Impact of Visual and Experiential Realism on Distance Perception in VR using a Custom Video See-Through System

Koorosh Vaziri
Dept. of Computer Science
University of Minnesota
umn@kvaziri.com

Peng Liu
Dept. of Computer Science
University of Minnesota
peng.liu916@gmail.com

Sahar Aseeri
Dept. of Computer Science
University of Minnesota
aseer002@umn.edu

Victoria Interrante
Dept. of Computer Science
University of Minnesota
interran@umn.edu

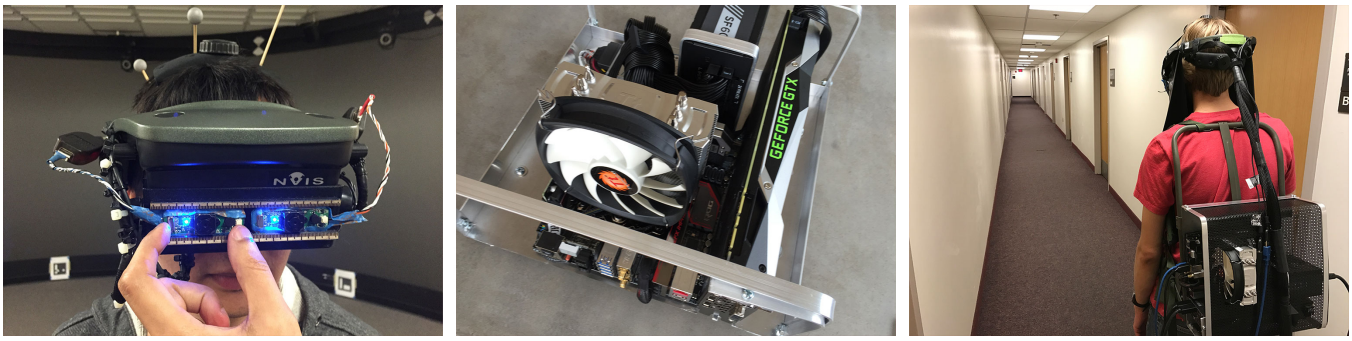

Figure 1: Left: 3D-Printed visor with adjustable cameras. Middle: Custom backpack computer. Right: Final system in use.

ABSTRACT

Immersive virtual reality (VR) technology has the potential to play an important role in the conceptual design process in architecture, if we can ensure that sketch-like structures are able to afford an accurate egocentric appreciation of the scale of the interior space of a preliminary building model. Historically, it has been found that people tend to perceive egocentric distances in head-mounted display (HMD) based virtual environments as being shorter than equivalent distances in the real world. Previous research has shown that in such cases, reducing the quality of the computer graphics does not make the situation significantly worse. However, other research has found that breaking the illusion of reality in a compellingly photorealistic VR experience can have a significant negative impact on distance perception accuracy.

In this paper, we investigate the impact of "graphical realism" on distance perception accuracy in VR from a novel perspective. Rather than starting with a virtual 3D model and varying its surface texture, we start with a live view of the real world, presented through a custom-designed video/optical-see-through HMD, and apply image processing to the video stream to remove details. This approach offers the potential to explore the relationship between visual and experiential realism in a more nuanced manner than has previously been done. In a within-subjects experiment across three different real-world hallway environments, we asked people to perform blind walking to make distance estimates under three different viewing conditions: real-world view through the HMD; closely registered camera views presented via the HMD; and Sobel-filtered versions of the camera views, resulting a sketch-like (NPR) appearance. We found: 1) significant amounts of distance underestimation in all three conditions, most likely due to the heavy backpack computer that participants wore to power the HMD and cameras/graphics; 2) a small but statistically significant difference in the amount of underestimation between the real world and camera/NPR viewing conditions, but no significant difference between the camera and NPR conditions. There was no significant difference between participants' ratings of visual and experiential realism in the real world and camera conditions, but in the NPR condition participants' ratings of experiential realism were significantly higher than their ratings of visual realism. These results confirm the notion that experiential realism is only partially dependent on visual realism, and that degradation of visual realism, independently of experiential realism, does not significantly impact distance perception accuracy in VR.

SAP '17, September 16-17, 2017, Cottbus, Germany

ACM ISBN 978-1-4503-5148-5/17/09…$15.00
https://doi.org/.1145/3119881.3119892

CCS CONCEPTS

• **Computing methodologies** → **Virtual reality**; *Perception*;

KEYWORDS

Virtual reality, spatial perception, non-photorealistic rendering

ACM Reference format:
Koorosh Vaziri, Peng Liu, Sahar Aseeri, and Victoria Interrante. 2017. Impact of Visual and Experiential Realism on Distance Perception in VR using a Custom Video See-Through System. In *Proceedings of SAP '17, Cottbus, Germany, September 16-17, 2017,* 8 pages.
https://doi.org/.1145/3119881.3119892

1 INTRODUCTION

The initial stage in the architectural design process is variously referred to as the conceptual, outline, or schematic design phase, in which various possibilities for the rough form and bulk of a building's structure are explored. Conceptual designs do not include photorealistic details because such details take time to model, and can distract from the more basic considerations of essential form and structure that are the primary focus at this point in the design process. Typically, conceptual designs are developed and communicated either as 2D sketches on sheets of paper or, possibly, as 3D wireframe models on a computer monitor.

We have observed, however, that these traditional modes of presentation unconsciously privilege attention to the external form of a building over attention to the occupants' experience of its interior space. Clients, in particular, may find it difficult to weigh alternative design directions on the basis of the lived experience that each can afford. As a result, priority may be given to exterior design features that later compromise interior comfort.

