Avatar Self-Embodiment Enhances Distance Perception Accuracy in Non-Photorealistic Immersive Virtual Environments

Lane Phillips, Brian Ries, Michael Kaeding and Victoria Interrante

Department of Computer Science, University of Minnesota

ABSTRACT

Non-photorealistically rendered (NPR) immersive virtual environments (IVEs) can facilitate the process of conceptual design in architecture by enabling preliminary design sketches to be previewed and experienced at full scale, from a first-person perspective. However, to this end it is critical to ensure the accurate spatial perception of the represented information, and many studies have shown that people typically underestimate distances in most IVEs, regardless of rendering style. In previous work we have found that while people tend to judge distances more accurately in an IVE that is a high-fidelity replica of their concurrently occupied real environment than in an IVE that it is a photorealistic representation of a real place that they've never been to, significant distance estimation errors re-emerge when the replica environment is represented in a non-photorealistic style. We have also previously found that distance estimation accuracy can be improved, in photo-realistically rendered novel virtual environments, when people are given a fully tracked, high fidelity first person avatar self-embodiment. In this paper we report the results of an experiment that seeks to determine whether providing users with a high-fidelity avatar self-embodiment in a non-photorealistically rendered virtual replica environment will enable them to perceive the 3D spatial layout of that environment more accurately. We find that users who are given a first person avatar in an NPR replica environment judge distances more accurately than do users who experience the NPR replica room without an embodiment, but not as accurately as users whose distance judgments are made in a photorealistically rendered virtual replica room. Our results provide a partial solution to the problem of facilitating accurate distance perception in NPR virtual environments, while supporting and expanding the scope of previous findings that giving people a realistic avatar self-embodiment in an IVE can help them to interpret what they see through an HMD in a way that is more similar to how they would interpret a corresponding visual stimulus in the real world.

CR Categories and Subject Descriptors: I.3.6 [Computer Graphics]: Methodology and Techniques; I.3.7 [Computer Graphics]: 3D Graphics and Realism – *virtual reality*.

Additional Keywords: spatial perception, immersive virtual environments, non-photorealistic rendering, first-person avatars.

1. INTRODUCTION

Immersive virtual environments can play a valuable role in the architectural design process by enabling people to preview designs from a first-person perspective before they are built [1, 3]. Although most often used to provide photorealistic walkthroughs of finished models, we believe that there is tremendous potential for greater use of virtual environments technology earlier in the design process, to facilitate the exploration of alternative conceptual approaches to the basic layout of the designed space.

Non-photorealistic rendering provides a promising solution to the problem of how to represent basic design ideas. Architects typically rely on rough 2D sketches at the beginning of the design process because the details necessary to produce a finished drawing have not yet been determined, and research has shown that sketch-style drawings are particularly effective at conveying the unfinished character of a design and inviting the consideration of changes [11].

A crucial goal in using immersive virtual environments technology is to enable people to experience the 3D space of a design from a first person perspective, so that they can naturally and intuitively interpret its spatial dimensions and layout. However studies have found that people typically do not judge egocentric distances in virtual environments as accurately as they do in the real world, and much effort has been devoted to investigating this problem and exploring potential solutions [6].

In this paper we address the question of how to facilitate accurate spatial perception in a non-photorealistic immersive virtual environment. In particular, we report the results of an experiment in which we test the impact of providing people with a first person avatar self-representation in the NPR IVE.

2. PREVIOUS WORK

Non-photorealistic rendering has long been advocated as a useful device for conveying a sense of uncertainty about an architectural design and for indicating that a given 3D model is preliminary and open to change [11]. Klein *et al*. [5] were the first to demonstrate a system for rendering virtual environments in a non-photorealistic style. Their method achieved a painterly representation through the use of dynamic stroke-based textures.

Gooch and Willemsen [2] were the first to investigate the accuracy of distance perception in non-photorealistically rendered IVEs presented via head mounted displays (HMDs). Their initial study considered a line-drawing representation, and they found that people underestimated distances in the NPR IVE. In subsequent work, they and others compared distance perception in the line-drawing style IVE with distance perception in a non-photorealistically textured virtual environment and in a simulated virtual environment created from panoramic photos, and found that participants underestimated distances in all three of these environments to a nearly equivalent extent [12].

