Presence and Discernability in Conventional and Non-Photorealistic Immersive Augmented Reality

William Steptoe*, Simon Julier, Anthony Steed

Department of Computer Science, University College London

Abstract

Non-photorealistic rendering (NPR) has been shown as a powerful way to enhance both visual coherence and immersion in augmented reality (AR). However, it has only been evaluated in idealized pre-rendered scenarios with handheld AR devices. In this paper we investigate the use of NPR in an immersive, stereoscopic, wide field-of-view head-mounted video see-through AR display. This is a demanding scenario, which introduces many real-world effects including latency, tracking failures, optical artifacts and mismatches in lighting. We present the *AR-Rift*, a low-cost video see-through AR system using an Oculus Rift and consumer webcams. We investigate the themes of consistency and immersion as measures of psychophysical non-mediation. An experiment measures discernability and presence in three visual modes: conventional (unprocessed video and graphics), stylized (edge-enhancement) and virtualized (edge-enhancement and color extraction). The stylized mode results in chance-level discernability judgments, indicating successful integration of virtual content to form a visually coherent scene. Conventional and virutalized rendering bias judgments towards correct or incorrect respectively. Presence as it may apply to immersive AR, and which, measured both behaviorally and subjectively, is seen to be similarly high over all three conditions.

Index Terms: H.5.1 [Information Interfaces and PresentationMultimedia Information Systems]: Artificial, augmented, and virtual realities— [I.3.7]: Computer Graphics—Three-Dimensional Graphics and RealismColor, shading, shadowing, and texture

1 Introduction

AR systems augment the real environment with virtual content in real-time [1]. A potential goal for AR is to integrate virtual and real imagery in such a way that they are visually indistinguishable [9]. One approach to this is to enhance graphical photorealism and illumination to match with the real world as closely as possible. This requires measurement of the illumination and material properties of the physical environment so that virtual objects can be shaded in such a way that they appear spatially- and temporally-consistent within the real-world viewport [3, 2]. Although this approach can produce highly convincing results, it is time consuming and cannot easily support dynamic changes in lighting or scene arrangement. An alternative approach is to change the view of the real world so that it appears closer to the computer-generated graphics. This can be achieved using non-photorealistic rendering, which applies artistic or illustrative filters to both the real-world imagery and the graphical content [4]. Unlike the appearance acquisition approach, which aims to enhance the photorealism of the virtual objects, stylized rendering diminishes the photorealism of both real and virtual content with the aim of homogenizing the overall appearance of the scene without the need for expensive measurement and precomputation.

*e-mail: w.steptoe@ucl.ac.uk

Compared to conventional real-time graphics techniques, NPR has been shown to make the task of discerning whether objects are real or virtual more difficult [5]. NPR applies image filters that silhouette edges, desaturate color or apply other non-photorealistic effects such as sketch-like, cartoonish or painterly rendering [7]. These disguise visual artifacts and inconsistencies commonly associated with real-time computer graphics such as aliasing, material appearance and imperfect illumination models. The filters are generally applied to the whole frame so both real and virtual content are transformed similarly. The influence of NPR on scene perception has only been investigated in idealized operating conditions in non-immersive AR with pre-rendered images and videos on standard desktop displays [5]. In head-mounted immersive AR, the issue of discernability becomes critical in forming a convincing experience.

To facilitate the experiment, we introduce a low-cost head-mounted video see-through immersive AR system, the AR-Rift. The system is based on an Oculus Rift Development Kit (DK1) and consumer cameras, coupled with an optical motion capture system for tracking as detailed in Section 3. It features wide field-of-view (FOV) stereoscopic rendering, providing a first-person embodied perspective supporting rich proprioception and exploration. However, the system also exhibits suboptimal characteristics common to operational AR systems such as different latency between video and tracking data, tracking failures, manually configured scene relighting, motion blur and optical distortions. Such inconsistencies may provide information in object discernability tasks and may distract users from feelings of presence within the AR environment.

Our study explores discernability and presence over three rendering modes in immersive head-mounted video see-through AR. Discernability refers to a user's ability to correctly distinguish between objects that are real (the object exists in the physical environment) and objects that are virtual. We investigate presence by assessing behavioral realism as is common in the virtual environments (VE) literature [13]. We define presence as applied to immersive AR as the perceptual illusion of non-mediation as is discussed further in Sections 2 & 6. This concept has not been rigorously applied to AR due to the demanding usability requirements for highly immersive systems. Presence will become a central aspect of immersive AR as the technology emerges at a consumer level. Thus our study also explores how its evaluation may be approached.