Immersive design systems have the potential not only to allow architects to explore design ideas from a first-person perspective, but also to enable clients to experience these outlined spaces firsthand, so that they can provide more informed feedback during the earliest stages of the design process, when changes are easier and less expensive to make.

Non-photorealistic rendering (NPR) techniques are ideal for displaying conceptual building models. Classical research in human-computer interaction has shown that presenting a wireframe building model in a sketch-like style conveys a greater sense of flexibility in the design that stimulates deeper engagement with the design ideas and more openly invites modifications [Schumann et al. 1996].

A fundamental concern with using VR to preview conceptual designs, however, is the need to ensure that people will be able to successfully derive an accurate appreciation of the volumetric extents of the roughly designed 3D spaces from their immersive experience in the sketchily-rendered 3D models. Historically, few reported VR implementations have successfully afforded accurate egocentric distance judgments [Renner et al. 2013]. Although recent research has found some evidence of improved distance estimation performance with newer HMDs [e.g. Young et al. 2014], we are unaware of any non-photorealistic virtual reality scenario in which significant distance underestimation has not been observed [Peer and Ponto 2017].

The research we present in this paper aims to contribute new insights into this classical problem by addressing it from a novel perspective. Specifically, we seek to better understand the potential for enabling accurate spatial perception in non-photorealistically rendered immersive virtual environments by considering the question of spatial perception accuracy within the context of a higher level of experiential realism than has been previously studied in VR-based experiments.

In particular, rather than varying the textural details of a virtual 3D model to assess the impact of rendering quality in VR, we use a custom-built, dual-camera attachment to an optical-see-through HMD to provide people with live visual input at three different levels of visual realism, in three different real world hallway environments. The display conditions were: (1) a direct view of the real world (seen with the visor removed and HMD turned off); (2) a dual-video view of the real world (obtained from two tiny cameras placed immediately in front of the person's eyes and displayed in the HMD); and (3) a line-drawing-style version of reality obtained by applying real time image-processing filters to the live video feed. We then compared participants' average distance perception accuracy across these three different display conditions, and analyzed those results in the context of peoples' subjective ratings of the visual and functional realism of each immersive experience.

2 BACKGROUND AND RELATED WORK

2.1 Perception

Much prior effort has been devoted to elucidating the root causes of distance underestimation in VR, e.g. [Interrante et al. 2006; Willemsen et al. 2009], and exploring potential solutions or workarounds [Kuhl et al. 2009; Mohler et al. 2006; Ries et al. 2008].

Willemsen and Gooch [2002] were the first to directly investigate the impact of the quality of the computer graphics rendering on distance perception accuracy in VR. They asked 12 participants to indicate perceived egocentric distances via blind walking from a fixed home base to targets on the floor at various distances away in three differently-rendered versions of the same hallway, with the viewing conditions presented in randomized order. The three stimulus conditions were: (1) unmediated real world viewing; (2) a view of a panoramic stereo photograph, previously obtained from a height-matched vantage point, displayed on an HMD; and (3) a highly detailed 3D computer graphics rendering, also displayed on the HMD. They found that accuracy was nearly perfect in the real world condition, but significantly underestimated to an equivalent extent in the image-based and computer-rendered conditions.

Gooch and Willemsen [2002] were also the first to explicitly consider the accuracy of egocentric spatial perception in non-photorealistically-rendered (line-drawing style) immersive virtual environments. Similarly to their earlier study, they asked 12 participants to make egocentric distance judgments via direct blind walking in two differently-rendered versions of the same hallway, but this time the conditions were: (1) unmediated real world viewing; and (2) a 3D line-drawing-style rendering of a highly detailed 3D model of the same hallway, in which silhouette edges and crease lines were displayed in black over a white background. As before, they found that results were nearly perfect in the real world, but significantly underestimated with the line-drawing rendering.

Thompson et al. [2004] subsequently replicated these studies in a different (foyer) environment using a between-subjects experimental design. They found that distance judgments were accurate (on average) for participants in the real world condition, but significantly underestimated to an equivalent extent by participants in the stereo panorama, textured 3D model, and wireframe 3D model conditions. These results suggest that the quality of the computer graphics rendering does not, on its own, significantly impact distance perception accuracy in VR. Specifically, in a VR scenario where distances are already being underestimated, reducing the representational fidelity of surface details doesn't appear to significantly compound the problem, at least when an action-based metric is used. Kunz et al. [2009] have found that verbally reported distance estimates may be more severely impacted.

In 2006, Interrante et al. [2006] discovered that people were able to achieve close to real-world-equivalent distance perception accuracy in VR when they were immersed in a highly photorealistic 3D replica environment that evoked a compelling illusion of directly seeing the actual physical environment through the HMD. However, when Phillips et al. [2009], considering potential applications in early-stage architectural design, assessed distance perception accuracy under the same co-located conditions using a sketch-like VR rendering style, they found that people significantly underestimated egocentric distances, compared to in the real world. This suggested that reducing the representational fidelity of surface details in a VR scenario where distances were *not* already being underestimated could make things worse.

In followup studies, Phillips and Interrante [2011] expanded on their earlier work, comparing distance judgments, in a between-subjects design, between three different conditions of texture fidelity on the same 3D model: (1) textures directly created from photographs of participants' concurrently-occupied physical environment; (2) textures obtained by hand-drawing black lines over the most prominent edges in the photographs, adding a coarse grid texture to the floor, and setting all other pixels to white; and (3) textures obtained by replacing all of the white pixels in the line-drawing textures with content from the original photographs. They found that distances were significantly underestimated in each of the two NPR conditions compared to when the unedited photographic textures were used. From these results, they concluded that the illusion of "reality" is fragile in VR experiences, and that once broken, peoples' ability to make accurate judgments of egocentric distances will be significantly impaired, regardless of the extent to which the visual resemblance to reality is diminished.