Our group discovered that people do *not* tend to significantly underestimate distances in IVEs when the virtual environment is a high fidelity replica of the same physical environment that the observer knows that s/he is concurrently occupying [4]. However, we found that significant errors in distance estimation *do* occur when this replica environment is rendered in a non-photorealistic style [7]. Several possible explanations can be considered. On the one hand, participants could be mis-judging distances when the replica environment is rendered as a line-drawing because of a lack of sufficient detail in the line drawing representation. The statistics of the images available to the viewer in this case are very different from the statistics of natural images, and it is possible that the impoverished NPR representation lacks critical cues that our visual system relies upon when making spatial judgments. However it is also possible that participants are underestimating distances in the NPR IVE because its lack of realism interferes with their ability to feel *present* in the virtual environment, and that this lack of presence in turn affects their interpretation of the visual input provided by the head mounted display.

Mel Slater and others have long advocated providing participants in a virtual environment with a virtual body in order to enhance their sense of presence [10]. In recent work [8], our group has found that people who are given a high fidelity avatar self-embodiment in a novel, highly realistic virtual environment are able to judge distances in that environment significantly more accurately than are people who are immersed in the same VE without an embodiment. However, follow up studies found that this performance boost was compromised when either an impoverished dot avatar representation, or single point tracking, as opposed to full-body tracking, was used [9].

If people are failing to accurately judge distances in our non-photorealistically rendered virtual replica room because of a lack of detail in the rendering, then one possible solution would be to use a different style of non-photorealistic rendering to represent our models. If, however, the decrease in performance is due to a lack of presence in the NPR IVE, then we might be able to improve the situation by providing people with a high fidelity avatar self-embodiment. The study we report here seeks to evaluate the potential of taking this latter approach.

3. OUR EXPERIMENT

The goal of our experiment was to determine whether participants would be able to judge egocentric distances more accurately in an NPR immersive virtual environment, presented via a head-mounted display, when they are immersed in that environment with a high fidelity avatar self-embodiment, as opposed to experiencing the environment from the default perspective of a dis-embodied viewer.

3.1. Method

We used direct 'blind walking' to assess the accuracy with which participants judged distances in our non-photorealistically rendered virtual replica room and in the real world. We used a mixed within- and between-subjects design, in which half of the participants experienced the virtual environment with an avatar embodiment and half did not, and all participants made judgments in both the virtual environment and in the real world as a control.

3.1.1. Participants

Twelve new participants (11 male, 1 female) took part in this experiment. Ten (9 male, 1 female) experienced the NPR VE with an avatar self-embodiment and the other two experienced the without-avatar condition; the latter's results were combined with the results of the 8 participants in our earlier experiment [7] run under identical conditions, to yield a total of 20 results for our present analysis. All participants were University of Minnesota students, ranging in age from 18 to 33 years old. They were recruited via either a hand-held sign at a high-traffic sidewalk, a posted sign on our lab door, or an announcement in one of the authors' large undergraduate classes. Each participant was compensated with a $10 gift certificate to a national retail chain.

3.1.2. Apparatus

The experiment was conducted in our Digital Design Consortium Laboratory, consisting of a large rectangular room with a large, curved screen on one side. This space was 30′ long, and 25′ wide at its widest point, narrowing to a width of 16.5′ at each end.

The virtual environment was modeled in SketchUp as an exact 3D replica of the real room, and was rendered with non-photorealistic textures derived by hand-tracing over the major edges in photographs of the original space. Our goal was to emulate the style of the non-photorealistic environment used in [2, 12], although the contours in that case were derived from the model geometry rather than from textures. Figure 1 shows the high-fidelity version of the replica virtual environment used in our earlier experiments and figure 2 shows the non-photorealistic version of that environment used in this study. The model was rendered using OpenGL in the without-avatar condition and using a modified version of the OGRE API in the with-avatar condition, as the latter system provided useful support for the avatar animation. The appearance of the virtual environment was identical in both cases and we did not notice any differences in rendering speed between the two conditions. Rendering was done on a desktop computer with a 2.42 GHz AMD Athlon 64 X2 dual core processor 4800+ with 2.0GB of RAM and an nVidia Quadro FX 4500 video card with 512 MB of graphics RAM.