Three rendering modes provide an approach to explore the influence of NPR on discernability and presence in immersive AR. We term the modes *conventional*, *stylized*, and *virtualized* (illustrated in Figures 2–4). The conventional mode does not post-process the image, showing unaltered video feeds and uses standard real-time graphics algorithms. The stylized mode applies an edge-detection filter to silhouette edges of objects within the full image frame, which includes both video and graphics. The virtualized mode presents an extreme stylization by both silhouetting edges and removing color information. The three modes represent points on a spectrum that both transform and diminish photorealism. The appearance of the virtualized mode, which would generally be unsuitable for normal AR applications, provides an extreme condition to explore the influence of a highly non-photorealistic scene appearance on metrics relating to discernability and presence.

2 BACKGROUND

2.1 Presence as Non-mediation in Immersive AR

Presence is central to the study of immersive display technology. It can be defined as a user's psychological response to patterns of sensory stimuli, resulting in the psychological sensation of "being there" in the computer-generated space [15]. Slater suggests that that immersive virtual reality (VR) systems can be characterized by the sensorimotor contingencies that they support, referring to the actions that a user can carry out in order to perceive the VE [14]. Moving one's head and eyes to change gaze direction, or bending down in order to see underneath a virtual object are sensorimotor contingencies typically supported by an immersive VR system.

AR grounds interaction within the local physical environment, so any suitable definition of presence as applied to immersive AR must emphasize the *seamless* integration of virtual content with the real environment. In this context, Lombard and Ditton's definition that presence (in VEs) is "the perceptual illusion of non-mediation" [11] is salient to the concept in AR. In other words, the "machinery" of the AR experience should not be evident to the user. MacIntyre notes that two classes of factors can influence this sense of non-mediation: consistency of the content, and technical factors [12]. Consistency of content refers to the importance of the behavior of virtual objects remaining consistent with a user's expectations. This relates to Slater's theory of plausibility in VEs [14]. Technical factors refer to the immersive elements of the system and are connected with both graphical fidelity and relighting and tracking latency. A theory of presence as non-mediation as applied to immersive AR should consider both consistency and content. It would hypothesize that higher degrees of presence may be fostered when virtual and real content is integrated effectively to form a perceptually-unified environment. Presence in immersive AR, then, relates to fostering in users a perceptual state of non-mediation, which arises from a high level of technologically-facilitated immersion and environmental consistency, and which in turn may give rise to realistic behavior and response. These relate to the degree of presence that is being experienced. Similar to the approach in the VE literature, presence should be assessed through measurement of both quantitative data, for instance tracked body movement or physiological response, and qualitative insights from questionnaires and interviews.

Our second experimental task, described in Section 4.2 is inspired by acrophobia experiments studied in VR [13], and later trialled in immersive AR [6]. In those studies, participants report feeling frightened and some report vertigo when faced with the virtual pit. Some participants will not walk out onto the ledge that surrounds the pit and ask to stop the experiment. A few participants walk out over the pit as if it were covered by glass, but this requires conscious mustering of will [13]. Inferred from physical movement behavior and other measures, these findings indicate that participants behave towards the virtual pit as if it could be real. Our study proposes a more subtle and physically-plausible scenario suitable for investigating immersive AR by assessing walking behavior through a cluttered environment.

2.2 Non-Photorealistic Rendering

NPR applies transformation filters to modify the appearance of an image. Early examples were developed to create artistic effects such as cartoon-style rendering [10] or to transform the world to look more like a painting [8]. Fischer et al. demonstrated that NPR can be used to reduce the discernability of virtual objects in AR applications, thereby visually homogenizing a mixed reality scene [5]. The detail of the image was reduced using a color filter and edge detection to strongly highlight discontinuities. In a series of perceptual studies, it was demonstrated that the use of NPR reduced users' ability to correctly identify virtual and real content.

We are not aware of previous studies assessing NPR in discrimination tasks or presence in operational immersive AR. Rather, studies have used non-interactive stimuli, such as pre-rendered videos or still images. Real-time performance has been cited as the reason for this. Fischer et al. reported that a system was developed using ARToolKit and is capable of real-time operation [5]. However, no systematic user studies were reported. In this paper we present an immersive system that performs to real-time requirements and use it to investigate accuracy of discernability when judging real and virtual objects and also on sense of presence and embodiment.

Figure 1: AR-Rift, consisting of an Oculus Rift DK1 and mounted modified Logitech C310 webcams.