A remaining open question from all of this prior research is: to what extent, and under what conditions, might we eventually hope to be able to provide the necessary affordances for an accurate spatial understanding of the interior volumes of sketchily-rendered 3D virtual architectural models, experienced via HMDs? We know of no instance yet in which participants have been able to make accurate distance judgments in a non-photorealistic immersive virtual environment. However, if enabling spatial perception accuracy depends primarily on constructing a compelling illusion of *functional realism* [Ferwerda 2003], it may be possible to achieve this independently of a high level of visual realism.

In the experiments reported in this paper, we begin to explore this question by investigating how distance judgments are affected by the degradation of visual realism under conditions of high experiential realism, where more accurate results are expected. This approach is similar in spirit to that taken by Legge et al. [2016], who studied, among other things, the impact of visual blur on judgments of room size in real world environments.

2.2 Video-See-Through Hardware Design

A video-see-through (VST) HMD consists of two main parts: a stereo imaging system and a display system. Bajura et al. [1992] were among the first to attach a camera to the front of a head-mounted display to enable a mixed reality experience, and many others have done so since — typically using two cameras to provide a more compelling sensation of 3D space [Fuchs et al. 1998; State et al. 2005, 1996; Steed et al. 2017; Takagi et al. 2000].

When building a VST HMD, it is important to carefully consider the design of the camera placement. Parallax (apparent displacement) will occur if the camera images are obtained from centers of projection that are offset, in any direction, from the positions of the viewer's eyes. The resulting distortion of the visual field can cause discomfort as well as interfere with the accurate perception of location, size and distance in the scene [Held and Banks 2008; Woods et al. 1993]. With respect to convergence angle, the cameras and displays can either be oriented straight ahead (parallel configuration) or jointly angled inward, so that the lines of sight intersect at a pre-defined distance (toed-in configuration). Several authors report various advantages to the parallel configuration, particularly when the locus of attention is beyond 1-2m [Takagi et al. 2000; Woods et al. 1993]. As an alternative to using physical convergence to ensure sufficient stereo overlap at near distances, State et al. [2001] describe a system that uses two parallel cameras with extremely wide fields of view. Since our focus is on architectural applications, which primarily involves attention to far distances, we also choose a parallel camera layout.

Because the cameras cannot physically occupy the exact same locations as a user's eyes, it is impossible to achieve a perfectly parallax-free system without using mirrors, although horizontal and vertical parallax can be avoided by placing the cameras directly in front of the eyes. Edwards et al. [1993] discuss design considerations for the construction of an optically optimal video see through head-mounted display system that uses two miniature cameras in conjunction with folded mirrors to ensure that the cameras' centers of projection exactly match the eyes' own physical locations, along with lenses that both compensate, partially, for the distortion caused by the optics of the HMD and enable the camera field-of-view (FOV) to exactly match the FOV of the display system. State et al. [2005] describe the design and implementation of a fully orthoscopic (distortion-free) and parallax-free VST HMD constructed from commercially-available components. They use a custom-designed 3D printed mount to support the cameras and mirrors. Although it has both been shown geometrically and confirmed in user studies that avoiding parallax in the depth direction is essential to enabling accurate size and distance perception in the near field, because our use case primarily involves attention to far distances, we elected to pursue a simpler design that foregoes the use of mirrors and focus instead on eliminating parallax in the horizontal and vertical directions only.

2.3 NPR Software

A wide variety of image-based artistic rendering methods have been developed over the years; an excellent survey is provided by Kyprianidis et al. [2013].

There are three key issues to consider with regard to implementing a sketch-like NPR method suitable for use in a stereo VST HMD. First, the algorithm must be computationally efficient, to allow frames to be transmitted with minimal latency. Second, temporal coherence should be preserved, to avoid distracting amounts of flicker. Finally, stereo coherence should also be ensured to minimize binocular rivalry [Kim et al. 2013].

For reasons of efficiency, we did not consider methods that required inferring depth from stereo disparity. Our low latency requirements also ruled out the use of any methods that required the integration of information collected over non-trivial lengths of time. That left us with the category of methods in which intensity gradients are used to stylize a source input, ideally emphasizing the most visually important edges in the scene.

The most common local edge detection techniques are Sobel, Canny, and Difference of Gaussian (DoG). Other methods include Extended DoG (XDoG), and Flow-based DoG (FDoG). The Sobel operator computes a value at pixel (*i*,*j*) as the intensity difference between the vertically and horizontally surrounding pixels [Sobel and Feldman 1968]. The Canny operator uses a multi-stage process in which: 1) a low-pass filter is used to blur the image; 2) a Sobel-like operator is used to compute the intensity gradients at each pixel in the blurred image; 3) the resulting edges are thinned using non-maximum suppression; 4) two globally-defined threshold values are used to classify the remaining pixels into one of three categories: strong edges; weak edges; or not edges; 5) through a process of hysteresis, weak edge pixels are either used to complete strong edges or are removed [Canny 1986]. Because of the nature of the gradient operator–which is a local operation done in the image frame space–stylization algorithms that rely on Sobel and Canny operators will suffer from frame-to-frame and stereo incoherence. Although the Canny operator may be somewhat more robust to spurious edges than the Sobel filter, it has the disadvantage of producing edges whose thickness is independent of the edge scale, and the resulting edge lines are still prone to discontinuity. The Difference of Gaussian (DoG) operator has the potential to produce results that are more amenable to NPR stylization. The DoG method essentially works by subtracting a more highly blurred version of an original image from a less highly blurred version, acting like a band-pass filter. The XDoG method applies thresholding to a sharpened DoG result, producing a two-tone image [Winnemöller et al. 2012]. The FDoG method uses flow along an edge tangent field [Kang et al. 2007] to achieve smoother, more coherent lines.