Figure 1: A screen shot of the photorealistic virtual environment used in our prior experiments, showing the photographic textures that were drawn over to define the line-drawing textures used in the NPR rendering.

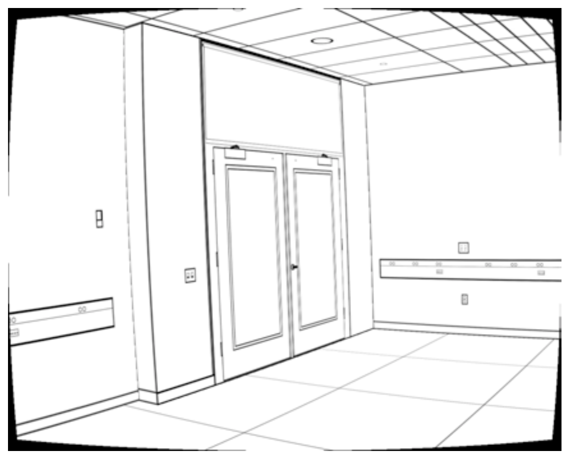

Figure 2: A screen shot of the non-photorealistic virtual environment used in this experiment.

The virtual environment was presented using an nVisor SX head mounted display, which has two 1280 x 1024 screens (one for each eye) that provide 100% stereo overlap across a manufacturer-specified 60° monocular field of view. Foam blinders attached to the HMD blocked participants' view of the external environment. The HMD weighed about 1 kg and was connected by a 15′ cable to a video control unit that was strapped to a wheeled cart. An experimenter managed the cables to prevent them from getting in the participant's way or from weighing down behind their head. Real time viewpoint tracking was provided by a HiBall 3000 optical ceiling tracker in the without-avatar condition, and, in the with-avatar condition, by a 12-camera Vicon MX40 motion capture system tracking six retro-reflective markers attached to the HMD. A portable radio tuned to static was used to mask auditory cues to participants' locations.

Participants in the with-avatar condition wore a black body suit over their clothes, to which we attached 41 retro-reflective markers to enable full body tracking. Additional markers were attached to participants' hands and feet. The avatar used in the experiment was based on a generic texture-mapped mesh from TurboSquid. All participants were embodied using the same mesh, which was manually skinned to the default Vicon skeleton in an offline pre-process. The skeleton was then custom-fitted to each participant via a calibration procedure, and the avatar mesh was scaled to fit. Particular care was taken to ensure that the avatar width was appropriate to the user, and that the lengths of the avatar's feet were scaled to match the participant's.

3.1.3. Procedure

Each participant individually entered our lab and read written instructions describing the experiment. The ten participants in the with-avatar condition went through an additional pre-process in which they put on the motion capture suit and performed a range-of-motion exercise to calibrate the Vicon system. All participants were assisted in placing the HMD on their heads and adjusting it to fit.

Each participant performed 20 trials of blind walking in the virtual environment, followed by 10 trials of blind walking in the real room as a control*. At the start of each trial, participants in the with-avatar condition were shown one virtual tape mark on the ground immediately in front of their virtual feet and another at a random distance 8–20′ in front of them. They were instructed to look down and align their toes with the starting mark, and then take visual aim at the ending mark. This procedure provided a subtle means of ensuring that all of the participants who had an avatar saw their avatar at the start of each trial. Figure 3 shows what this looked like. Participants in the without-avatar condition were shown an ending tape mark only, as they had no visible feet to align with a starting tape mark, and we did not want to artificially draw their attention to their lack of an embodiment. The participants in each condition were only made aware of the protocol for their condition and were unaware of what other conditions were being tested. When participants felt confident that they knew the location of the ending tape mark, they would close their eyes, say 'ready', and then walk with their eyes closed to where they thought the ending tape mark was. At the 'ready' signal, the experimenter would press a key to record the participant's starting position and blank the display. When participants arrived at where they thought the target mark was, they stopped and said 'done', and the experimenter pressed a different key to record their ending position while leaving the display blacked out. With the display still blank, the participant was directed to walk (with their eyes still shut) to a new position to start the next trial.