3 AR-RIFT

3.1 System Overview

The AR-Rift, shown in Figure 1, is an immersive head-mounted video see-through AR display comprising of low-cost, commercially available components. A comprehensive description of the methodology and process of designing, building and calibrating the display, including limitations can be found in [17]. The display consists of an Oculus Rift DK1 VR head-mounted display (HMD) mounted with two modified Logitech C310 webcams forming a wide FOV stereoscopic camera. The Rift DK1 features a 7" RGB LCD panel with a resolution of 1280×800 and a refresh rate of 60Hz. Side-by-side stereoscopy is achieved by rendering left and right eye images to each half of the panel, resulting in 640(h)$\times$800(v) pixels per eye. The $2\times$ aspheric lenses provide a FOV approximately 90° horizontal and 110° vertical depending on eye relief and inter-pupillary distance (IPD). Each lens focuses on one horizontal half of the panel at a perceived distance of approximately 3.66m. The stereo pair features approximately a 95% horizontal overlap.

The optical and technical stereo camera specifications are approximately in parity with the those of the Rift DK1. The sensor resolution is 1280×960, exceeding the Rift DK1's per-eye resolution. The C310's stock lenses were replaced with lenses providing a 120° FOV, slightly exceeding the Rift's 110° on the vertical dimension. The C310 has a stated specification of 30 FPS, but it is able to operate reliably up to 45 FPS. As we are aiming for a low-cost implementation, we have not considered professional machine vision cameras, so our system does not allow for camera synchronization or other desirable features. We opted to use low-cost components to enable wider reimplementation and experimental replicability.

Reflective markers positioned on the AR-Rift (Figure 1) are tracked by a NaturalPoint OptiTrack optical motion capture system with twelve Flex 3 cameras operating at 100 FPS. The working tracking volume is approximately 80% of the $6\times4\times3$m lab. The Motive v1.5 software computes six degrees-of-freedom (6DoF) position and orientation of the HMD, which is transmitted via UDP to our software framework developed using Unity 4. Versions of the Oculus Rift beyond DK1 will include positional tracking capability, thereby potentially eliminating the need for the external tracking system used in our setup.

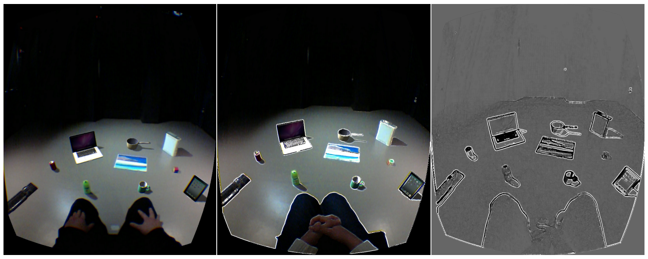


Figure 2: Point-of-view screen-captures during the object discernability task in each condition. Left to right: conventional, stylized, virtualized. Real objects: keyboard, shampoo, poster, Xbox, Rubik's Cube. Virtual objects: Coke can, Macbook, pan, mug, iPad.

Figure 3: Close-up of real and virtual pan object in each condition. Left to right: conventional, stylized, virtualized. Real versions are shown to the right of the virtual versions in each image.

3.2 Performance

The calibration process rectifies the video feeds and aligns the stereo image to the 6DoF tracking pose. This results in visual collocation of physical and virtual objects over the full FOV. The application performs consistently at >120 FPS on an Intel i7 CPU, 16GB RAM and an nVidia GTX 680 running Windows 7. The machine features one USB 3.0 and two USB 2.0 controllers, allowing both the OptiTrack motion capture system and the Unity application with the two camera feeds to operate from the same machine.

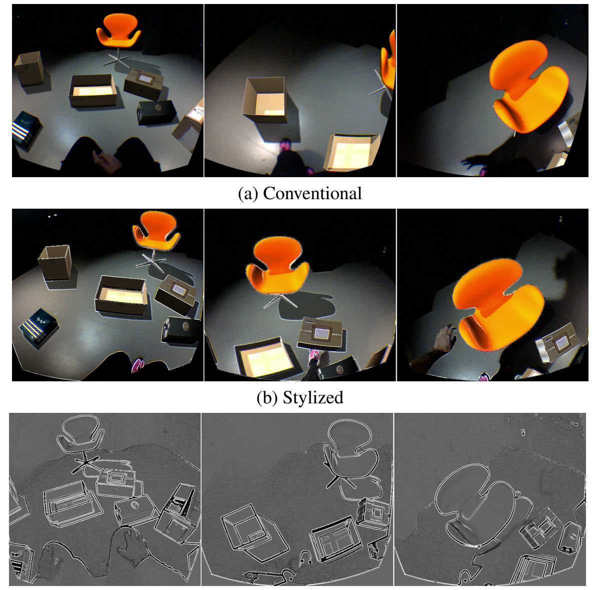
The latency of camera images and tracking data are 100ms and 60ms respectively. This results in visual discrepancy between the motion of the physical space as seen in the video and that of virtual objects. In cases of rapid head movement, this can also result in stationary virtual and real objects being seen to overlap one another. It is possible to delay the tracking data to synchronize with the camera images. This will be considered in future work. For this experiment, however, we investigate the impact of rendering mode in an imperfect setup common to operational AR systems.

