You Invaded my Tracking Space! Using Augmented Virtuality for Spotting Passersby in Room-Scale Virtual Reality

Julius von Willich
TU Darmstadt, Germany
willich@tk.tu-darmstadt.de

Markus Funk
TU Darmstadt, Germany
makufunk@hotmail.com

Florian Müller
TU Darmstadt, Germany
mueller@tk.tu-darmstadt.de

Karola Marky
TU Darmstadt, Germany
marky@tk.tu-darmstadt.de

Jan Riemann
TU Darmstadt, Germany
riemann@tk.tu-darmstadt.de

Max Mühlhäuser
TU Darmstadt, Germany
max@tk.tu-darmstadt.de

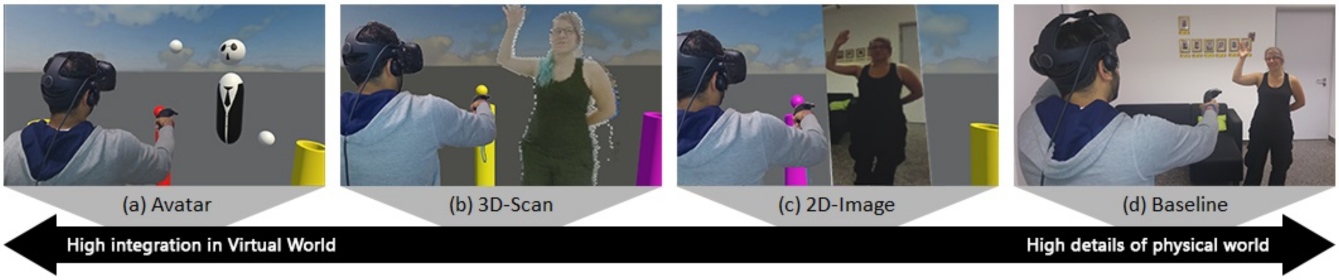
(a) Avatar
(b) 3D-Scan
(c) 2D-Image
(d) Baseline
High integration in Virtual World
High details of physical world

Figure 1. The visualization of passersby that we used in the study. (a) A passerby is represented using a *Avatar* that matches the games general style. (b) A colored *3D-Scan* visualization showing the passerby with different amount of details. (c) A *2D-Image* that is taken from the head-mounted camera in the user's HMD. (d) The *Baseline*, which is pointing at passersby in the physical world.

ABSTRACT

With the proliferation of room-scale Virtual Reality (VR), more and more users install a VR system in their homes. When users are in VR, they are usually completely immersed in their application. However, sometimes passersby invade these tracking spaces and walk up to users that are currently immersed in VR to try and interact with them. As this either scares the user in VR or breaks the user's immersion, research has yet to find a way to seamlessly represent physical passersby in virtual worlds. In this paper, we propose and evaluate three different ways to represent physical passersby in a Virtual Environment using Augmented Virtuality. The representations encompass showing a *3D-Scan*, showing a *Avatar*, and showing a *2D-Image* of the passerby. Our results show that while a *2D-Image* and a *Avatar* are the fastest representations to spot passersby, the *Avatar* and the *3D-Scan* representations were the most accurate.

Author Keywords

VR; AR; Passersby Visualization

DIS '19, June 23–28, 2019, San Diego, CA, USA

ISBN 978-1-4503-5850-7/19/06...$15.00
DOI: https://doi.org/10.1145/3322276.3322334

CCS Concepts

•**Human-centered computing** → **Virtual reality; Pointing;** *User studies;*

INTRODUCTION

Room-scale VR systems using head-mounted displays (HMDs) are now commercially available and have found their way into users' living rooms (e.g., the HTC Vive). These systems can create very immersive experiences for the user wearing the HMD. As such systems are usually set up in central places in the user's home, e.g., the living room [40], there might be other persons that cannot or do not want to be part of the VR experience. This leads to two problems: First, these other persons will at some point invade a user's tracking space to interact with them. A concept on how this could look is presented in Figure 2. Such interaction will, of course, create a distraction for the VR user which will ideally be minimized using Augmented Virtuality (AV) approaches such as those presented in this paper. Second, the VR users lose track of the physical world and therefore might get hurt or break something. In order to avoid injuries and damages, information from the physical world can be displayed directly in the Virtual Environment (VE). This is known as AV [24]. Previous research has investigated using AV for displaying virtual keyboards to improve text entry [18], reaching one's coffee mug [2], or displaying persons that want to interact with VR users [23]. The latter work compared opaque to non-opaque overlays to represent passersby. Thereby, the opaque representation creates the highest awareness but also distracts the VR user the most. In this paper, we will focus on aware-

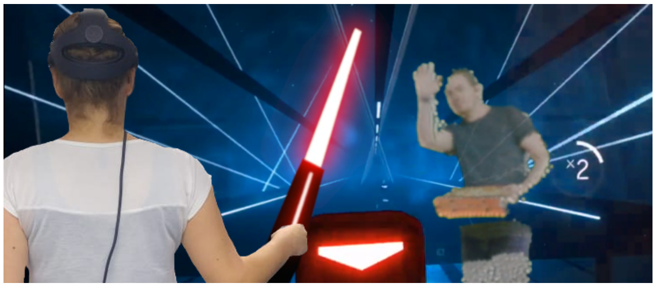

Figure 2. A user is completely immersed in a VR game. However, she can notice passersby that invade her tracking space through Augmented Virtuality.

ness in terms of spatial awareness as knowing about where and how another person is positioned.

We build upon this previous research and further investigate which AV visualization for displaying passersby in VR is best suited for locating and interacting with passersby and for keeping focused on the task at hand. In this paper, we compare three visualizations with different levels of detail and one *Baseline*. These approaches can be placed on a scale from full integration into the VE with the *Avatar* to a high level of detail from the physical world with the *2D-Image*:

- a *Avatar* that matches the VE's visual style, not linked to the passerby's appearance,
- a *3D-Scan* that only shows the passersby in high quality but does not show any environment,
- a *2D-Image* revealing full details of the passersby and the physical environment.

We compare the visualizations with regard to how accurately participants can locate the passerby, how distracted they are from the task at hand and how quickly the passerby is located. Being able to locate a passerby accurately enables the user in VR to avoid them and to hold a natural conversation, keeping approximate eye contact.