Klein et al. [2000] were among the first to promote the use of expressively non-photorealistic rendering styles in VR. They demonstrated a system in which temporal- and scale-coherent textures were interactively displayed on architectural interior surfaces, and dark lines accentuated prominent creases in the geometry, evoking a sensation of being immersed in a 3D painting. Fischer et al. [2005] used cartoon stylization to reduce the overall level of apparent realism of real world content to help virtual objects appear less distinguishable from real objects in an augmented reality context.

3 HARDWARE DESIGN

3.1 Adjustable IPD Camera Bracket

The first step in building our VST system was to design and print a specialty stereo camera mount (shown in Figure 1) that attaches two cameras to the front of the HMD at the height of the user's eyes, and allows the cameras to move freely from side to side along two rails, so that their left/right positions can be adjusted according to each participant's IPD. This mounting piece replaces the blinder piece of the nVisor ST50, and can be removed to allow real-world optical see-through observations.

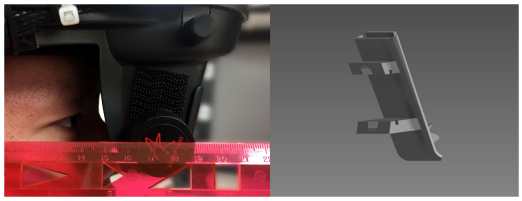

Figure 2: Left: Photo showing the tilt on the nVisor ST50 front panel and ~50mm eye-edge offset; Right: CAD model of the custom-built blinder with camera mount.

As seen in Figure 2, the front of our HMD is angled with respect to the ground, yet the cameras need to face straight ahead. Although we measured everything as precisely as manual tools allow, achieving the perfect angle proved to require an iterative process. We started by 3D printing half of the mount, so that we could observe the camera images with one eye and compare them with the optical view seen through the HMD by the other eye, to verify the correctness of the alignment. After 3D printing our first design, we observed that the camera angle was slightly pitched forward, which caused a vertical offset in the camera image with respect to what was seen through the optical view. To determine the exact amount of tilt correction needed, we used a variant of the FOV calculation technique described in [Jongerius 2015]: we attached an alignment pattern to the wall at eye level and physically observed the amount of vertical displacement between the camera and optical images from multiple distances (Figure 3). Specifically, we used the equations below to first solve for depth distance K between the camera origin and the optical origin, and then solved for the angle α representing amount of camera tilt responsible for the observed vertical offset. The subscripts a and b refer to measurements taken at two different distances, x represents the distance from the camera to the wall, and y represents the vertical displacement between the optical image and the camera image:

$$\frac{y_a}{x_a + K} = \frac{y_b}{x_b + K} \Rightarrow K = \frac{y_a * x_b - y_b * x_a}{y_b - y_a}, \quad \alpha = \tan^{-1}\left(\frac{y_a}{x_a + K}\right)$$

To make this procedure more robust, we repeated the process for multiple pairs of distances and averaged the results. After applying the obtained correction to our design, we printed the mount again and verified that the camera and optical views were closely aligned.

It is worth mentioning that when the camera mount was printed horizontally, we found that the camera rails and visor attachment features broke after just a few uses. Printing vertically–although

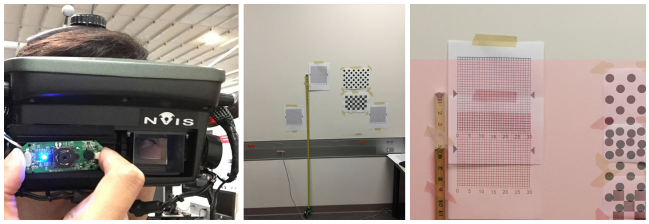

Figure 3: Left: Enabling the bi-ocular viewing; Middle: the optical view; Right: a zoomed-in illustration of the superposition of the camera view (tinted pink).

it required much more support material–provided a stronger and sturdier result.

3.2 Matching Screen Size and Cropping Factor

The next step in our system implementation was to ensure that the image size and aspect ratio of our camera feed was matched to that of our HMD screens. To keep costs low, we chose to use Logitech C615 USB webcams, which can capture video at a resolution of up to 1920x1080 pixels over a 74° diagonal field of view. However, the nVisor ST50 displays images at a resolution of 1280x1024 over a 50° diagonal field of view. Obviously we could not simply crop 1280x1024 pixels out of the available 1920x1080 camera image and display them directly on the HMD, because that cropped camera image would represent content acquired over a 58.6° diagonal FOV. The correct amount of crop needed to span a 50° FOV would be 1065x852; if we take that sub-image and up-sample it to 1280x1024, everything should look normal.

However, there is then the problem of the cameras being offset by about 50mm from the eyes in the depth direction, which causes magnification at all depth distances. To give a sense of the magnitude of the problem, at a distance of 1m the magnification factor from a 50mm camera-eye offset would be 1.05, at 3m it would be 1.0167, and at 5m it would be 1.01, tending towards 1.0 at infinity. Although there is no general way to correctly mitigate the resulting distortion, we elected to adjust the crop area, with the help of the optical view, so as to obtain zero displacement of points at a distance of 10 ft., which leads to magnification at distances closer than 10 ft. and minification at distances farther than 10 ft. While this solution is not perfect, we were able to subjectively verify, in bi-ocular viewing with one eye seeing the camera feed and the other eye viewing the real world, that the fused results did not exhibit noticeable mismatch while freely viewing objects across a wide range of distances at and beyond arm's length, up to the limits of our room size.

Note that to minimize the eye-to-camera offset, and to be able to move the cameras inside the adjustable IPD rails, we extracted the cameras from their shells, stripped down their boards from previous wires, and soldered on new thin USB cables and connectors.