After completing the virtual world trials, participants performed ten similar trials in the real room as an individual control. We used each participant's performance in the real world to calibrate their performance in the virtual world. This enabled us to control for individual variation in each participant's baseline distance estimation accuracy using blind walking*. Participants removed the HMD and the motion capture suit before conducting the real world trials, but they continued to wear the radio playing static. The basic procedure was the same as in the virtual world, except that we used arbitrarily placed strips of cloth to mark the starting and ending points for each trial. Participants in the without-avatar condition did not wear shoes during the virtual-world trials (so that we could track their feet), but were required to wear their shoes during the real-world trials, so that they could not feel the cloth strips. Participants also wore a blindfold during the real world trials to ensure that they were not able to peek at the floor any time during or after a trial. The distances between tape marks and the distances walked were measured in the real world with a tape measure.

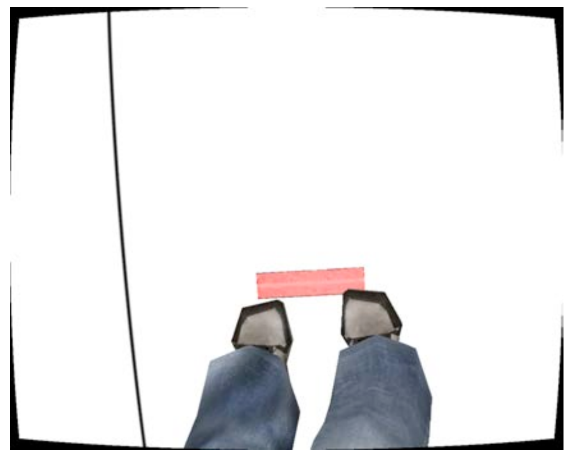

Figure 3: A screen shot showing what a participant in the with-avatar condition might see at the beginning of a trial.

3.2. Results

Figures 4 and 5 show the results of the 20 total participants in the without- and with- avatar conditions respectively. The results of 8 of the participants in the without-avatar condition were previously presented in an earlier paper [7], but are shown here again for comparison, augmented with results from two additional participants to achieve balance in the number of participants in each condition.

In each case, participants overall significantly underestimated distances in the non-photorealistically rendered virtual world relative to in the real world {$F(1,18) = 27.026$, $p < 0.001$, in the no-avatar condition and $F(1,18) = 16.608$, $p < 0.001$ in the with-avatar condition}. It is important to note that while one or more of the individual participants in each case did *not* perform significantly differently in the virtual world than in the real world, what we are measuring here is how the participants performed overall, as a group. This indicates that giving people an avatar

* In a previous unpublished study specifically looking at order effects, we found no difference between participants who made distance judgments in our high fidelity virtual replica room before making distance judgments in the real room and those who performed the real world trials first. However we did find significant order effects in another earlier unpublished study looking at the effects of texture degradation on distance judgment accuracy. In that study, participants who performed real world trials first, followed by trials in a high fidelity version of the VE, and then in a detail-poor VE, performed significantly more accurately in the degraded VE than did participants who had to make distance judgments in the degraded VE first, possibly because they were using insights gathered from their experience in one or both of the detail-rich environments to inform their subsequent judgments in the texturally impoverished version. As we were only using the real world judgments as a calibration control in our present study, we wanted to ensure that they were done after the VE trials to avoid inadvertently biasing our VE findings.

* While many studies have found that, on average, people tend to accurately judge distances in the real world using blind walking, on an individual basis some people will consistently walk short and some will consistently walk long. By using each person as their own control we are able to concentrate our analysis on the performance *difference*, for each individual, in the real and virtual worlds, thus factoring out these individual differences from our analysis.

self-embodiment in a non-photorealistically rendered virtual environment does *not* enable them to perceive distance in that environment as accurately as they can perceive distances in the real world.

