Gaze Analysis of Avatar-based Navigation with Different Perspectives in 3D Virtual Space

Jooyeon Lee
Information & Interaction Design,
Underwood International College
Yonsei University
406-840 Incheon, Korea
jooyeonlee@yonsei.ac.kr

Manri Cheon, Seong-Eun Moon, Jong-Seok Lee
School of Integrated Technology &
Yonsei Institute of Convergence Technology
Yonsei University
406-840 Incheon, Korea
{manri.cheon, se.moon, jong-seok.lee}@yonsei.ac.kr

ABSTRACT
This paper explores the relations between perspectives of navigation and visual perception in 3D virtual space, by analyzing avatar-based navigation with the eye-gaze data. We examine how different perspectives and types of avatars affect the users' scopes of visual perception within 3D virtual environments. Throughout this research, we attempt to draw possible connections between the perspectives and cognitive patterns of visual perception. We propose that manipulating perspectives of avatars or those of users has immediate effects on the users' scopes of visual perception and patterns of visual attention.

Author Keywords
Navigation; Virtual Space; Perspective; Human Agent Interaction; Eye-tracking; Gaze Analysis.

ACM Classification Keywords
H.1.2. User/Machine Systems: Human Factors; H.5.1 Multimedia Information Systems: Artificial, augmented, and virtual realities

INTRODUCTION
As the advent of augmented reality (AR) and virtual reality (VR) technologies has opened up the vast opportunities for 3D technologies, digital devices and applications that help people to navigate and search information in 3D digital environments now draw huge attentions. With depth information, 3D space can certainly display richer contexts and elements than 2D space can, and a well-designed 3D application can create a great sense of presence and agency, delivering fascinating user experience with sophisticated contents that 2D space cannot afford to present.

Unfortunately, the human cognitive system has limited capacity, and excessive information presented in 3D space can overburden the cognitive system and hinder information retrieval [1]. For those reasons, it is important to keep the balance between the availability of information and the ability of human to process information [5]. In order to make use of 3D space, most of the researches have focused on placing huge amount of data in compact 3D space [1] or pushing the limit of human ability to augment cognitive capacities [6]. However, there have been few studies on limiting the scope of perception to enhance cognitive ability to help people concentrate on what is important, while navigating through such vast informative space in 3D environments with an agent or an avatar. A proper replacement of an avatar within 3D space help users to modify their viewpoints of the 3D world, and such effects can be useful under certain circumstances. In this paper we discuss how avatar-based navigations with difference perspectives in 3D space provide a guidance of what to look at, and analyze how the scopes of the visual perception are projected onto the 3D environments with respect to the perspectives of avatars and users' frame of reference.

HAI 2015, October 21-24, 2015, Daegu, Kyungpook, Republic of Korea

DOI: http://dx.doi.org/10.1145/2814940.2814981

STUDY DESIGN
As our objective is to study the changes in visual perception according to the perspectives of avatars, we prepared video clips with two sets of 3D environments, a road and an art gallery, in both first-person (1PP) and third-person (3PP) perspectives. The video clips were recorded with a simple scenario. Each sequence ends as the avatar or the viewpoint of the camera arrives at the other side of the wall or the road. As in Figure 1, eight sequences were shown to the subjects: (a) a 1PP sequence with no avatar on a road, (b) a 3PP sequence with a car closer to the viewpoint of the subjects, (c) a 3PP sequence with a car farther from the viewpoint of the subjects, (d) a 1PP sequence with no avatar in an art gallery, (e) a 1PP sequence with a tool (gun) as the extended part of subjects' virtual arms, (f) a 3PP sequence with a human avatar, (g) a 3PP sequence with a robot avatar closer to the viewpoint of the subjects, and lastly (h) a 3PP sequence with a robot avatar farther from the viewpoint of the subjects. The video clips of Set 1 and Set 2 have lengths of 5 seconds and 25 seconds, respectively. All of them have a frame rate of 24 Hz.