Our results reveal that the *Avatar*, and the *3D-Scan* representations were the most accurate when it comes the representing the passerby's position. The *2D-Image* and the *Avatar* are the fastest representations to spot and locate passersby. This shows that providing more information from the physical world does enhance the location time, but worsens the accuracy. We conclude that in order to prevent injuries the *2D-Image* and the *Avatar* are the most promising due to the reduced time.

RELATED WORK

We group related approaches according to two categories. First, we describe *interaction opportunities between the physical and virtual reality* and second, we provide an overview of systems using *Augmented Virtuality*.

Interaction between Physical and Virtual Reality

Research has focused on bringing the physical and virtual worlds closer together. Initial systems suggested bringing digital information into the physical world: for example, giving a physical representation to digital information [15] or overlaying physical objects with additional virtual information [32] or providing a reverse cave [14]. Other research projects have focused on enhancing the interaction between physical and virtual locations [6] by creating a hybrid interaction space [34]. However, ultimately the vision for these systems is to reach a One Reality [33], where the physical world and the virtual world seamlessly intertwine.

Although it is possible that multiple users to share one physical tracking space [22, 27, 19, 35, 26], most room-scale VR systems are currently only used by one person that is immersed in the VE. Thus, related research has developed possibilities to let non-HMD users be part of the VR experience by projecting parts of the VE into the physical environment [7], or showing the VE on a face-mounted display [8]. Such a device can also be used to show passersby a virtual face that visualizes a user's eye-gaze [21, 3] while in VR. Passersby can even enhance the experience of the person in VR by manually providing haptic feedback [4, 5, 9], or by being able to modify the VEs [13, 39]. In contrast to the VR user's actions that are visualized in the physical world, most of these experiences do not provide a back channel from the VE to the real world by visualizing passersby in the VE this circumstance can be easily rectified.

Augmented Virtuality

In the Reality-Virtuality continuum by Milgram et al. [24], two different categories of systems between the physical reality and the virtual reality are defined: Augmented Reality (AR), which overlays the real world with information of the virtual world, and Augmented Virtuality (AV), which overlays the virtual world with information from the real world. Although systems that are using AV have been around for the last two decades, they have not received as much attention as AR or VR [36]. However, we assume that through the recent popularity of room-scale VR systems, also AV will gain more popularity.

Over the years, research has introduced many example scenarios for AV. For example, in 1997 the *Windows on the World* [37] system was introduced. It enables a user that is immersed in a VE to look out of a virtual window to perceive the physical world. Further, Regenbrecht et al. [31, 29] use video streams of real faces in a virtual environment to hold an AV video conference. Further systems are using AV to steer an unmanned aerial vehicle (UAV) in the physical world through a VR environment [28], to collaborate with others [12], to assist in an image-guided neurosurgery [25], or to remotely inspect [41] and finding hazards in physical environments [1]. AV can also be used to enhance VR by introducing haptics from the physical world into the virtual world [38]. For example, Knierim et al. [17] actively navigate haptic props in the physical world to overlay virtual objects, while Hettiarachchi and Wigdor [11] overlay physical objects with digital models.

Recently, research projects introduced using AV to enhance room-scale VR with features that simplify an everyday usage of VR systems. For example, Budhiraja et al.[2] enable the user to reach a physical coffee mug while being immersed in VR, while Knierim et al. [18] visualize the physical keyboard while a user is immersed in VR to improve the typing performance. Most related to our work is the work of McGill et al. [23], who compared a partially opaque visualization of passersby, a fully opaque visualization of passersby, and a baseline without visualization of passersby. Their results show that a fully opaque visualization leads to a higher awareness, but also to a higher distraction through passersby. Whether high awareness or low distraction is desired, depends on the application. For entertainment applications, avoiding distraction is the main objective, for industrial applications, instant, and high awareness is key.

Building on the results of McGill et al. [23], we experiment with presenting passersby in more different visualizations from being fully integrated into the VE using a *Avatar* that matches the VE's style and moves according to the passerby's movement, to a *2D-Image*, a *3D-Scan*, and a *Baseline* condition, requiring the participants to remove their headset.

VISUALIZING PASSERSBY

In this paper, we explore different visualizations of passersby in VE. We classify the passersby visualizations from a high level of physical reality to a high level of integration into the VE. In the following, we present three visualizations of passersby that enable different levels of integration into the VE. First we give information on our *prototype implementation*, then we explain the visualization methods *Avatar*, *3D-Scan* and *2D-Image* and their integration in our prototype in detail.

Prototype Implementation

For our evaluation, we implemented all visualization methods in Unity 2018.1.3f1 using an HTC Vive. To detect passersby, we use a Microsoft Kinect V2 and its built-in skeleton tracking. The Kinect's field of vision includes part of the tracking space outside of the play area, namely the area of our room from which the passerby could approach. We place the Kinect on a tripod outside of the play area and thus the HTC Vive's Chaperone borders in order to avoid knocking it over by accident. An HTC Vive tracker is attached to the Kinect in order to track its position and orientation in 3D space which is why it is placed inside the tracking space. While this works fine in our controlled setup, it is not suited for uncontrolled environments where passersby can approach from any angle. For the future, we envision environment-mounted depth cameras for tracking passersby, similar to related work [16, 20].

Avatar

Representing the passerby as a *Avatar* offers the highest level of integration into the VE. The visual design of the *Avatar* can match the VE's style. We assume that matching the style of the passerby visualization to the VE will not distract the participants as much as presenting passersby in more detail. For animating the *Avatar*, we move the *Avatar* according to skeleton data of the passerby that we measured using a Microsoft Kinect. The aim of this animation is the retain some degree of non-verbal communication, making the *Avatar* feel more like a real person. We further made sure that the size of the *Avatar* matches the size of the passerby in the physical world. This was done in order to ensure correct mapping of the *Avatar*'s virtual and the passerby's real nose, which is later used to evaluate the accuracy of the users' pointing. We chose this spatial measurement over subjective measurements such as the influence of the passerby visualization to have objectively comparable results. The *Avatar* is depicted in Figure 1 on the far left side.