4 Experiment

4.1 Rendering Condition Design

Three rendering modes were designed to explore the influence of appearance on discernability and presence in the AR-Rift. The modes are illustrated in Figures 2–4. Conventional rendering presents unmodified video and standard graphical rendering techniques to display the AR scene. The stylized mode uses a Sobel edge filter to highlight horizontal and vertical edges [16]. This edge detection is performed on the full image frame and modifies the appearance of the scene in a subtle but noticeable way. The virtualized mode extends the stylized shader to desaturate the color of the image using Cg's *saturate()* function. Edges are exaggerated and clarity of lighting and shading from both real virtual light sources are diminished.

The three rendering modes were designed to represent positions on a scale of photorealism. The modes both transform and diminish the appearance of the scene. The conventional mode puts virtual content at a disadvantage in terms of its perceptual marriage with the surrounding physical environment. Differences in the lighting and shading appearance of virtual and real objects as well as rendering artifacts such as aliasing are obviously apparent. The stylized mode aims to unify the appearance of virtual and real objects by disguising



(a) Conventional

(b) Stylized

(c) Virtualized

Figure 4: Sequences during the second experimental task in each condition. The task is to stand up and walk to the chair. Boxes appear to be scattered around the environment. The chair and all boxes are virtual. Figure 5 shows paths taken in each condition.

such inconsistencies through embossed edges. The environment is still salient as the physical reality, but the virtuality of objects is less visually obvious. Finally, the virtualized mode provides an inverse to the conventional mode by diminishing the environment to an extent that it could be perceived as virtual. In this mode, real objects may appear less seamless in the environment than the virtual objects.

4.2 Task Summary and Data Collection

Our experiment features two tasks exploring discernability and presence over the three rendering conditions. The first task relates to discernability, where participants are required to judge each of ten objects as real or virtual. Five objects were real and five were virtual, with object virtuality balanced over participants and conditions. Objects were common to home and office environments: computer keyboard, Coke can, Macbook Pro, bottle of shampoo, printed A3-sized poster, 20cm saucepan, mug, Xbox 360, Rubik's Cube, iPad. Virtual and physical counterparts of objects were sourced. Virtual versions were sourced from Trimble 3D Warehouse and are representative of mid-to-high quality modeling and shading. The VE was manually relit so virtual luminance and shadows were qualitatively similar to that of the real environment when viewed in the AR-Rift. We measured binary response accuracy for each of the ten objects.

The second task assessed users' sense of presence by measuring behavior relating to the extent to which the mixed reality environment is acted upon as the salient physical reality. Participants are given the task of walking from their current seated position to sit on a chair located around 2m opposite them. Between their starting position and target chair, however, there appear to be a number of medium-large cardboard boxes scattered around the floor. These boxes and also the target chair are virtual: the 'stage' area is void of physical objects. The scene is illustrated in Figure 4. We capture behavior in this task by tracking foot position using a pair of shoes fitted with reflective markers at 60 Hz using the OptiTrack system.

Following both tasks, participants completed the questionnaire shown in Table 2 relating to the experience in terms of visual quality, presence and embodiment, and system usability. Questions were answered on a 1–5 Likert scale with either Virtual–Real or No–Yes representing the ends of the scale.

Table 1: Discernability accuracy for each object in each condition and when real and virtual. Normalized to 1.

	Keyboard	Coke Can	Macbook	Shampoo	Poster	Pan	Mug	Xbox	Rubik's	iPad	**Overall**
Conventional	0.7	0.8	0.7	0.7	0.7	0.7	0.8	0.7	0.8	0.7	0.73
Stylized	0.4	0.5	0.5	0.7	0.4	0.5	0.7	0.6	0.7	0.6	0.56
Virtualized	0.3	0.4	0.3	0.7	0.3	0.4	0.5	0.3	0.2	0.4	0.38
When Real	0.27	0.73	0.4	0.6	0.73	0.6	0.73	0.6	0.73	0.53	0.59
When Virtual	0.67	0.4	0.6	0.8	0.2	0.47	0.6	0.47	0.4	0.6	0.52