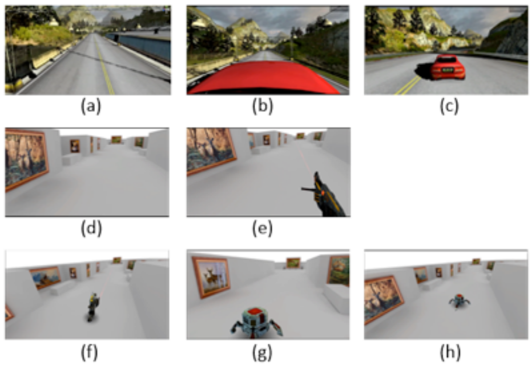


Figure 1. Two sets of video clips: Set 1 – a road environment with (a) no avatar, (b) 3PP car-near, and (c) 3PP car-far; Set 2 – an art gallery environment with (d) no avatar, (e) 1PP human, (f) 3PP human, (g) 3PP robot-near, and (h) 3PP robot-far.

Participants

Five healthy volunteer subjects with normal visual acuity (two male and three female; age=20-24 years; M=21.5 years; SD=1.48) participated in the experiment, which took an hour. Before the experiment, all participants were fully informed about the purpose of the study and the procedure of the experiment.

Experiment Procedure

Eye tracking data were recorded by Smart Eye Pro 5.9 system with three cameras, which have a sampling rate of 60 Hz. The video clips were displayed on a computer monitor (21x12 inches, 1980x1080 pixels), which was placed 22-27 inches apart from each participant. After careful camera and gaze calibrations, participants were asked to watch two sets of video clips with a 10-minute intermission between them and several 10-second breaks in-between consecutive sequences. The order of the video shown was: Set 1 (a)-(c)-(b), and Set 2 (d)-(h)-(g)-(e)-(f).

RESULTS

We down-sampled the eye-gaze data from 60 Hz to 24 Hz to match the frame rate of the video clips. In order to compare each sequence's overall frequency of ranges of saccades, we calculated the distances between two eye-gaze points on consecutive frames of all subjects' data, and then normalized each subject's distance data within a range of 0 and 1. Then, we cumulated all subjects' normalized distances to obtain histograms (Figures 2 and 3).

Although not huge differences were shown in the histograms, the means of the saccadic distances of 1PP sequences, which are (a), (d), and (e), were higher than those of 3PP sequences (Table 1). This indicates that the scopes of perception in first-person perspectives tend to be more wide and arbitrary than those of third-person perspectives.

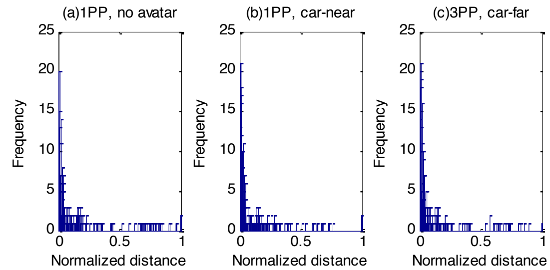


Figure 2. Histograms of normalized distances for Set 1.

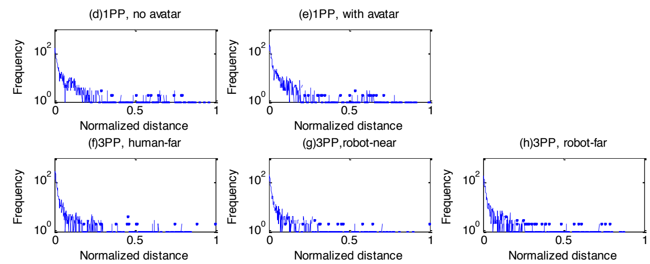


Figure 3. Histograms of normalized distances for Set 2.

Set	1PP w/o Avatar	1PP w/ Avatar	3PP Near	3PP Far
1	(a) 0.109	-	(b) 0.083	(c) 0.090
2	(d) 0.087	(e) 0.067	-	(f) 0.055
			(g) 0.057	(h) 0.065

Table 1. Average normalized distances.

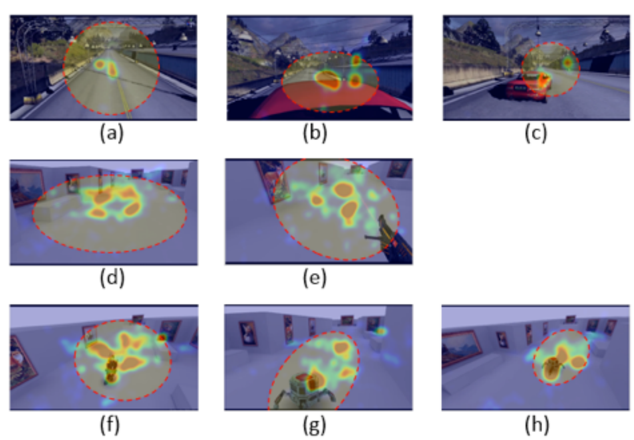


Figure 4. Heat maps for each sequence.