3.3 Backpack Computer

The next component we needed in our VST system was a computer to process the views from the two USB cameras and send the outputs to the HMD's displays. In order to be able to easily use the system in multiple locations outside of our lab, we wanted the display to be portable. We initially looked into using Raspberry Pis, but they were not capable of supporting the necessary frame rate. We then considered using a gaming laptop, but ultimately decided to build our own backpack computer in order to be able to have access to the most powerful graphics processor available at that time.

As seen in Figure 4, our portable computing system consists of two parts: an aluminum backpack frame and a custom computer case. We chose the USGI military-issued A.L.I.C.E. (All-purpose Lightweight Individual Carrying Equipment) backpack frame, which is relatively light–around 3.5 lb.–and versatile. It comes with optional removable cargo shelves, but as those would have added 1.6 lb. we designed our own lightweight computer case that bolts

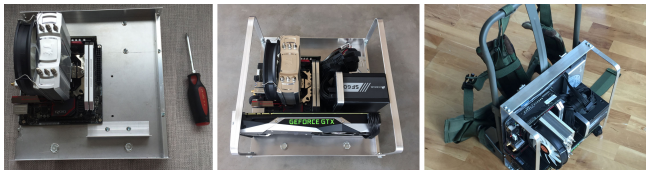

Figure 4: Mounting parts on the custom back-plate, and final backpack computer (without cover) on the ALICE frame.

directly into the frame with wing-nuts for easy attachment and removal. The computer case consists of a 12"x12" weldable 16 gauge metal sheet for the back plate with 1-1/4" strips of angled aluminum for the sides, three 1/2" flat aluminum bars for the front, and a black lincane aluminum sheet for the cover, plus motherboard standoffs, and a number of screws, washers, nuts, and bolts. All of these parts together weigh around 4 lb.

For the computer, we started with a mini ITX form factor mainboard, the ASUS ROG z170 Maximus Impact with built-in wi-fi, and used brass standoffs to mount it professionally to the back plate. We used two Corsair Dominator 16GB memory modules on the mainboard for a total of 32GB of RAM, and placed a 12cm Thermaltake fan on the CPU. A 400GB Intel 750 U.2 NVMe solid state drive (SSD) is mounted in the top-right corner, and a 600W Corsair SFX form factor power supply is mounted directly below it. Finally, a full sized nVidia GTX 1080 graphics card is placed inside the mainboard's PCI-e x16 slot. The total weight for all of the computer components is 7.5 lb, and with the case, everything comes to about 11.5 lb. Adding the backpack frame, we reach a total weight of 15lbs.

3.4 Software

Inspired by the elegant results demonstrated by Winnemöller et al. [2012], we initially explored the use of Flow-based XDoG algorithms to achieve spatially and temporally-coherent aesthetic line-drawing-style renderings from our camera images. Specifically, we started by updating Kyprianidis's open-source CUDA code [Kyprianidis 2015] and adding multi-threading stereo rendering features. To make the code thread-safe, we updated it first with pThreads and then with OpenMP libraries as we found OpenMP to have slightly better performance. However, the resulting XDoG implementation was still too slow for VR, even using the nVidia GTX-1080 graphics card. To improve performance, we down-sampled the input before

Figure 5: Top: Flow-based XDoG CUDA implementation. Bottom: Flow-based XDoG CUDA with threshold.

sending it to the GPU for processing, and then up-sampled the output before sending it to the HMD for display. A sampling ratio of 2.0 to 2.5 produced the results shown in Figure 5 (top). Thresholding this output, we were able to achieve two-tone images like those shown in Figure 5 (bottom).

Ultimately, however, we decided that a simple edge-detection filter would be more suitable for the purposes of our planned experiment, as the images produced by that method appeared closer in spirit to the line-drawing-style renderings employed in the prior NPR distance perception studies. We used the OpenCV Sobel algorithm to render the two webcam streams in parallel with the help of OpenMP. On our custom-built computer, we can easily maintain a rendering speed of >30 fps with this approach, although we still have to contend with the latency from the Logitech cameras.

3.5 System Latency

In our video-see-through system, latency refers to the delay between the time an image is captured by our cameras and the time it is shown to the user on the HMD. Latency has long been linked to cybersickness [Stanney et al. 1998], and even small amounts of latency have been shown to reduce people's sense of presence in VR [Meehan et al. 2003]. Although there is negligible latency associated with the processing of the camera images in our implementation, we found that our webcams have notable latency in sending images. To reduce this latency, we ultimately decided to transmit all of the video at half of the original resolution. Even so, we ended up with a transmission latency of 120-140 milliseconds. This latency affects both processed and unprocessed images, and becomes especially noticeable if the user moves their head quickly.

We also acknowledge that the cameras that we used, the Logitech C615, have a rolling shutter. Because a rolling shutter captures different parts of the scene at different points in time, fast camera movements can lead to blurry images, and fast-moving objects can appear blurred even to a stationary camera [Liang et al. 2008].

Fortunately the design of our present experiment does not require or encourage fast head movements, and the observed environment does not feature any fast-moving objects, but for more general-purpose applications in the future we plan to switch to using USB 3.0 cameras with global shutters, despite their considerably greater expense.

4 USER STUDY

The goal of our experiment, as explained in the Introduction, was to seek further insight into the ultimate potential for enabling architects and their clients to obtain an accurate and intuitive first-person appreciation of the interior volumetric extents of alternative 3D conceptual building designs presented in a sketch-like style via immersive virtual environments technology. We approached this question by comparing participants' distance perception accuracy in a real-world viewing scenario under three different conditions of computer-mediated visual quality degradation.