3D-Scan

The *3D-Scan* visualization adds more details of the passerby's physical appearance to the visualization. It is inspired by related work that proposes point clouds to represent physical objects or persons [20, 30]. This visualization shows the point cloud data of the passerby's body and hides the data from the physical environment. In our implementation, we use the Microsoft Kinect's depth data and combine it with the RGB data stream to obtain a colored *3D-Scan*. We further use the passerby's position from the skeleton data to remove data points that are part of the environment. Again, we made sure that the visualization of the person in the VE through the *3D-Scan* matches the actual size of the passerby in the physical world.

Our *3D-Scan* visualization comes closes to the visualization used by McGill et al. [23]. As their study showed that an opaque visualization of others led to a significantly higher awareness, we also choose to use an opaque visualization for our *3D-Scan* visualization.

2D Image

Similar to the "Real World Windowed" approach, proposed in related work [2], we created a *2D-Image* that shows a passerby in life-size in the VE. This visualization conveys the most details from the physical world, as it uses an RGB video feed showing the passerby. For displaying the video feed in the VE, we use a plane facing the user, which is placed at the passerby's physical position. As we want to display the passerby in a correct angle facing the participants, we used the front-facing camera of the HTC Vive headset as the video source for displaying the *2D-Image*. We further cropped the video to only show a rectangle around the participants' position, which we retrieved from the Microsoft Kinect's skeleton data. Using this data, we made sure that the person in the *2D-Image* matches the size of the person in the physical world which, in conjunction with placing the image at the position of the tracked person, ensured a correct overall representation.

EVALUATION

We conducted a user study to evaluate the effect of the different AV visualizations of passersby. Our study aims to prove the following two hypotheses:

H_1 Using higher details of the physical world in passersby visualization leads to a higher spatial awareness in the participants.

H_2 Using higher details of the physical world in passersby visualization leads to a higher distraction from the played game.

Task: A Ball Game

For the user study, we implemented a simple game to immerse the participants in a VE. The game consists of eight tubes placed around the center of the HTC Vive's play area, surrounded by see-through walls to limit the physical tracking space. Once the game is started, colored balls are spawned throughout the play area. The balls are bouncing around, which requires the participants to move through the VE in order to pick them up. The game's goal is to pick up the balls and place them in a tube that matches the ball's color (see

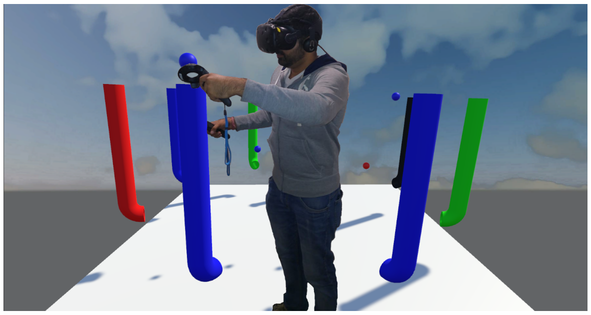

Figure 3. We created a simple ball game for immersing our participants in a VE in our study. Their task is to place the balls in the tubes with the same color.

Figure 3). It aimed to disorient the users so they would have to look for the passerby.

Study Design

We designed our user study following a repeated measures design with the passersby visualization as the only independent variable having four levels (*Avatar*, *2D-Image*, *3D-Scan*, and a *Baseline* without any overlay). As dependent variables, we measured the two dimensional offset from the perceived passerby position to the actual passersby position (d_{offset}), the time the participants required to locate the passersby (t_{locate}), the score that was achieved in the ball game, and the perceived cognitive load measured with the Raw NASA Task Load Index (RTLX) [10] score. Furthermore, we used a custom questionnaire with questions regarding the passerby's representation.

Baseline

We designed the *Baseline* condition as follows. A visual change in the play area, tinting it red, notifies the participant about the presence of a passerby. This cue was chosen to not give away the passerby's position, as verbally addressing would, and is used in the other conditions as well. The participant has to take off the headset and point at the nose of the passerby using the HTC Vive's controller. With this, we aim to simulate a verbal interaction, including eye contact, between the participant and the passerby as it would occur in a real scenario. The nose was chosen to simulate eye contact over the actual eyes to avoid confusion, e.g., which eye to focus and still retain a common point of reference. Afterward, the participant needs to re-enter the VE by equipping the headset again and continue with the task.

Procedure

In this section, we detail our study procedure. Our study follows the guidelines of the ethics commission at our institution.

Welcome and Demographics

At the beginning of the study, we welcomed the participants and explained the study's purpose to them. We explained to them which data is collected during the experiment. Afterward, we asked the participants to sign a consent form and to provide demographics.

Ball Game Familiarization

Once the questionnaire and consent form were filled out, the participants were given the opportunity to familiarize themselves with the ball game.

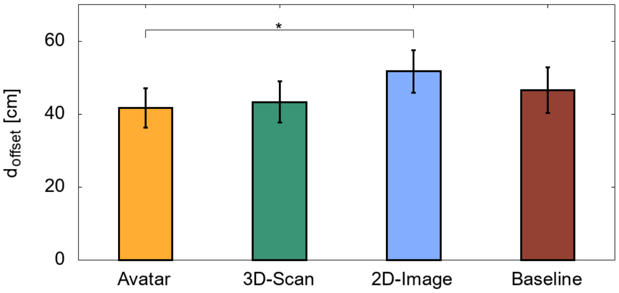

Figure 4. The average offset d_{offset} that the participants had in locating the passerby's nose. All error bars indicate the standard error. The asterisk (*) indicates a statistically significant difference between the visualizations.

Interaction and Tracking

Once the participants felt sufficiently familiar with the game and the VE, we presented the different visualizations and started recording data. A Balanced Latin Square randomization determined their order. To assess the two-dimensional offset from the perceived passerby position to the actual passersby position (d_{offset}), we tracked the position of the examiner's head, who also posed as the passerby, with an HTC Vive Tracker. The offset to the position of the experimenter's nose was calibrated using an HTC Vive controller. This ensured that the position of the examiner's nose, which was the fix-point the participants should locate, was always correctly tracked. Once the necessary calibration was performed, we started the ball game with a total duration of four minutes. After one, two, and three minutes into the game, the examiner took one of three possible passerby positions, which were selected, so they are equally far away from the HTC Vive's tracking space's center, according to a Balanced Latin Square. The examiner then triggered a visual indicator in the game, prompting the participants to locate the examiner and point at their nose. The examiner's representation was only visible once the visual indication was active and both, the passerby's representation and the red tint, vanished once the participants located the passerby. This also ensured that participants received feedback about whether their input was recorded or not, to ensure recording a reaction. After four minutes, the game finished and the participants were asked to fill out an RTLX [10] questionnaire as well as our custom questionnaire evaluating the current passerby visualization. We repeated this procedure for all three passersby visualizations and the *Baseline* condition.