With the eye-tracking data, heat maps and gaze plots were obtained for each sequence. Figure 4 shows the heat maps over all subjects for each sequence. Figures 5 to 10 are the

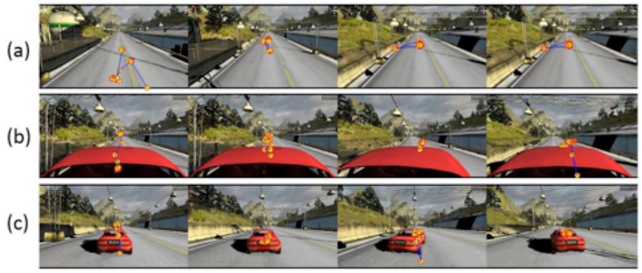


Figure 5. Gaze plots for Set 1: (a)PP1 without avatar, (b)PP3 car-near, (c)PPv3 car-far

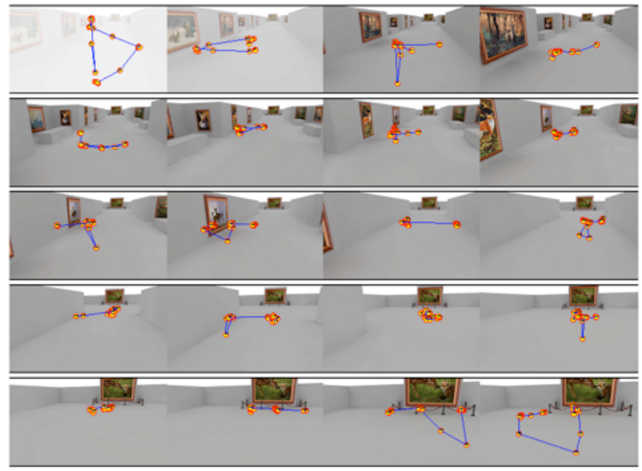


Figure 6. Gaze plots for Set 2: (d)PP1 without avatar

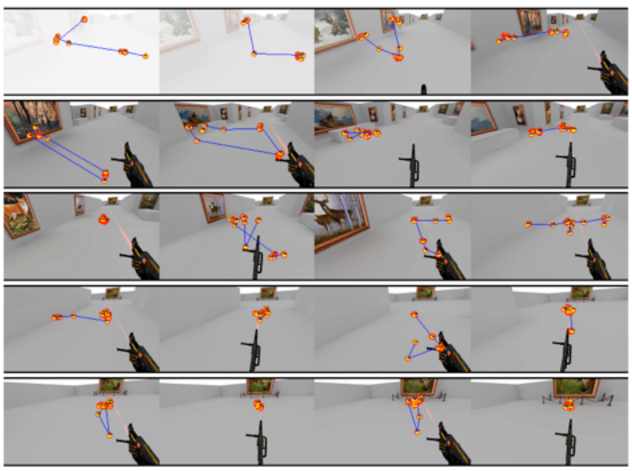


Figure 7. Gaze plots for Set 2: (e)PP1 with avatar

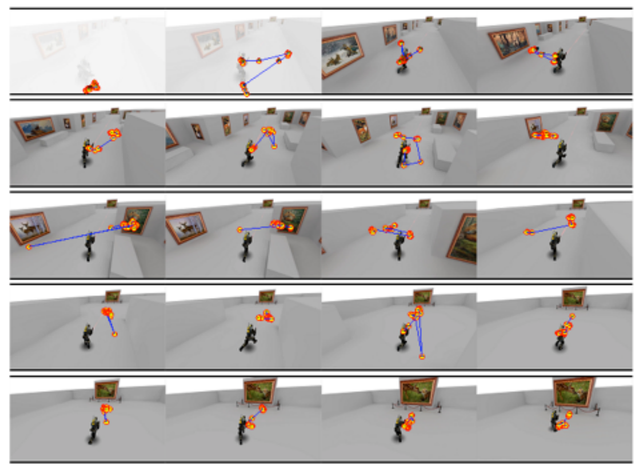


Figure 8. Gaze plots for Set 2: (f)PP3 human-far

Figure 9. Gaze plots for Set 2: (g)PP3 robot-near

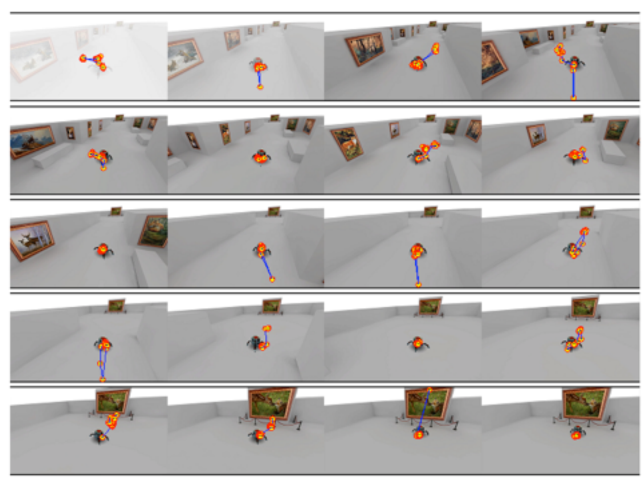


Figure 10. Gaze plots for Set 2: (h)PP3 robot-far

samples of individual (subject 5) gaze plots, where a screen shot of the gaze plots was taken every 30 frames for qualitative analysis.

Overall, if a sequence had 1PP, gaze points showed arbitrary saccadic movements that seemed to scan the environment from the subjects' own point of views outside of the digital environment. The gaze plots are dispersed, and the ranges of the navigating areas of the subjects are relatively large. If a sequence had 3PP, however, gaze points were concentrated to both of the avatar's position and the peripheral space, or namely, the avatar's peripersonal (reachable) space [4]. In cognitive science, first-person perspective is referred to as an egocentric perspective, constituted by subject-to-object relation, and third-person perspective is referred to as an allocentric perspective, constituted by object-object relations [8]. Comparing the two gaze plots in Figures 6 and 10, we can indeed observe such differences between 1PP and 3PP, as the users' scopes of visual perception were adjusted according to the subject-object relations or object-object relations.