However, further analysis of the difference in the virtual-minus-real world errors between the with- and without-avatar conditions reveals that the participants who were given an avatar self-embodiment in the NPR IVE performed significantly more accurately in the virtual environment than in the real world, as a group, than did the participants who experienced the NPR IVE without an embodiment {$F(1,18) = 5.118$, $p = 0.036$}. The average difference between the relative errors in participants' virtual world and real world distance judgments was -14.70% for the participants who experienced the NPR IVE with a first-person avatar self-embodiment and -25.97% for the participants who experienced the NPR IVE in the default, dis-embodied manner. We can conclude from this that while providing people with a first-person avatar does not solve the problem of distance underestimation in non-photorealistically rendered virtual environments, it does help.

3.3. Discussion

Although it appears clear that providing people with a first-person avatar in a non-photorealistic immersive virtual environment enables them to perceive distances in that environment more accurately than if they were not embodied, the explanation for this result remains open to interpretation. On the one hand, it is possible that providing people with an avatar self-embodiment enabled them to feel more 'present' in the virtual environment, which in turn encouraged them to act upon what they saw through the HMD in a way that was more similar to how they would act in the real world, despite the 'unreality' of what they saw, than participants who experienced the NPR IVE from a disembodied point of view that evoked greater uncertainty in how to interpret the unfamiliar visual input. But other explanations cannot be definitively ruled out. For instance, it is possible, but unlikely in light of the findings of [9], that participants were able to use the familiar size cues provided by the high fidelity virtual body representation as a referent for calibrating their perception of scale in the NPR IVE. It is also possible that some other, unaccounted for difference in the protocol between the with- and without-avatar conditions is responsible for the performance difference, although we feel that this is highly unlikely in the absence of any clear theoretical basis for concern.

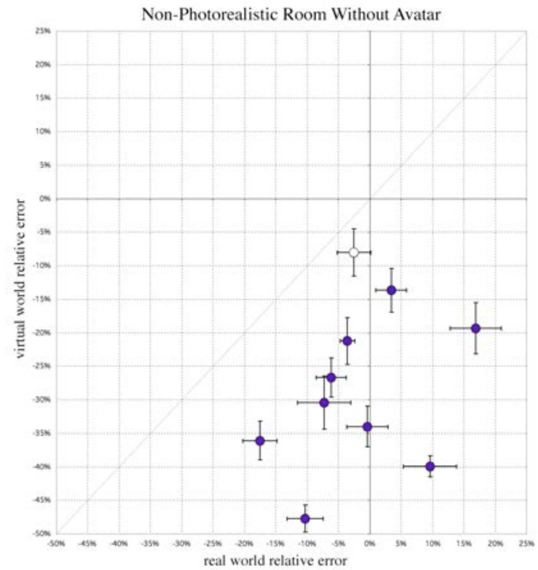

Figure 4: A point plot showing the results for the ten participants who did not have an avatar self-embodiment in the VE. Filled circles indicate results that were significantly different in the real and virtual worlds. Error bars bound the 95% confidence intervals.

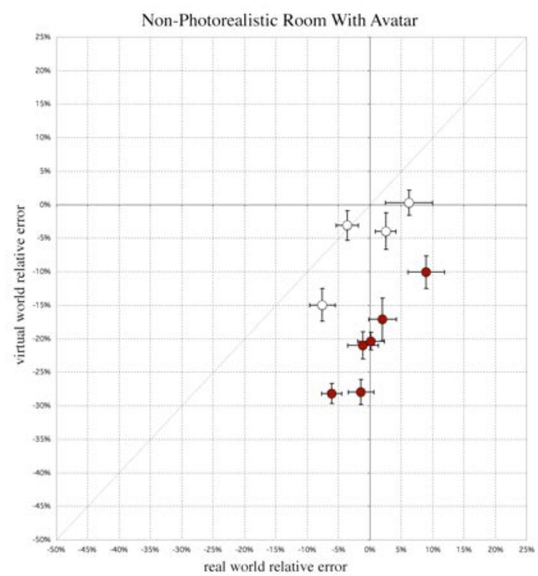

Figure 5: A point plot showing the results for the ten participants who were given a first-person avatar self-embodiment in the VE.