Condition Ranking and Debriefing

After concluding all four conditions, the participants were asked to fill out our questionnaire about the visualization styles and provide additional qualitative feedback using a semi-structured interview. Finally, the participants were given the opportunity to ask questions.

Participants

We recruited 16 participants for our user study through our university's mailing list and did not provide any compensation for taking part in the study. Six of them identified as female and 10 identified as male. The participants were between 20 and 38 years old ($M = 25.06$ years, $SD = 4.22$ years) and were students of various subjects and university employees. Two participants did not have any previous VR experience.

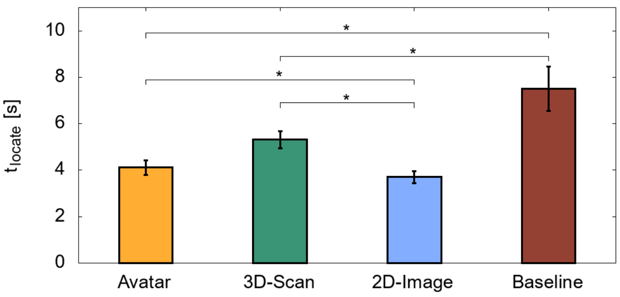

Figure 5. The average time t_{locate} that the participants needed to locate the passerby's nose. All error bars indicate the standard error. The asterisk (*) indicates a statistically significant difference between the visualizations.

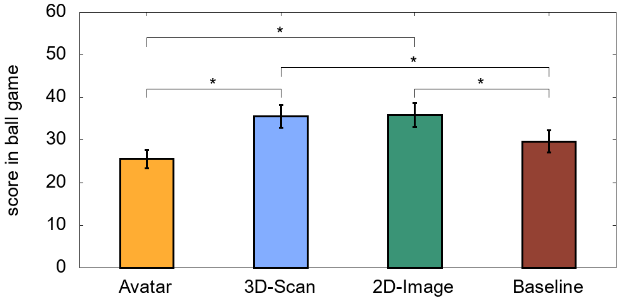

Figure 6. The average score that the participants reached in the study when playing the ball game according to the different visualizations. All error bars indicate the standard error. The asterisk (*) indicates a statistically significant difference between the visualizations.

Quantitative Results

We compared the passersby visualizations in VR with a one-way repeated measures ANOVA. Mauchly's test showed that the sphericity assumption was violated for t_{locate}. Therefore, we used the Greenhouse-Geisser correction to adjust the degrees of freedom ($\varepsilon = .373$ for t_{locate}). We further used the Bonferroni correction for all post-hoc tests.

Target Offset

First, we analyzed the offset between the actual target and the target the participants pointed at, d_{offset}. As a point of reference for the participants, we chose the examiners nose, which was represented in all visualizations. The *Avatar* ($M = 41.72$cm, $SD = 21.36$cm) led to the lowest d_{offset}, followed by the *3D-Scan* ($M = 43.26$cm, $SD = 22.59$cm), the *Baseline* ($M = 46.55$cm, $SD = 25.05$cm), and the *2D-Image* ($M = 51.72$cm, $SD = 23.14$cm). A repeated measures ANOVA revealed a significant difference between the conditions ($F(3,141) = 2.893$, $p = .038$). The post-hoc test using pairwise comparisons showed a significant difference in d_{offset} between the *Avatar* and the *2D-Image* ($p < 0.05$). The effect size estimate shows a small effect ($\eta^2 = .058$). The results are depicted in Figure 4.

Location Time

Considering the time it took the participants to locate the passerby, t_{locate}, the *2D-Image* led to the fastest time ($M =$ 3.7s, $SD =$ 1.05s), followed by the *Avatar* ($M =$ 4.12s, SD = 1.25s), the *3D-Scan* ($M =$ 5.31s, $SD =$ 1.47s), and the *Baseline* condition ($M =$ 7.51s, $SD =$ 3.81s). A repeated measures ANOVA revealed a significant difference between the conditions ($F(1.12,16.79) = 13.023$, $p < .01$). The post-hoc pairwise comparisons revealed a statistically significant difference between the *Baseline* vs. *2D-Image*, *Baseline* vs. *Avatar*, *2D-Image* vs. *3D-Scan*, and *Avatar* vs. *3D-Scan* (all $p < 0.05$). The effect size estimate shows a large effect ($\eta^2 = .237$). The results are depicted in Figure 5.

Ball Game Score

When analyzing the score that was achieved in the ball game using the different passersby visualizations, the *3D-Scan* ($M = 35.88$, $SD = 11.53$) and the *2D-Image* ($M = 35.63$, SD = 10.66) resulted in similar scores, while the *Baseline* ($M =$ 29.69, $SD = 10.44$) and the *Avatar* ($M = 25.56$, $SD = 8.62$) led to lower scores. A repeated measures ANOVA revealed a significant difference between the conditions $F(3,45) =$ 14.452, $p < .001$. The pairwise post-hoc comparisons revealed a statistically significant difference between the following conditions: *Baseline* vs. *2D-Image*, *Baseline* vs. *3D-Scan*, *2D-Image* vs. *Avatar*, and *Avatar* vs. *3D-Scan* (all $p < 0.05$). The effect size estimate shows a large effect ($\eta^2 = .491$). The results are also shown in Figure 6.

RTLX Score

Finally, regarding the perceived cognitive load measured by the RTLX score, the *Avatar* led to the lowest perceived cognitive effort ($M = 29.95$, $SD = 16.33$), followed by the *2D-Image* ($M = 31.35$, $SD = 18.33$), the *3D-Scan* ($M = 31.98$, $SD = 16.62$), and the *Baseline* ($M = 34.11$, $SD = 19.03$). However, a repeated measures ANOVA test could not find a statistically significant difference between the conditions ($p > 0.05$). The effect size estimate shows a medium effect ($\eta^2 = .074$).