There are two types of mechanisms involved in controlling the scope of visual perception in addition to the perspectives of the users [7]. Bottom-up process is driven by the perceived stimulus, while top-down process is caused by demands of attention. As avatars somehow displayed strong visual saliencies, the avatars themselves attracted the subjects' attentions, but once the subjects got used to the presence of the avatars, subjects seemed to navigate the virtual environment related to the avatars' reachable space or peripersonal space. In case of 1PP with extended part of the avatar or tool shown, gaze points displayed capricious saccadic movements similar to the saccades of 1PP with some constraint, as shown in Figures 7 and 4(e). Unlike 3PP, the extended part of the avatar, which displayed a strong visual saliency, did not attract subjects' attention but rather worked as a visual barrier or an extended body schema [3] that limits the subjects' scope of visual perception and redirect the subjects' attention with contextual cues [2].

In summary, we can argue that the different perspectives with avatars, especially with the spatial information of the avatars related to the 3D virtual space, help users to adjust their scopes of the visual perception within 3D environments. Thus, it would be possible to effectively manipulate the perceptual scope of users by adjusting the presence and locations of avatars in 3D virtual space. This can be useful in designing 3D virtual space with particular purpose in applications. For instance, when it is required to narrow the perceptual area to unburden users' cognitive load, it would be helpful to use 3PP with an avatar. And if it is required to navigate the environment without constraint, it would be better to use 1PP. With 1PP the scope of the perceptual area can be limited by using a pointing device (gun in our case) as well.

CONCLUSION

In this work we analyzed the eye-gaze data from avatar-based navigation and the correlation between perspectives and perception within 3D virtual environment. We drew interesting interpretation on first-person perspective and third-person perspective and how these perspectives can improve 3D navigation effectively by changing the scope of visual perception. This result can be beneficial to 3D applications where navigation is important.