4.1 Method

Using a within-subjects design, we asked participants to make egocentric distance judgments using direct blind walking under three

Figure 6: Top: Stills from the raw camera feed showing each of the three different hallways. Bottom: Each hallway as it appeared in the NPR viewing condition.

different conditions of visual input: real-world view (V1), unprocessed camera view (V2), and line-drawing-style NPR camera view (V3), in three different physical hallways in our building, denoted H1, H2 and H3. The hallways, shown in Figure 6 (top), were all structurally similar, but varied slightly in their width, color, and level of adornment, so that there could be no mistaking that the locations were different. Figure 6 (bottom) shows what each hallway looked like in the NPR viewing condition. Participants experienced each viewing condition in a different physical hallway, and we varied both the assignments of viewing conditions to hallways and the order in which the viewing conditions were experienced in a balanced way between subjects. Specifically, the first six participants experienced the combinations (H1/V1), (H2/V2), (H3/V3) in shuffled order, the next six participants experienced (H1/V2), (H2/V3), (H3/V1) in shuffled order, and the final six participants experienced (H1/V3), (H2/V1), (H3/V2) in shuffled order. With this design, we aimed to enable a within-subjects comparison of distance estimation accuracy between rendering styles while avoiding carry-over effects from prior exposure to the same environment under different viewing conditions. Pooling the data across subjects, we also had the ability to verify the absence of a separate significant effect of the different hallway environments on distance judgment accuracy. Participants wore the HMD and backpack computer in all three viewing/environment conditions, without the blinder in V1 and with the custom adjustable-IPD camera mount in V2 and V3, along with disposable ear plugs to mask audio input from the surrounding environment. The experiment was conducted with the approval of our university's Institutional Review Board (IRB).

4.1.1 Participants. We recruited a total of 22 participants for this experiment from our local University community, through a combination of personal contacts and posted flyers. After the first three participants, we recognized an error in our protocol and were required to start over, and one additional participant elected to discontinue the experiment after the first trial and was replaced, resulting in a total of 18 (12 male, 6 female) who completed the experiment. None of the participants were members of our lab.

The 18 participants ranged in age from 18 - 64 years old (μ = 25.8 ± 10.8), and were demographically varied. Each participant was compensated with a $15 gift card to an online retailer.

4.1.2 Procedure. Participants arrived one at a time to our lab and were given written instructions describing the experiment procedure and asked to sign a consent form. We first screened for low vision by asking participants to read a line of letters, from a distance of between 3-4 meters, displayed on a 15" monitor at a computed visual acuity of 20/60. If participants wore corrective lenses in the HMD, we allowed them to wear those lenses during the vision test. We then screened for stereo vision ability by showing participants three different random-dot stereograms of increasing complexity on the HMD and asking them to describe what they saw. All participants passed both tests. Next, participants put on the backpack computer and ear plugs, and we adjusted the positions of the cameras on the VST mount to match the locations of their pupils, first ensuring that the pupils were at the center of the person's eyes. We then attached opaque black fabric around all of the edges of the HMD to prevent any light leakage or inadvertent peeking at the floor. Finally, we asked participants to close their eyes, and we directed them, either by gripping their elbow or asking them to hold on to a prop, out of our lab, into an elevator, and out to one of the test hallways. We ensured that the hallways were unoccupied by scheduling participants on evenings and weekends, and by posting an associate around the corner to make sure that the experiment was not interrupted.

In each environment, participants were asked to perform blind walking distance judgments to targets indicated by wide pieces of masking tape, placed one at a time at six different distances in front of them: 5m, 5.5m, 6m, 7m, 7.5m, and 8m. The distances of 5.5m and 7.5m were shown first and treated as training trials, with the results discarded. The remaining four trials were presented in random order, determined ahead of time using a computer program. Participants did not receive any feedback about their performance at any time during the experiment. The starting location for each trial was arbitrarily varied, and targets were placed manually by measuring out the appropriate distance in front of the participant from wherever they stood, while they were blindfolded. The starting location was also marked. Once the two experimenters were out of view, the participant was instructed to open their eyes and, in V1, to lift their blindfold; in V2 and V3 the graphics were made visible on the HMD. Participants then took visual aim at the target, said "ready," then closed their eyes and replaced the blindfold or had the graphics set to black, and walked to where they thought the target was. An aluminum bracket was placed adjacent to their toes at their stopping location, and the walked distance was measured with a laser. While still blindfolded, the participant was then led by one of the experimenters to a different location in the hallway from which to start the next trial, while the second experimenter removed the tape marks.

After each block of trials, the participant was led, still blindfolded and with their eyes closed, back to the lab for a short break, where they were given some refreshments and asked to fill out a simulator sickness questionnaire. At the end of the entire experiment we asked participants to fill out a survey form in which they rated, on a 7-point scale, the visual and functional realism of their experience

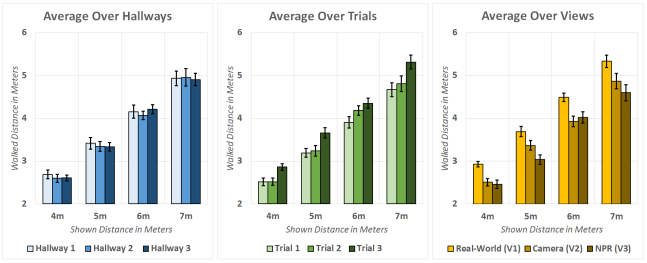

Figure 7: Left: No significant difference across different hallways. Middle: Accuracy tended higher in later blocks. Right: Real-world better than camera and NPR, but no significant difference between the latter two.

during each block of trials, as well as their level of "presence" in each environment. We took care to identify the conditions by number rather than by potentially biasing terms such as "real world" or "NPR." We also asked participants to answer some more open-ended questions, from which we sought to infer the extent to which they may have been fooled into thinking they were experiencing a virtual model rather than a real environment in some of the conditions. The entire experiment took about 90 minutes.