4. Future Work

Our present study focused on the use of a high fidelity avatar in light of the findings of [9] which showed this avatar representation to enable significantly improved performance over two other more simplified types of embodiment. However we think it would be interesting, in future work, to investigate the benefits of providing people immersed in a non-photorealistically rendered virtual environment with a non-photorealistically rendered avatar representation, for greater consistency with the NPR theme. Rendering everything seen with the same representational style might offer a stronger basis for the plausible interpretation of the unfamiliar rendering style as being a consequence of a global 'filter' applied equally to all input, facilitating the effort of looking past the differences and interpreting the essential signal conveyed in the NPR VE similarly as in the real world.

In future work we are also interested in exploring the use of alternative NPR representational styles. We remain concerned that the particular rendering style chosen for this study produces images whose statistics are very different from the statistics of natural images, and we foresee some potential benefit in using a non-photorealistic representation that involves more natural variation in the spatial frequencies in the input signal.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation through grants IIS-0313226 and IIS-0713587, and by the University of Minnesota through the Digital Technology Center and the Linda and Ted Johnson Digital Design Consortium Endowment and Lab Setup Funds.

REFERENCES

[1] ANDERSON, L., ESSER, J. AND INTERRANTE, V. 2003. A virtual environment for conceptual design in architecture. In *Proc. IPT-EGVE 2003*, Zurich, Switzerland, 57–63.

[2] GOOCH, A. A., AND WILLEMSEN, P. 2002. Evaluating Space Perception in NPR Immersive Environments. In *Proc. 2nd Int'l Symp. Non-Photorealistic Animation and Rendering,* 105-110.

[3] HENRY, D. AND FURNESS, T. 1993. Spatial perception in virtual environments. In *Proc. VRAIS 93,* 33–40.

[4] INTERRANTE, V., ANDERSON, L. AND RIES, B. 2006a. Distance perception in immersive virtual environments, revisited. In *Proc. IEEE Virtual Reality 2006*, 3–10.

[5] KLEIN, A., LI, W., KAZHDAN, M., CORRÊA, W., FINKELSTEIN, A., AND FUNKHOUSER, T. 2000. Non-Photorealistic Virtual Environments. In *Proc. SIGGRAPH 2000*, 527–534.

[6] LOOMIS, J. M., AND KNAPP, J. M. 2003. Visual perception of egocentric distance in real and virtual environments. In Virtual and Adaptive Environments, L. J. Hettinger & M. W. Haas (Eds.), 21–46.

[7] PHILLIPS, L., RIES, B., INTERRANTE, V., KAEDING, M., AND ANDERSON, L. 2009. Distance Perception in NPR Immersive Virtual Environments, Revisited. In *Proc. APGV* 09, 11–14.

[8] RIES, B., INTERRANTE, V., KAEDING, M., AND ANDERSON, L. 2008. The effect of self-embodiment on distance perception in immersive virtual environments. In *Proc. VRST*, 167–170.

[9] RIES, B., INTERRANTE, V., KAEDING, M., AND PHILLIPS, L. 2009. Analyzing the Effect of a Virtual Avatar's Geometric and Motion Fidelity on Ego-Centric Spatial Perception in Immersive Virtual Environments. In *Proc. VRST 09*, 59–66.

[10] SLATER, M., AND USOH, M. 1994. Body centred interaction in immersive virtual environments. In *Artificial Life and Virtual Reality*, N. Magnenat-Thalmann and D. Thalmann (Eds.), 125–148.

[11] STROTHOTTE, T., PREIM, B., RAAB, A., SCHUMANN, J., AND FORSEY, D. 1994. How to render frames and influence people. *Computer Graphics Forum*, 13, 3, 455–466.

[12] THOMPSON, W. B., WILLEMSEN, P., GOOCH, A. A., CREEM-REGEHR, S. H., LOOMIS, J. M., AND BEALL, A. C. 2004. Does the quality of the computer graphics matter when judging distances in visually immersive environments. *Presence: Teleoperators and Virtual Environments,* 13, 5, 560–571.