ACKNOWLEDGMENTS

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the "IT Consilience Creative Program" (IITP-2015-R0346-15-1008) supervised by the IITP (Institute for Information & Communications Technology Promotion).

REFERENCES

1. Cockburn, A., and McKenzie, B. Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*, (2002), 203-29.
2. Chun, M. M. Contextual cueing of visual attention. *Trends in Cognitive Sciences 4*, 5 (2000), 170-177.
3. Holmes, N. P., and Spence, C. The body schema and the multisensory representation(s) of peripersonal space. *Cognitive Processing 5*, 2 (2004), 94–105.
4. Làdavas, E. Functional and dynamic properties of visual peripersonal space. *Trends in Cognitive Sciences 6*, 1 (2002), 17-21.
5. Morel, B., and Ach, L. Attention-aware intelligent embodied agents. In *Human Attention in Digital Environments*, Roda, C. (ed.) Cambridge University Press, Cambridge, UK, (2011), 147-165.
6. Parush, A. Navigation and orientation in 3D user interfaces: The impact of navigation aids and landmarks. *International Journal of Human-Computer Studies 61*, 3 (2004), 375-395.
7. Räihä, K.-J., Hyrskykari, A., and Majaranta, P. Tracking visual attention and adaptive applications. In *Human Attention in Digital Environments*, Roda, C. (ed.) Cambridge University Press, Cambridge, UK, (2011), 166-185.
8. Vogeley, K., and Fink, G. R. Neural correlates of the first-person perspective. *Trends in Cognitive Sciences 7*, 1 (2003), 38-42

Gaze Analysis of Avatar-based Navigation with Different Perspectives in 3D

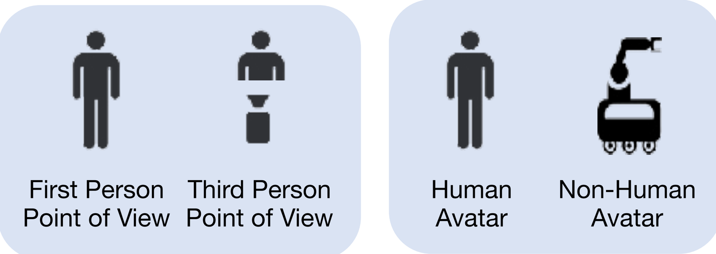
Jooyeon Lee
Information & Interaction Design, Underwood International College, Yonsei University, Korea

Manri Cheon, Seong-Eun Moon, Jong-Seok Lee
School of Integrated Technology & Yonsei Institute of Convergence Technology, Yonsei University, Korea

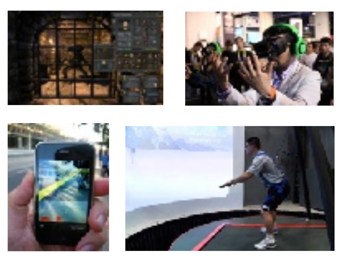
Abstract

This paper explores the relations between perspectives of navigation and visual perception in 3D virtual space, by analyzing avatar-based navigation with the eye-gaze data. We examine how different perspectives and types of avatars affect the users' scopes of visual perception within 3D virtual environments. Throughout this research, we attempt to draw possible connections between the perspectives and cognitive patterns of visual perception. We propose that manipulating perspectives of avatars or those of users has immediate effects on the users' scopes of visual perception and patterns of visual attention.

Types of Avatar in 3D Virtual Space



3D Virtual Space Applications



Study Design

Participants
2 Males and 3 Females
(age=20-24 years, M=21.5 years, SD=1.48)

Experiment Environments
Smart Eye Pro 5.9 system with three cameras (a sampling rate of 60 Hz)
Computer Monitor (21x12 inches, 1980x1080 pixels)