4.2 Results

Figure 7 summarizes our findings. First of all, we note that participants walked significantly farther for the more distant targets, which establishes a baseline level of credibility for the experiment. We can also see that the average distances walked were similar across the three different hallway environments, verifying that the subtle environmental differences did not have a significant effect. Overall, we note that participants significantly underestimated distance in all three conditions, possibly due to an influence of the heavy backpack [Bhalla and Proffitt 1999]. Walked distances also tended to increase over time (p = 0.015).

Most important to the goals of our experiment, is the finding of a significant main effect of viewing condition on the distance estimation accuracy (p = 0.001). TukeyHSD tests revealed a significant difference both between the real-word vs. camera view conditions (p = 0.021), and the real-world vs. NPR view conditions (p = 0.002), but not between the camera and NPR conditions (p = 0.686).

The survey results showed equivalent ratings for presence and functional realism, so we only further consider the two realism questions. As expected, participants described significantly higher levels of both visual and functional realism in the real world condition than in the camera or NPR conditions. They also described higher levels of both visual and functional realism in the camera condition than in the NPR condition. However, what stands out is that while participants' ratings of functional realism were nearly identical to their ratings of visual realism in both the real world and camera conditions, their ratings of functional realism were significantly higher than their visual realism in the NPR case.

We did not observe any significant incidence of cybersickness, and even though we did not explicitly ask, many participants volunteered that they thought they were seeing a filtered camera view in our NPR condition. One, who had a computer vision background, even recognized that we were using a Sobel filter.

5 DISCUSSION AND CONCLUSIONS

The results of this experiment are encouraging with respect to our ultimate goals, in that a severe degradation of visual realism did not cause a significant decrease in peoples' distance perception accuracy. This finding is consistent with the results of Thompson et al. [2004], though the mode of graphics presentation (image-processed live video vs. rendered 3D model) is quite different. We are disappointed that overall accuracy levels were not higher, but recognize that multiple compromising factors likely contributed to that outcome. Besides the energetic demands of the heavy backpack and awkward ergonomics of the HMD, there were the problems of latency in the camera feed, and parallax due to the offset between the camera origins and the eyes. However, hopefully such problems can be more easily avoided as the capabilities of VR technology advance. In future work, we plan to develop a fully parallax-free VST mount with mirrors and USB 3.0 cameras for our SMI eye-tracking-enabled HTC Vive.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation through grants II-NEW 1305401 and CHS: Small 1526693, and by the Linda and Ted Johnson Digital Design Consortium Endowment. S. Aseeri was supported by a King Abdulaziz University Scholarship from the Saudi Arabian Cultural Mission (SACM).

REFERENCES

Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. 1992. Merging virtual objects with the real world: seeing ultrasound imagery within the patient. *Computer Graphics (Proc. ACM SIGGRAPH)* 26, 2 (July 1992), 203–210.

Mukul Bhalla and Dennis R Proffitt. 1999. Visual–motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance* 25, 4 (Aug. 1999), 1076–1096.

John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8, 6 (Nov. 1986), 679–698.

Emily K. Edwards, Jannick P. Rolland, and Kurtis P. Keller. 1993. Video see-through design for merging of real an virtual environments. In *Proc. Virtual Reality Annual International Symposium (VRAIS)*. 223–233.

James A. Ferwerda. 2003. Three varieties of realism in computer graphics. In *Proc. SPIE 5007, Human Factors and Electronic Imaging VIII*.

Jan Fischer, Dirk Bartz, and Wolfgang Straßer. 2005. Stylized augmented reality for improved immersion. In *Proc. IEEE Virtual Reality (VR)*. 195–202.

Henry Fuchs, Mark A. Livingstone, Ramesh Raskar, D'nardo Colucci, Kurtis Keller, Andrei State, Jessica Crawford, Paul Rademacher, Samuel H. Drake, and Anthony A. Meyer. 1998. Augmented reality visualization for laproscopic surgery. In *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 934–943.

Amy A. Gooch and Peter Willemsen. 2002. Evaluating space perception in NPR immersive environments. In *Proc. Non-Photorealistic Animation and Rendering (NPAR)*. 105–110.

Robert T. Held and Martin S. Banks. 2008. Misperceptions in stereoscopic displays: a vision science perspective. In *Proc. Applied Perception in Graphics and Visualization (APGV)*. 23–32.

Victoria Interrante, Brian Ries, and Lee Anderson. 2006. Distance perception in immersive virtual environments, revisited. In *Proc. IEEE Virtual Reality (VR)*. 3–10.

Jerry Jongerius. (accessed July 12, 2015). *Measuring Lens Field of View (FOV) (aka: Locating the Lens Entrance Pupil)*. http://www.panohelp.com/lensfov.html

Henry Kang, Seungyong Lee, and Charles K Chui. 2007. Coherent line drawing. In *Proc. Non-Photorealistic Animation and Rendering (NPAR)*. 43–50.

Yongjin Kim, Yunjin Lee, Henry Kang, and Seungyong Lee. 2013. Stereoscopic 3D line drawing. *ACM Transactions on Graphics* 32, 4 (July 2013), Art. 57.

Allison W. Klein, Wilmot Li, Michael M. Kazhdan, Wagner T. Corrêa, Adam Finkelstein, and Thomas A. Funkhouser. 2000. Non-photorealistic virtual environments. In *Proc. ACM SIGGRAPH*. 527–534.

Scott A. Kuhl, William B. Thompson, and Sarah H. Creem-Regehr. 2009. HMD calibration and its effects on distance judgments. *ACM Transactions on Applied Perception* 6, 3 (Aug. 2009), Art. 19.

