasks in a cave. In *Proc. VR 2011*, pp. 91–94. IEEE, 2011.

[39] D. Wang, A. Vincent, P. Blanchfield, and R. Klepko. Motion-compensated frame rate up-conversion – part ii: New algorithms for frame interpolation. *IEEE Trans Broadcast*, 56(2):142–149, 2010.

[40] C. Ware and R. Balakrishnan. Reaching for objects in vr displays: lag and frame rate. *ACM Trans Comput-Hum Interact*, 1(4):331–356, 1994.

[41] Z. Yin and R. R. Mourant. The perception of optical flow in driving simulators. In *Proc. The 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Big Sky, Montana*, pp. 176–182, 2009.

[42] S. Zhai and R. Woltjer. Human movement performance in relation to path constraint - the law of steering in locomotion. In *Proc. VR 2003*, pp. 149–156, IEEE, 2003.

[43] X. Zhang and I. MacKenzie. Evaluating eye tracking with iso 9241 - part 9. In J. Jacko, editor, *HCI Intelligent Multimodal Interaction Environments*, volume 4552 of *LNCS* , pp. 779–788. Springer Berlin Heidelberg, 2007.

[44] D. J. Zielinski, R. Kopper, R. P. McMahan, W. Lu, and S. Ferrari. Intercept tags: enhancing intercept-based systems. In *Proc. VRST 2013*, pp. 263–266. ACM, 2013.