Video Clips
Set 1 (5 seconds) and Set 2 (25 seconds) with a frame rate of 24 Hz

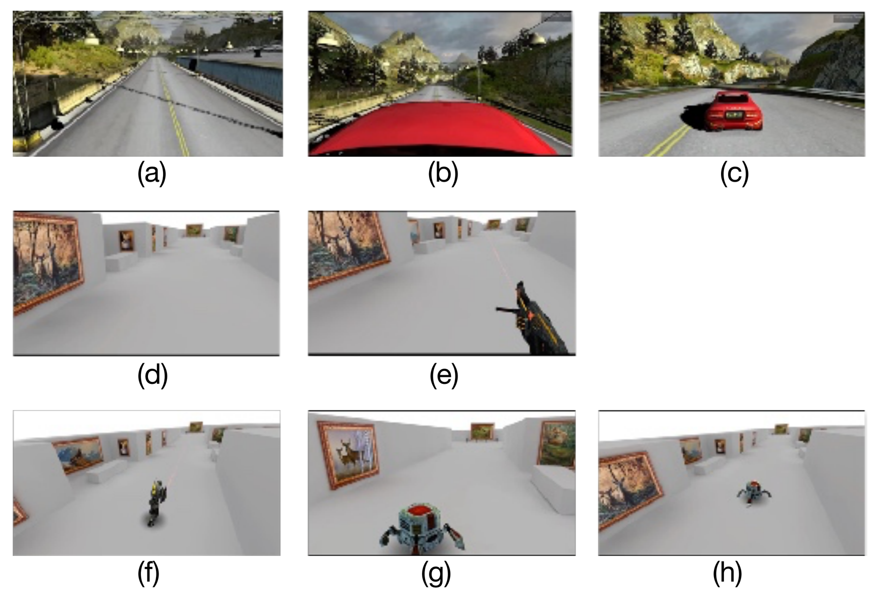


Figure 1. Two sets of 3D environments, a road and an art gallery, in both first-person (1PP) and third-person (3PP) perspectives.

Eight Sequences in Two Sets

(a) 1PP sequence with no avatar on a road
(b) 3PP sequence with a car closer to the viewpoint of the subjects
(c) 3PP sequence with a car farther from the viewpoint of the subjects
(d) 1PP sequence with no avatar in an art gallery
(e) 1PP sequence with a tool (gun) as the extended part of subjects' virtual arms
(f) 3PP sequence with a human avatar
(g) 3PP sequence with a robot avatar closer to the viewpoint of the subjects, and lastly
(h) 3PP sequence with a robot avatar farther from the viewpoint of the subjects

Results

We down-sampled the eye-gaze data from 60 Hz to 24 Hz to match the frame rate of the video clips. In order to compare each sequence's overall frequency of ranges of saccades, we calculated the distances between two eye-gaze points on consecutive frames of all subjects' data, and then normalized each subject's distance data within a range of 0 and 1. Then, we cumulated all subjects' normalized distances to obtain histograms (Figures 2 and 3).

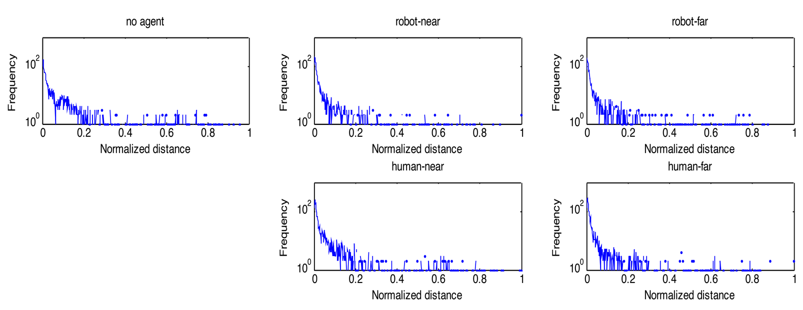


Figure 2. Histograms of normalized distances for Set 1.

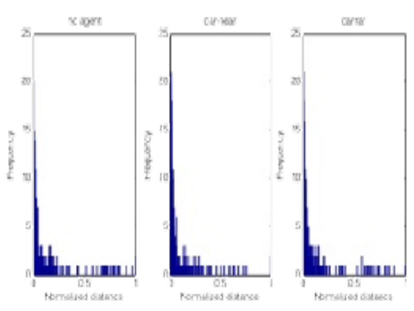


Figure 3. Histograms of normalized distances for Set 2.

Set	1PP w/o Avatar	1PP w/ Avatar	3PP Near	3PP Far
1	(a) 0.109	-	(b) 0.083	(c) 0.090
2	(d) 0.087	(e) 0.067	-	(f) 0.055
			(g) 0.057	(h) 0.065

Table 1. Average normalized distances.