Likert Questionnaire Results

After each condition, we provided the participants with a questionnaire aiming to evaluate experiences on a 5-point Likert-scale (1: strongly disagree, 5: strongly agree). Figure 7 depicts the gathered data for all questions. We analyzed the data using Friedman's test. When significant effects where revealed, we used Wilcoxon's rank sum test with Bonferroni corrections for pairwise post-hoc analysis.

Q1: *"Was it easy for you to locate the passerby?"*
The analysis revealed a significant difference between the conditions ($\chi^2(3) = 14.74$, $p < .01$). Post-hoc tests showed significant lower ratings for the *2D-Image* compared to the three other conditions (all $p < .05$).

Q2: *"Were you able to locate the passerby accurately?"*
Friedman's test showed a significant effect between the conditions ($\chi^2(3) = 21.43$, $p < .001$). Post-hoc tests confirmed significantly lower ratings for the *2D-Image* condition compared to the other conditions ($p < .01$ for *Baseline*, $p < .05$ for *3D-Scan* and *Avatar*).

Q3: *"Was the position of the passerby clear to you?"*
The analysis showed a significant effect ($\chi^2(3) = 25.42$, $p < .001$) between the conditions. Post-hoc tests confirmed significantly lower ratings for the *2D-Image* condition compared to the *3D-Scan* ($p < .05$) and the *Baseline* ($p < .001$). We further found significantly lower ratings for the *Avatar* condition compared to *Baseline* ($p < .001$).

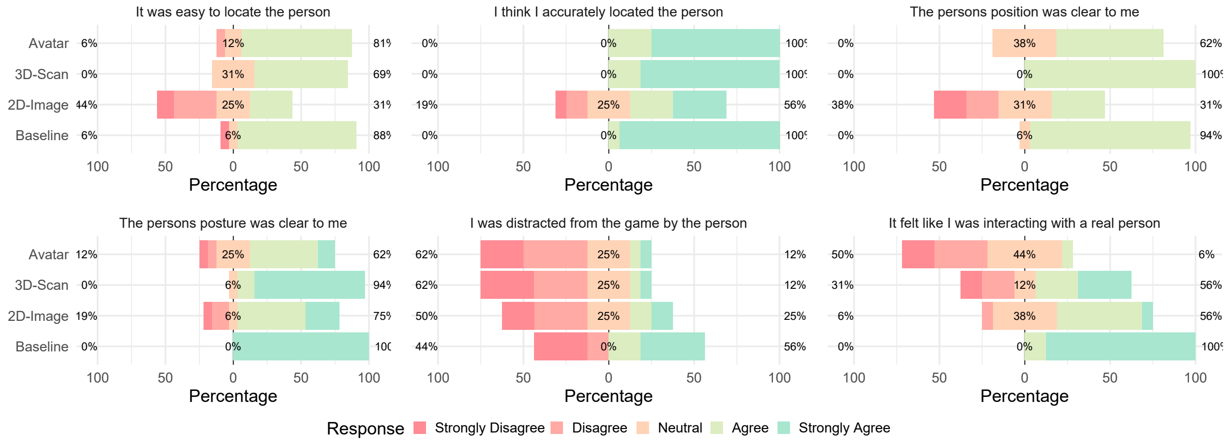

Figure 7. The answers to our questionnaire using a 5-point Likert scale. Red shades indicate disagreement, orange is neutral and green shades represent agreement. Furthermore, the bars are centered around the neutral point.

Q4: *"Was the posture of the passerby clear to you?"*
Again, the analysis showed a significant effect between the conditions ($\chi^2(3) = 29.7$, $p < .001$). Post-hoc tests confirmed significantly lower ratings for the *2D-Image* condition compared to *3D-Scan* ($p < .05$) and *Baseline* ($p < .001$). We further found significantly lower ratings for the *Avatar* condition compared to the *3D-Scan* ($p < .01$) and *Baseline* ($p < .001$) conditions.

Q5: *"Were you distracted by the person?"*
Despite slightly higher ratings for the *Baseline* condition, Friedman's test did not show significant effects ($\chi^2(3) = 4.94$, $p > .05$).

Q6: *"Did you feel like you were interacting with a real person?"*
Again, the analysis showed a significant effect ($\chi^2(3) = 30.54$, $p < .001$). As expected, the *Baseline* condition was rated significantly better than the other conditions ($p < .01$ compared to *3D-Scan*, $p < .001$ for the other conditions). Further, we found significantly lower ratings for the *Avatar* compared to the *2D-Image* ($p < .01$).

Qualitative Results

We collected qualitative feedback through semi-structured interviews at the end of the study. The interviews revealed that the participants generally liked the idea of passersby visualization in VEs. A sample comment given by a participant: *"Being able to see passersby in-game is a great idea to not get scared when someone is tapping on one's shoulder while I am immersed in VR"* (P3).

Regarding the *Avatar* representation, P12 liked *"that the passerby looked like a part of the game when visualized as a 3D-Model"*. However, another participant *"would have wished that the 3D-Model had arms instead of just floating hands"* (P10).

The participants expressed mixed feelings regarding the *2D-Image*. While one participant found the appearing *2D-Image* *"kind of scary"* (P7), another participant felt that *"Seeing my own hand on the video feed is irritating"* (P2). On the other hand, one participant liked that *"it is possible to change the direction from which I see the passerby, as the camera is head-mounted"* (P9).

Participants were impressed by the *3D-Scan*'s quality of detail. During the study, a participant shouted out: *"Wow, this looks surprisingly real"* (P8). Further, P12 commented that *"the 3D-Scan is so detailed, I can even see the motif on your t-shirt"*.

Regarding the *Baseline* condition, the participants disliked that *"[they] had to remove the headset every time [they] want to interact with persons outside VR"* (P3, P4, P16). One participant stated that taking off the HMD is cumbersome and that *"an extra hand to take off the headset is required"* (P15). Further, a participant who usually wears glasses was *"happy to have taken off glasses before this trial, as they would have dropped otherwise"* (P9).