4.3 Procedure

Thirty participants (14 female) with normal or corrected-to-normal vision and no previous immersive HMD experience were recruited from UCL's student and staff population. The experiment used a between-subjects design with the three rendering modes as the independent variable, resulting in ten participants per condition. Participants were not told about the rendering mode being a manipulation. Following reading an instruction handout and providing consent, participants were seated in a chair at the center of the 6×4×3m lab. The floor was initially empty of physical objects. The participant donned the motion capture shoes followed by the AR-Rift, which was adjusted for comfort. The simulation application in the specific rendering condition was started and the AR-Rift turned on. The participant acclimatized to the experience of seeing through the cameras, observing their surroundings and own body. The experimenter then instructed the participant to stand and walk around the lab before sitting back in the chair. None of the 30 participants reported unsteadiness or sickness during this period and were all happy to continue with the experiment.

The experimenter turned off the AR-Rift to act as blindfold and placed noise-canceling headphones on the participant so they could not observe arrangement of the stage area. The experimenter placed five physical objects at predefined marked positions on the floor (the remaining five would be virtual). The headphones were removed and the AR-Rift turned back on. The participant saw the ten objects positioned around them as illustrated in Figure 2. Starting from the left-most object (keyboard), the experimenter asked the participant to judge if they thought each object was real or virtual. There was no time-limit imposed on the judgements and the seated participant was free to use any combination of visual and perceptual cues.

Following the object discernability task, the display was again turned off and headphones re-placed. The experimenter cleared the stage area and then waited for two minutes (a plausible amount of time needed to arrange the chair and boxes had they been real) before removing the headphones and turning the display back on. The participant was confronted with the second environment as shown in Figure 4, and when ready, stood and made their way over to the chair. The experimenter stopped the participant in the event they actually tried to sit on the virtual chair. The AR-Rift and motion-capture shoes were removed. The participant completed the post-experimental questionnaire and short interview with the experimenter. The experimental procedure was around 20 minutes.

5 Results

5.1 Discernability Accuracy

The upper half of Table 1 shows discernability accuracy for each object and overall mean in each of the three rendering conditions. The overall mean accuracy for conventional rendering is 73%, for stylized it is 56%, and for virtualized rendering it is 38%. We calculated a Pearson Chi-Square test (asymptotic 2-sided) using SPSS with frequency of correctness for each rendering mode. A significant main effect of rendering mode was found between conditions ($p < 0.001$), with inter-group significance between all modes: conventional and stylized ($p = 0.012$), conventional and virtualized ($p < 0.001$) and stylized and virtualized ($p = 0.011$). No main effect between object type was found ($p = 0.669$).

While this analysis shows no difference in accuracy between object type, it is interesting to compare accuracy in terms of virtuality.

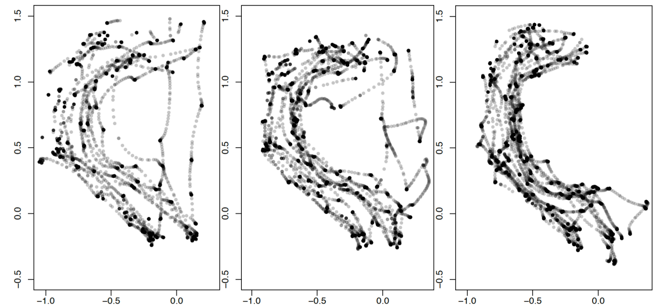


Figure 5: Paths taken from starting position (*x=0, y=-0.25*) to target position (*x=-0.25, y=1.25*) in each condition. Left to right: conventional, stylized, virtualized. Points represent left and right foot positions visualized at 60Hz. More dense areas indicate slower movement, particularly grounded footsteps. Units are in meters.

While the real and virtual pairs represent the same object and are visually similar, they nevertheless have qualitatively different appearance. The lower half of Table 1 shows discernability accuracy for each object's virtual and real counterparts and overall mean for all conditions. The overall mean accuracy is 59% when the objects were real and 52% when virtual. A Pearson Chi-Square test does not reveal a significant overall main effect ($p = 0.393$). Focusing on each object in turn, differences are revealed between real and virtual counterparts only for the keyboard ($p = 0.028$) and the poster ($p = 0.002$). There are near-significant differences between real and virtual versions of the Coke can and the Rubik's Cube ($p = 0.069$ for both). There was no bias towards more objects being identified as either virtual or real over the three visualization modes, with 49%, 44% and 46% of objects judged to be real in conventional, stylized and virtualized modes respectively.