Although not huge differences were shown in the histograms, the means of the saccadic distances of 1PP sequences, which are (a), (d), and (e), were higher than those of 3PP sequences (Table 1). This indicates that the scopes of perception in first-person perspectives tend to be more wide and arbitrary than those of third-person perspectives.

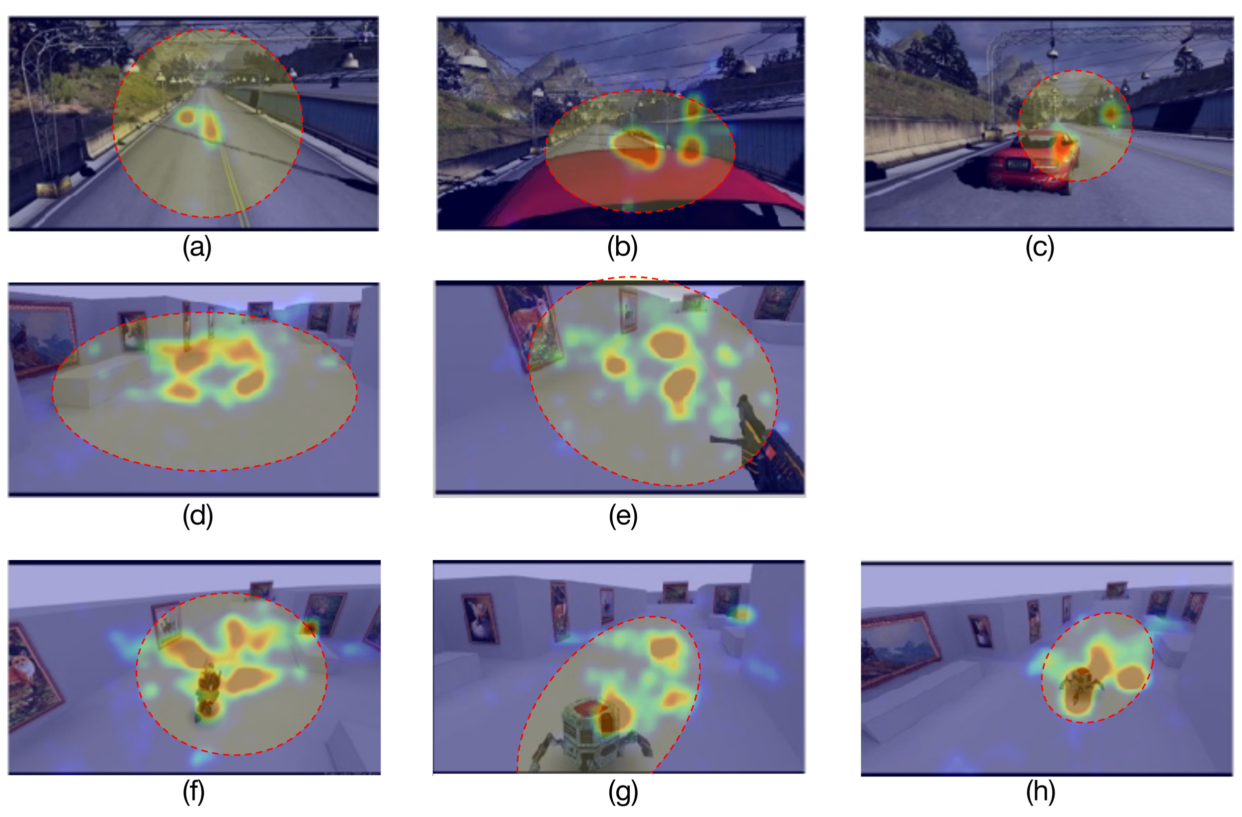
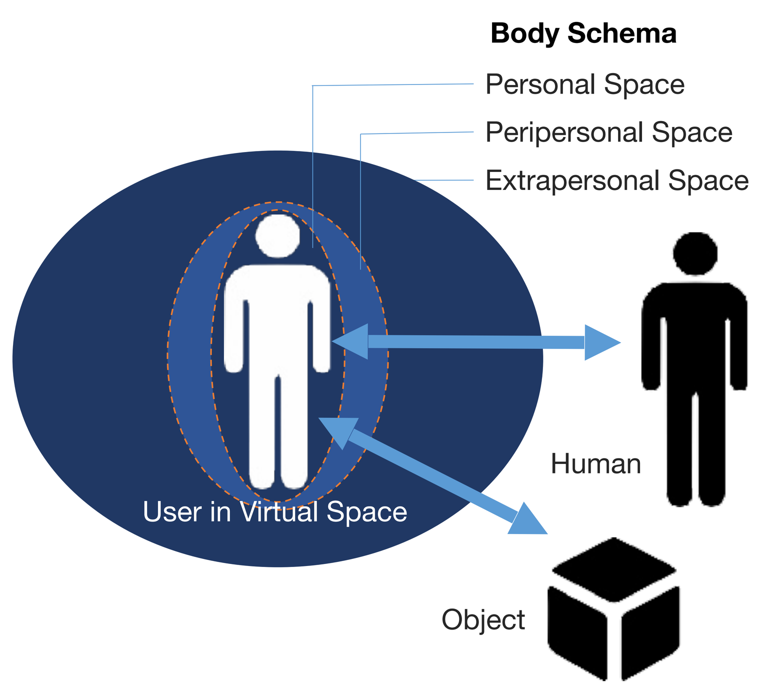