DISCUSSION

In this section, we discuss the findings of our user study with regards to our previously introduced research questions:

H$_1$ Using higher details of the physical world in passersby visualization leads to a higher spatial awareness of the passersby.

H$_2$ Using higher details of the physical world in passersby visualization leads to a higher distraction from the played game.

Spatial Awareness

Considering the participants' accuracy d_{offset} for pointing at the passersby, we could show that the *2D-Image* scores significantly worse, compared to the *Avatar*. This contradicts the assumption that more details from the real world increase passersby detection accuracy. Considering the participants' perceived detection accuracy, we can see the *Baseline* and *3D-Scan* scoring the highest, with *2D-Image* scoring the lowest. The *Avatar* scores worse then the *3D-Scan* and *Baseline* in the questionnaire, but better in the objective measurements. This might be explained by the additional reference offered in the *3D-Scan* and *Baseline* approach, which gives the participants a point of reference. Being able to reference the seen passerby could increase the trust in their pointing accuracy.

Regarding the location time t_{locate}, we could show that both, the *3D-Scan* and the *Baseline* conditions, have a higher time compared to the *2D-Image* and the *Avatar*. While we expected the *Baseline* condition to result in a higher t_{locate}, the high t_{locate} for the *3D-Scan* condition is an interesting finding. A possible explanation for this is that participants using the *3D-Scan* tried to point more accurately and therefore required more time.

The Likert scale questions mainly showed that the *2D-Image* was perceived harder to locate passersby (*Q1*) and was perceived to be more inaccurate (*Q2*) by the participants. Also, the position (*Q3*) and the posture (*Q4*) of the passersby were not clear to the participants using the *2D-Image* compared to the other visualizations. Q6 revealed a significant difference in the users' perception of whether they are interacting with a real person considering the *2D-Image* and the *Avatar*. Although the *Avatar* led to a significantly lower offset in locating the passerby (d_{offset}), the participants still rated the *2D-Image* to feel more like interacting with a real person.

The *Avatar* and the *3D-Scan* were the most accurate showing that the additional information from the physical world from the *2D-Image* does not help the VR user in passersby location accuracy. Thus, we cannot support H_1 using the data gathered with our custom questionnaire. However, the *2D-Image* reduces the location time which is crucial to avoid injuries.

Distraction

Regarding H_2, which assumes conveying more information about the physical world leads to a higher distraction from the virtual environment, we could show that the results concerning the *2D-Image* do not support this theory. We measured the distraction through the participants' scores in the ball game, the perceived cognitive load using the RTLX score, and by asking the participants whether they were distracted by the passersby visualization (*Q5*). The idea behind using the score as a measure for distraction is that participants need time to get back into the game after locating the passerby. The more alienated they are from the task by the passerby, the longer this takes, and thus the score is lower. While *Q5* and the RTLX did not show any significant effect, the *2D-Image* and the *3D-Scan* led to the highest ball game scores. We assume that the rectangular base shape of the *2D-Image*, leading to a slightly larger visual addition into the VE made it easier to locate the overlay itself, leading to less time away from the game. This is further supported by the *2D-Image*'s t_{locate} being significantly lower than to the *Baseline* and *3D-Scan*. Though the data is not significant, we could observe a trend concerning the users' answers about the *3D-Scan*. *Q1* shows that it was easier to locate, had an easily interpretable posture *(Q4)* and was not distracting for the users *(Q5)*. This could mean that it was easier for the participant to passively locate the passerby, allowing them to keep some focus on the task. The similar score compared to *2D-Image* but with a higher t_{locate}, though not significant, supports this idea. To sum up, using objective measures, we could not show that conveying more details about the physical worlds leads to a higher distraction from the VE. Subjectively, however, participants felt distracted by the *2D-Image* and the *Baseline*, thus H_2 could be confirmed. The assumed degradation of performance could not be observed.

Limitations

It has to be mentioned that our study comes with a few minor limitations. Our study setup uses one Microsoft Kinect, which limits the area from which passersby can approach the tracking area to one side of the room. However, related approaches have shown that an environment mounted multi Kinect setup is feasible [20, 16]. Further, we only tested the behavior of the passersby visualization for a simple ball game task. We acknowledge that different VEs might yield different results.

CONCLUSION

In this paper, we compared three passersby visualizations that show users that are immersed in VR that a person is invading their tracking space through a user study with 16 participants. We compared the passersby visualizations (*Avatar*, *3D-Scan*, and *2D-Image*), to a *Baseline* condition without visualization. The study results revealed that, although users did not prefer a *2D-Image*, the *2D-Image* and the *Avatar* led to the fastest time for spotting the passersby. Further, we found that the *Avatar* visualization is most accurate for locating passersby and every approach, except for the *2D-Image* feel very accurate.

Interaction Suggestions

Our findings lead to multiple design recommendations for different use cases when the main use case is the interaction. We suggest using a visualization similar to the *Avatar* for physical interaction between the user in VR and the passersby, such as handing an object. The high accuracy measured in our experiment will be beneficial in such cases. For social interaction, such as talking with a passerby, using an approach with high perceived position and, more importantly, posture accuracy is suggested. Trading physical accuracy, which is not necessarily needed when talking to each other, for perceived accuracy, which makes for a more natural conversation can prove beneficial. We suggest using the *3D-Scan* approach if interaction quality is more important than instantly starting the conversation. If an immediate response is more important than the quality of the interaction, the *2D-Image* approach should be used.

Integration Suggestions

Using passersby integration with the goal of reducing distraction can best be achieved using either the *3D-Scan* or the *Avatar* approaches. For seamless integration into the VE an *Avatar* using the same visual style as the application is most promising. Participants did not feel distracted by our *Avatar* and also did not feel like interacting with a real human. This could be leveraged to camouflage passersby as non-player characters in a VR game. The *3D-Scan* approach is best suited for keeping users engaged in the VE while still presenting them with a real-world person, recognizable as such. In our experiment, participants reported being not distracted by the *3D-Scan*, with identical scores to the *Avatar*. The impression of interacting with a real person though was better than for the *Avatar*.

Future Work

In future work, we want to address the limitations by testing passerby visualizations in a setup using multiple Kinects and test a 360° coverage of detecting passersby. Further, we want to try the passerby visualizations in more complex VEs. Additionally, the effects concerning high t_{locate} and high score for the *3D-Scan* approach could be revised with additional experiments.