5.2 Movement Paths

Figure 5 shows the paths of all participants during the second task when moving through the environment illustrated in Figure 4 in each condition. The target chair and all boxes were virtual and the stage area was clear of physical objects. Each point in Figure 5 represents a left or right foot position, recorded at 60Hz. Points are semi-transparent. Participants started at their seated position at *x=0,y=-0.25*, where the left and right feet are clearly distinguishable. They were asked to walk from this position to the target chair located at *x=-0.25,y=1.25*. Slower movements, particularly grounded footsteps are visible as denser and darker areas on the plots. Participants in the conventional rendering condition took fewer steps and moved more quickly than those in the stylized and particularly the virtualized conditions. The chosen route to the target chair was similar over all conditions, with most participants taking the clearest path through the boxes that curves out to the left.

5.3 Questionnaire

Questions 1–7 use the Virtual–Real scale and Questions 8–17 use the No–Yes scale. Questions are listed in Table 2. Medians and interquartile ranges (IQR) are shown. A one-way ANOVA reveals significant difference in five of the questions. Q2 relates to if the body looked real or virtual. The overall significance is $p = 0.018$,

Table 2: Post-experimental questionnaire relating to visual fidelity, presence, embodiment, and system usability. Responses were recorded on a *1→5* Likert scale, with either *Virtual→Real* or *No→Yes* anchoring the low and high ends of the scale respectively. Responses showing medians, interquartile ranges and ANOVA tests for each condition.

			Conventional		Stylized		Virtualized		
			Median	IQR	**Median**	IQR	**Median**	IQR	*P* value
	Q1	I had the feeling that my body was...	**5**	0	**4.5**	2	**4.5**	1.75	0.166
	Q2	My body looked...	**5**	0.75	**3**	1.75	**3**	1.5	*0.018*
1: Virtual	Q3	I had the feeling that the room was...	**5**	0	**3.5**	2.75	**4**	1.75	*0.013*
↓	Q4	The room looked...	**5**	0	**2**	2.75	**3**	0	*<0.001*
5: Real	Q5	I had the feeling that the boxes were...	**2**	1.75	**3**	1	**4**	1.5	*0.01*
	Q6	I had the feeling that the chair (target) was...	**1**	1	**2.5**	2.75	**5**	0.75	*<0.001*
	Q7	Overall, the experience felt...	**3**	1	**3**	1.75	**4**	1	0.29
	Q8	I had the feeling that I was in the room	**5**	0	**5**	0	**5**	1	0.572
	Q9	I had the feeling I was surrounded by objects	**4**	1	**4**	1	**5**	1	0.417
	Q10	I was worried about tripping over the boxes	**3**	2.75	**2.5**	1	**4**	1.75	0.102
1: No	Q11	My body appeared to be the right size	**5**	0.75	**5**	1	**5**	0	0.571
↓	Q12	My body appeared to be in the right place	**4.5**	1	**4**	1.75	**5**	0	0.249
5: Yes	Q13	I felt dissociated from my body	**2.5**	2.75	**1**	1.75	**2**	0	0.299
	Q14	I felt less able than normal	**4.5**	1	**3.5**	2	**3.5**	1.75	0.123
	Q15	I felt vulnerable	**3**	1.75	**3.5**	1.75	**2**	3	0.643
	Q16	I felt disorientated at some points	**3**	1.75	**3.5**	2	**2**	1.5	0.217
	Q17	I felt sick at some points	**1.5**	1	**2.5**	2	**1**	1.5	0.361

with post-hoc Tukey tests indicating differences between the higher-rated conventional mode and both stylized and virtualized conditions (both $p = 0.035$). Q3 relates to if the room felt real or virtual. Overall a significance of $p = 0.013$ was found, with significance between conventional and stylized rendering ($p = 0.01$). Q4 asked how the room looked on a scale from virtual to real. Overall significance is $p < 0.001$ with significance lying between conventional and both stylized ($p < 0.001$) and virtualized ($p = 0.004$) conditions. Several outliers are found in both conventional and virtualized conditions, however, indicating a lack of consensus. Q5 assessed if boxes in the second task were felt to be real or virtual, with a main effect of $p = 0.01$ and group significance lying between virtualized and both conventional ($p = 0.026$) and stylized ($p = 0.016$) rendering. The final significant main effect was found with Q6 ($p < 0.001$), assessing the feeling of the target chair being real or virtual, with significances between virtualized and both conventional ($p < 0.001$) and stylized ($p = 0.002$) conditions.