Benjamin R. Kunz, Leah Wouters, Daniel Smith, William B. Thompson, and Sarah H. Creem-Regehr. 2009. Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception, & Psychophysics* 71, 6 (Aug. 2009), 1284–1293.

Jan Eric Kyprianidis. 2012 (accessed May 25, 2015). *XDoG-DEMO*. http://code.google.com/p/xdog-demo

Jan Eric Kyprianidis, John Collomosse, Tinghuai Wang, and Tobias Isenberg. 2013. State of the "Art": a taxonomy of artistic stylization techniques for images and video. *IEEE Trans. Visualization and Computer Graphics* 19, 5 (May 2013), 866–885.

Gordon E. Legge, Rachel Gage, Yihwa Baek, and Tiana M. Bochsler. 2016. Indoor spatial updating with reduced visual information. *PLoS ONE* 11, 3 (March 2016).

Chia-Kai Liang, Li-Wen Chang, and Homer H. Chen. 2008. Analysis and compensation of rolling shutter effect. *IEEE Trans. Image Proc.* 17, 8 (2008), 1323–30.

Michael Meehan, Sharif Razzaque, Mary C. Whitton, and Frederick P. Brooks, Jr. 2003. Effect of latency on presence in stressful virtual environments. In *Proc. IEEE Virtual Reality (VR)*. 141–148.

Betty J. Mohler, Sarah H. Creem-Regehr, and William B. Thompson. 2006. The influence of feedback on egocentric distance judgments in real and virtual environments. In *Proc. Applied Perception in Graphics and Visualization (APGV)*. 9–14.

Alex Peer and Kevin Ponto. 2017. Evaluating perceived distance measures in room-scale spaces using consumer-grade head mounted displays. In *Proc. IEEE Symp. on 3D User Interfaces (3DUI)*. 83–86.

Lane Phillips and Victoria Interrante. 2011. A little unreality in a realistic replica environment degrades distance estimation accuracy. In *Proc. IEEE Virtual Reality (VR)*. 235–236.

Lane Phillips, Brian Ries, Victoria Interrante, Michael Kaeding, and Lee Anderson. 2009. Distance perception in NPR immersive virtual environments, revisited. In *Proc. Applied Perception in Graphics and Visualization (APGV)*. 11–14.

Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. 2013. The perception of egocentric distances in virtual environments-a review. *Comput. Surveys* 46, 2 (Nov. 2013), Article No. 23.

Brian Ries, Victoria Interrante, Michael Kaeding, and Lee Anderson. 2008. The effect of self-embodiment on distance perception in immersive virtual environments. In *Proc. Virtual Reality Software and Technology (VRST)*. 167–170.

Jutta Schumann, Thomas Strothotte, Stefan Laser, and Andreas Raab. 1996. Assessing the effect of non-photorealistic rendered images in CAD. In *Proc. ACM SIGCHI*. 35–41.

Irwin Sobel and Gary Feldman. 1968. A 3x3 isotropic gradient operator for image processing. (1968).

Kay M. Stanney, Ronald R. Mourant, and Robert S. Kennedy. 1998. Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments* 7, 4 (Aug. 1998), 327–351.

Andrei State, Jeremy Ackerman, Gentaro Hirota, Joohi Lee, and Henry Fuchs. 2001. Dynamic virtual convergence for video see-through head-mounted displays: maintaining maximum stereo overlap throughout a close-range work space. In *Proc. International Symposium on Augmented Reality (ISAR)*. 137–146.

Andrei State, Kurtis P. Keller, and Henry Fuchs. 2005. Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. In *Proc. Int. Symp. Mixed and Augmented Reality (ISMAR)*. 28–31.

Andrei State, Mark A. Livingstone, William F. Garrett, Gentaro Hirota, Mary C. Whitton, Etta D. Pisano (MD), and Henry Fuchs. 1996. Technologies for augmented-reality systems: realising ultrasound-guided needle biopsies. In *Proc. ACM SIGGRAPH*. 439–446.

Anthony Steed, Yonathan Widya Adipradana, and Sebastian Friston. 2017. The AR-Rift 2 Prototype. In *Proc. IEEE Virtual Reality (VR)*. 231–232.

Akinari Takagi, Shoichi Yamazaki, Yoshihiro Saito, and Naosato Taniguchi. 2000. Development of a stereo video see-through HMD for AR systems. In *Proc. International Symposium on Augmented Reality (ISAR)*. 68–77.

William B. Thompson, Peter Willemsen, Amy A. Gooch, Sarah H. Creem-Regehr, Jack M. Loomis, and Andrew C. Beall. 2004. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments* 13, 5 (Oct. 2004), 560–571.

Peter Willemsen, Mark B. Colton, Sarah H. Creem-Regehr, and William B. Thompson. 2009. The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Transactions on Applied Perception* 6, 2 (Feb. 2009), Art. 8.

Peter Willemsen and Amy A. Gooch. 2002. Perceived egocentric distances in real, image-based, and traditional virtual environments. In *Proc. IEEE Virtual Reality*. 275–276.

Holger Winnemöller, Jan Eric Kyprianidis, and Sven C. Olsen. 2012. XDoG: an extended difference-of-Gaussians compendium including advanced image stylization. *Computers & Graphics* 36, 6 (Oct. 2012), 740–753.

Andrew J. Woods, Tom Docherty, and Rolf Koch. 1993. Image distortions in stereoscopic video systems. In *Proc. SPIE 1915, Stereoscopic Displays and Applications IV*. 36–48.

Mary K Young, Graham B Gaylor, Scott M Andrus, and Bobby Bodenheimer. 2014. A comparison of two cost-differentiated virtual reality systems for perception and action tasks. In *Proc. ACM Symposium on Applied Perception (SAP)*. 83–90.