REFERENCES

[1] Alex Albert, Matthew R. Hallowell, Brian Kleiner, Ao Chen, and Mani Golparvar-Fard. 2014. Enhancing Construction Hazard Recognition with High-Fidelity Augmented Virtuality. *Journal of Construction Engineering and Management* 140, 7 (2014), 04014024. DOI: http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000860

[2] Pulkit Budhiraja, Rajinder Sodhi, Brett Jones, Kevin Karsch, Brian Bailey, and David Forsyth. 2015. Where's My Drink? Enabling Peripheral Real World Interactions While Using HMDs. *arXiv preprint arXiv:1502.04744* (2015). http://arxiv.org/abs/1502.04744

[3] Liwei Chan and Kouta Minamizawa. 2017. FrontFace: Facilitating Communication Between HMD Users and Outsiders Using Front-Facing-Screen HMDs. In *MobileHCI '17: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, 1–5. DOI: http://dx.doi.org/10.1145/3098279.3098548

[4] Lung-Pan Cheng, Patrick Lühne, Pedro Lopes, Christoph Sterz, and Patrick Baudisch. 2014. Haptic turk: a Motion Platform Based on People. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14 (CHI '14)*. ACM, New York, NY, USA, 3463–3472. DOI: http://dx.doi.org/10.1145/2556288.2557101

[5] Lung-Pan Cheng, Thijs Roumen, Hannes Rantzsch, Sven Köhler, Patrick Schmidt, Robert Kovacs, Johannes Jasper, Jonas Kemper, and Patrick Baudisch. 2015. TurkDeck: Physical Virtual Reality Based on People. In *ACM Symposium on User Interface Software & Technology*. ACM, New York, NY, USA, 417–426. DOI: http://dx.doi.org/10.1145/2807442.2807463

[6] Damien Clergeaud, Joan Sol Roo, Martin Hachet, and Pascal Guitton. 2017. Towards seamless interaction between physical and virtual locations for asymmetric collaboration. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17*. ACM, New York, NY, USA, 1–4. DOI: http://dx.doi.org/10.1145/3139131.3139165

[7] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017a. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4021—-4033. DOI: http://dx.doi.org/10.1145/3025453.3025683

[8] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2017b. FaceDisplay: Enabling Multi-User Interaction for Mobile Virtual Reality. In *Chi Ea '17 (CHI EA '17)*. ACM, New York, NY, USA, 369–372. DOI: http://dx.doi.org/10.1145/3173574.3173628

[9] Sebastian Günther, Florian Müller, Markus Funk, and Max Mühlhäuser. 2019. Slappyfications: Towards Ubiquitous Physical and Embodied Notifications. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. Glasgow, Scotland, UK. DOI: http://dx.doi.org/10.1145/3290607.3311780

[10] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. Sage Publications Sage CA: Los Angeles, CA, 904–908. DOI: http://dx.doi.org/10.1177/154193120605000909

[11] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16) (CHI '16)*. ACM, New York, NY, USA, 1957–1967. DOI: http://dx.doi.org/10.1145/2858036.2858134

[12] Charles E Hughes and Christopher B Stapleton. 2005. The Shared Imagination : Creative Collaboration in Augmented Virtuality Creative Collaboration and Mixed Reality Mixed Reality is not just Augmented Reality. *Human Computer Interaction International 2005 (HCII2005)* (2005), 22–27.

[13] H. Ibayashi, Y. Sugiura, D. Sakamoto, N. Miyata, M. Tada, T. Okuma, T. Kurata, M. Mochimaru, and T. Igarashi. 2015. Dollhouse VR: A multi-view, multi-user collaborative design workspace with VR technology. In *SIGGRAPH Asia 2015 Posters, SA 2015*. ACM, New York, NY, USA, 2–3. DOI: http://dx.doi.org/10.1145/2820926.2820948

[14] Akira Ishii, Masaya Tsuruta, Ippei Suzuki, Shuta Nakamae, Tatsuya Minagawa, Junichi Suzuki, and Yoichi Ochiai. 2017. ReverseCAVE: providing reverse perspectives for sharing VR experience.. In *ACM SIGGRAPH 2017 Posters on - SIGGRAPH '17*. ACM, New York, NY, USA, 28. DOI: http://dx.doi.org/10.1145/3102163.3102208

[15] H. Ishii and B. Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits, and atoms. In *Proceedings of the SIGCHI conference on Human factors in computing systems CHI 97*. ACM, New York, NY, USA, 234–241. DOI: http://dx.doi.org/10.1145/604045.604048

[16] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew D. Wilson, Eyal Ofek, Blair Macintyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. RoomAlive: Magical Experiences Enabled by Scalable, Adaptive Projector-Camera Units. In *ACM UIST*. ACM, New York, NY, USA, 637–644. DOI: http://dx.doi.org/10.1145/2642918.2647383

[17] Pascal Knierim, Thomas Kosch, Valentin Schwind, Markus Funk, Francisco Kiss, Stefan Schneegass, and Niels Henze. 2017. Tactile Drones - Providing Immersive Tactile Feedback in Virtual Reality through Quadcopters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17 (CHI EA '17)*. ACM, New York, NY, USA, 433–436. DOI: http://dx.doi.org/10.1145/3027063.3050426

[18] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 1–9. DOI: http://dx.doi.org/10.1145/3173574.3173919

[19] Eike Langbehn, Eva Harting, and Frank Steinicke. 2018. Shadow-Avatars: A Visualization Method to Avoid Collisions of Physically Co-Located Users in Room-Scale VR. (2018). http://basilic.informatik.uni-hamburg.de/Publications/2018/LHS18

[20] David Lindlbauer and Andy D Wilson. 2018. Remixed Reality: Manipulating Space and Time in Augmented Reality. In *Proc. of CHI (CHI '18)*. ACM, New York, NY, USA, 1–13. DOI: http://dx.doi.org/10.1145/3173574.3173703

[21] Christian Mai, Lukas Rambold, and Mohamed Khamis. 2017. TransparentHMD: revealing the HMD user's face to bystanders. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY, USA, 515–520. DOI: http://dx.doi.org/10.1145/3152832.3157813