6 Discussion

6.1 Discernability

Results from the discernability task demonstrate the effectiveness of the two non-photorealistic rendering modes to visually merge virtual objects within a real environment. The stylized mode was particularly successful, with a 56% mean accuracy near chance, which is both significantly lower than conventional (73%) and higher than virtualized (38%) modes. This indicates participants in the stylized condition were unable to discriminate physical and virtual objects. Judgments in the conventional mode had a mean of 73% accuracy, significantly higher than both other conditions. This suggests that the visual difference between real and virtual objects provided enough information to make more accurate judgments. An interesting phenomenon emerges in the virtualized condition, where the visual appearance of both real and virtual objects are transformed to such an extent that accuracy is reduced below chance to a mean of 38%. This indicates that the visual characteristics of the virtualized condition are either inadequate or are misleading to the judgment process. These results support the intended effect of the rendering modes.

Judgment accuracy in the stylized mode suggests that NPR can be effective in unifying the appearance of an environment in immersive AR despite the range of perceptual sensorimotor cues afforded by the system. Fischer et al.'s 2006 study reported 94% and 69% accuracy for conventional and stylized modes respectively [5]. Our results are generally lower, likely due to several differences in both technical and experimental setups as well using much higher-fidelity 3D models and shading. In particular, our system and experimental design allow for the use of latency and parallax during judgment. This refers to the strategy of using fast head movements or observation of optical distortions to determine variations between real and virtual objects. This information could then be used to judge objects according to one of two groups rather than purely on visual appearance. From post-experimental interviews, this strategy was used heavily by participants in both the stylized (7/10) and virtualized (8/10) conditions. The few (3/10) using this approach in the conventional condition indicates that visual appearance was sufficient to perform the task confidently. Lower accuracy results in the virtualized mode suggests the potential for these cues to mislead. This is elucidated by the number of participants who achieved "very high" (≥80%) or "very low" (≤20%) overall accuracies. A high 6/10 participants in the conventional mode match the former category, and inversely, 4/10 in the virtualized mode match the latter, while few participants in the stylized mode (2/10 and 1/10) fall into the categories respectively. This provides evidence for the virtualized mode reducing the effectiveness of cues otherwise used successfully in the conventional mode. The stylized mode achieves a balance in forming a unified scene despite the availability of these perceptual cues.

Certain objects were more difficult to judge when real or virtual (see Table 1). Only 27% of judgments were correct for the real keyboard. This was the first object to be judged, so inexperience may be a factor. However, when judging the virtual poster, only 20% of judgments were correct. The flatness of both the keyboard and poster results in shadow information neither being present or expected. Participants tended towards judging both real and virtual versions of the objects as real. Post-experimental interviews mentioned shadows as an important cue relied on by participants in all conditions. Other objects showing near-significant disparity were the Coke can and the Rubik's Cube, both showing bias towards being judged as real. The small size of these objects, coupled with the relatively low resolution of the Rift DK1 display may be influential. Also of note is the shampoo object, which was judged with 70% accuracy in all conditions. Its height and proximity to the participant enabled detailed scrutiny, including viewing from wide angles and leaning forwards to look down on the object. This provides a successful example of using the immersive system's supported sensorimotor contingencies to observe a scene beyond purely visual information.

6.2 Presence

Our second task provided a scenario to study presence in immersive AR by measuring ambulatory behavior. The motion tracking paths shown in Figure 5 indicate that participants navigated the environment similarly in all conditions by walking around the (virtual) boxes scattered around the environment to reach the (also virtual) chair, rather than walking directly through the boxes to reach the chair. This is the approach that we would expect if the task were to be presented in reality, and demonstrates that participants are acting on the environment as the salient version of reality. One participant in the conventional mode walked directly through the boxes (Figure 5(a)). One participant in the stylized mode also appears to do this (Figure 5(b)), however, this participant is carefully stepping between and over the boxes. All participants in the virtualized mode walked around the clearest path, avoiding the boxes. Responses to Q5 and Q6 shown in Table 2 show how the perceived virtuality of the boxes and chair varies significantly with visual appearance. Participants in the conventional mode generally believed the objects were virtual, and those in the virtualized mode to be real. The motion plots, however, show similar paths around the boxes when walking.

This behavior is comparable to that observed in the virtual pit experiments [13], with participants avoiding situations that would (if real) put themselves in danger. Our experiment measures response to a less extreme scenario, with motion plots demonstrating an aversion to the physically-impossible (placing a foot within a solid object) or unusual action of walking through objects rather than around them. Behavior is similar in all conditions, indicating a high degree of embodiment and presence in the AR environment despite the varying degree of perceived virtuality. Subjective responses to questions Q1–Q4 and Q7–Q9 support this conclusion. These findings support our definition of presence in immersive AR being the perceptual state of non-mediation arising from technologically-facilitated immersion and observed environmental consistency, and which in turn gives rise to behavioral realism.

While motion paths are similar between conditions, the characteristics of the ambulation varies with photorealism. Movements were most careful in the virtualized condition, with slower overall speed and an increased number of smaller steps than those in stylized and particularly conventional modes. This is illustrated in Figure 5 by less path variation and denser plots. The tendency for more careful movement in the virtualized condition is likely due to the diminished visual realism of the physical environment, with features and shadows being difficult to distinguish.

6.3 Usability

Questions Q11–Q17 relate to usability of the AR-Rift. No significant differences were found between conditions. Responses to Q11–Q13 show a high sense of embodiment when using the system in terms of participants' bodies feeling a normal size and correctly positioned. Q14 shows that participants felt relatively less able than normal, expected due to reduced visual acuity and peripheral vision and added latency. Q15 and Q16 indicate low-medium levels of feeling vulnerable and disorientated, and Q17 shows that a low level of motion sickness was experienced. These measures indicate relatively high system usability performance and is encouraging for future low-cost immersive AR systems.

7 Conclusions

Immersive AR embeds virtual content in the physical world and presents this mixed environment with a first-person perspective featuring head tracking, a wide FOV and stereoscopic rendering. This embodied scenario allows rich proprioceptive exploration similar to immersive VR. We presented the AR-Rift, a low-cost immersive video see-through AR HMD based on the Oculus Rift DK1 and consumer cameras. As this technology emerges at a consumer level, the concept of presence as the perceptual illusion of non-mediation will become salient. An approach to achieving this illusion is NPR and our user study demonstrated its effectiveness towards achieving perceptual unification of a scene so as to make it difficult to discern between real and virtual objects. Our presence scenario based on seminal studies in the VE literature observes similar behavior over the three rendering conditions, indicating high degrees of presence and embodiment. System usability results were encouraging for the future of low-cost immersive AR systems. Future work will explore presence and interaction in immersive AR and improve technical shortfalls of the current system.

Acknowledgements

Supported by the FP7 EU collaborative project BEAMING (248620) and the EPSRC-funded project VisAge.

References

[1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34–47, 2001.

[2] D. A. Calian, K. Mitchell, D. Nowrouzezahrai, and J. Kautz. The shading probe: Fast appearance acquisition for mobile ar. In *SIGGRAPH Asia 2013 Technical Briefs*, SA '13, pages 20:1–20:4, New York, NY, USA, 2013. ACM.

[3] P. Debevec. Image-based lighting. *IEEE Computer Graphics and Applications*, 22(2):26–34, 2002.

[4] J. Fischer, D. Bartz, and W. Straber. Stylized augmented reality for improved immersion. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, pages 195–202. IEEE, 2005.

[5] J. Fischer, D. Cunningham, D. Bartz, C. Wallraven, H. Bülthoff, and W. Straßer. Measuring the discernability of virtual objects in conventional and stylized augmented reality. In *Proceedings of the 12th Eurographics conference on Virtual Environments*, pages 53–61. Eurographics Association, 2006.

[6] M. Gandy, R. Catrambone, B. MacIntyre, C. Alvarez, E. Eiriksdottir, M. Hilimire, B. Davidson, and A. C. McLaughlin. Experiences with an ar evaluation test bed: Presence, performance, and physiological measurement. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 127–136. IEEE, 2010.

[7] M. Haller. Photorealism or/and non-photorealism in augmented reality. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry*, pages 189–196. ACM, 2004.

[8] A. Hertzmann. Painterly rendering with curved brush strokes of multiple sizes. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 453–460. ACM, 1998.

[9] G. Klein and D. Murray. Compositing for small cameras. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 57–60. IEEE Computer Society, 2008.

[10] J. Lander and G. CONTENT. Shades of disney: Opaquing a 3d world. *Game Developer Magazine*, 7(3):15–20, 2000.

[11] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2):0–0, 1997.

[12] B. MacIntyre, J. D. Bolter, and M. Gandy. Presence and the aura of meaningful places. *Presence: Teleoperators and Virtual Environments*, 6(2):197–206, 2004.

[13] M. Meehan, S. Razzaque, B. Insko, M. Whitton, and F. P. Brooks Jr. Review of four studies on the use of physiological reaction as a measure of presence in stressfulvirtual environments. *Applied psychophysiology and biofeedback*, 30(3):239–258, 2005.

[14] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549, 2009.

[15] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3(2):130–144, 1994.

[16] I. Sobel. An isotropic $3\times$ 3 image gradient operator. *Machine Vision for three-demensional Sciences*, 1990.

[17] W. Steptoe. AR-Rift, http://willsteptoe.com/post/66968953089/ar-rift-part-1. Website, 2013.