[22] Sebastian Marwecki, Maximilian Brehm, Lukas Wagner, Lung-Pan Cheng, Florian 'Floyd' Mueller, and Patrick Baudisch. 2018. VirtualSpace - Overloading Physical Space with Multiple Virtual Reality Users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18 (CHI '18)*. ACM, New York, NY, USA, 1–10. DOI: http://dx.doi.org/10.1145/3173574.3173815

[23] M McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A dose of reality: overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd Annual*. ACM, New York, NY, USA, 2143–2152. DOI: http://dx.doi.org/10.1145/2702123.2702382<http://dx.doi.org/10.1145/2702123.2702382>)

[24] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. <title>Augmented reality: a class of displays on the reality-virtuality continuum</title>. In *Telemanipulator and Telepresence Technologies*, Hari Das (Ed.). SPIE, 282–292. DOI: http://dx.doi.org/10.1117/12.197321

[25] Perrine Paul, Oliver Fleig, and Pierre Jannin. 2005. Augmented virtuality based on stereoscopic reconstruction in multimodal image-guided neurosurgery: Methods and performance evaluation. *IEEE transactions on medical imaging* 24, 11 (2005), 1500–1511. DOI: http://dx.doi.org/10.1109/TMI.2005.857029

[26] Iana Podkosova and Hannes Kaufmann. 2017. Preventing Imminent Collisions between Co-Located Users in HMD-Based VR in Non-Shared Scenarios.

[27] Iana Podkosova and Hannes Kaufmann. 2018. Mutual collision avoidance during walking in real and collaborative virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games - I3D '18*. ACM Press, New York, New York, USA, 1–9. DOI: http://dx.doi.org/10.1145/3190834.3190845

[28] Nathan Rackliffe. 2005. *An Augmented Virtuality display for improving UAV usability*. Technical Report. BRIGHAM YOUNG UNIV PROVO UT. 1–3 pages. DOI:http://dx.doi.org/10.1590/0047-2085000000024

[29] H. Regenbrecht, T. Lum, P. Kohler, C. Ott, M. Wagner, W. Wilke, and E. Mueller. 2004. Using augmented virtuality for remote collaboration. *Presence: Teleoperators and Virtual Environments* 13, 3 (2004), 338–354. DOI: http://dx.doi.org/10.1162/1054746041422334

[30] Holger Regenbrecht, Katrin Meng, Arne Reepen, Stephan Beck, and Tobias Langlotz. 2017. Mixed Voxel Reality: Presence and Embodiment in Low Fidelity, Visually Coherent, Mixed Reality Environments. In *Proceedings of the International Symposium on Mixed and (ISMAR)*. IEEE, Piscataway, NJ, USA, 90–99. DOI: http://dx.doi.org/10.1109/ISMAR.2017.26

[31] H. Regenbrecht, C. Ott, M. Wagner, T. Lum, P. Kohler, W. Wilke, and E. Mueller. 2003. An augmented virtuality approach to 3D videoconferencing. In *Proceedings - 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2003*. IEEE, Piscataway, NJ, USA, 290–291. DOI: http://dx.doi.org/10.1109/ISMAR.2003.1240725

[32] Jun Rekimoto and Katashi Nagao. 1995. The World through the Computer: Computer Augmented Interaction with Real World Environments. In *Proc 8th Ann ACM Symp User Interface and Software Technology*, Vol. pages. ACM, New York, NY, USA, 29–36. DOI:http://dx.doi.org/10.1145/215585.215639

[33] Joan Sol Roo and Martin Hachet. 2017a. One Reality: Augmenting How the Physical World is Experienced by combining Multiple Mixed Reality Modalities. In *Proceedings of the 30th ACM User Interface Software and Technology Symposium*. ACM, New York, NY, USA, 787–795. DOI: http://dx.doi.org/10.1145/3126594.3126638

[34] Joan Sol Roo and Martin Hachet. 2017b. Towards a hybrid space combining Spatial Augmented Reality and virtual reality. In *2017 IEEE Symposium on 3D User Interfaces, 3DUI 2017 - Proceedings*. IEEE, Piscataway, NJ, USA, 195–198. DOI: http://dx.doi.org/10.1109/3DUI.2017.7893339

[35] Anthony Scavarelli and Robert J. Teather. 2017. VR Collide! Comparing Collision-Avoidance Methods Between Co-located Virtual Reality Users. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*. ACM Press, New York, New York, USA, 2915–2921. DOI: http://dx.doi.org/10.1145/3027063.3053180

[36] Marc Aurel Schnabel, Xiangyu Wang, Hartmut Seichter, and Tom Kvan. 2007. From Virtuality to Reality and Back. *Proceedings of the International Association of Societies of Design Research* 1 (2007), 15.

[37] Kristian T Simsarian, K.-P. Åkesson, and Karl-Petter Akesson. 1997. Windows on the world: An example of augmented virtuality. In *Interface 1997, Sixth International Conference Montpellier 1997: Man-machine interaction*. 68–71.

[38] Keng Hua Sing. 2016. Garden : A Mixed Reality Experience Combining Virtual Reality and 3D Reconstruction. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16 (CHI EA '16)*. ACM, New York, NY, USA, 180–183. DOI: http://dx.doi.org/10.1145/2851581.2890370

[39] Aaron Stafford and Wayne Piekarski. 2006. Implementation of god-like interaction techniques for supporting collaboration between outdoor AR and indoor tabletop users. In *Proceedings of the 5th IEEE and*. IEEE, Piscataway, NJ, USA, 165–172. DOI: http://dx.doi.org/10.1109/ISMAR.2006.297809

[40] Alladi Venkatesh, Erik Kruse, and Eric Chuan-fong Shih. 2003. The networked home : an analysis of current developments and future trends. *Cognition, Technology & Work* 5, 1 (2003), 23–32. DOI: http://dx.doi.org/10.1007/s10111-002-0113-8

[41] Xiangyu Wang and Rui Chen. 2008. An Empirical Study on Augmented Virtuality Space for Tele-Inspection of Built Environments. *Tsinghua Science and Technology* 13, SUPPL. 1 (2008), 286–291. DOI: http://dx.doi.org/10.1016/S1007-0214(08)70163-9