Figure 4. Heat maps for each sequence.

Overall, if a sequence had 1PP, gaze points showed arbitrary saccadic movements that seemed to scan the environment from the subjects' own point of views outside of the digital environment. The gaze plots are dispersed, and the ranges of the navigating areas of the subjects are relatively large. If a sequence had 3PP, however, gaze points were concentrated to both of the avatar's position and the peripheral space, or namely, the avatar's peripersonal (reachable) space. In cognitive science, first-person perspective is referred to as an egocentric perspective, constituted by subject-to-object relation, and third-person perspective is referred to as an allocentric perspective, constituted by object-object relations [8]. Comparing the two gaze plots in Figures 6 and 10, we can indeed observe such differences between 1PP and 3PP, as the users' scopes of visual perception were adjusted according to the subject-object relations or object-object relations.



There are two types of mechanisms involved in controlling the scope of visual perception in addition to the perspectives of the users. Bottom-up process is driven by the perceived stimulus, while top-down process is caused by demands of attention. As avatars somehow displayed strong visual saliencies, the avatars themselves attracted the subjects' attentions, but once the subjects got used to the presence of the avatars, subjects seemed to navigate the virtual environment related to the avatars' reachable space or peripersonal space. In case of 1PP with extended part of the avatar or tool shown, gaze points displayed capricious saccadic movements similar to the saccades of 1PP with some constraint, as shown in Figures 7 and 4(e). Unlike 3PP, the extended part of the avatar, which displayed a strong visual saliency, did not attract subjects' attention but rather worked as a visual barrier or an extended body schema that limits the subjects' scope of visual perception and redirect the subjects' attention with contextual cues.

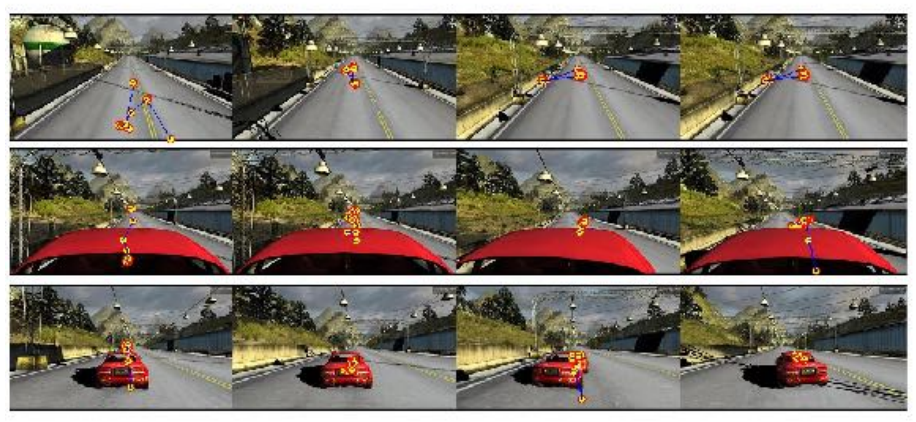


Figure 5. Gaze plots for Set 1: (a)PP1 without avatar, (b)PP3 car-near, (c)PPv3 car-far

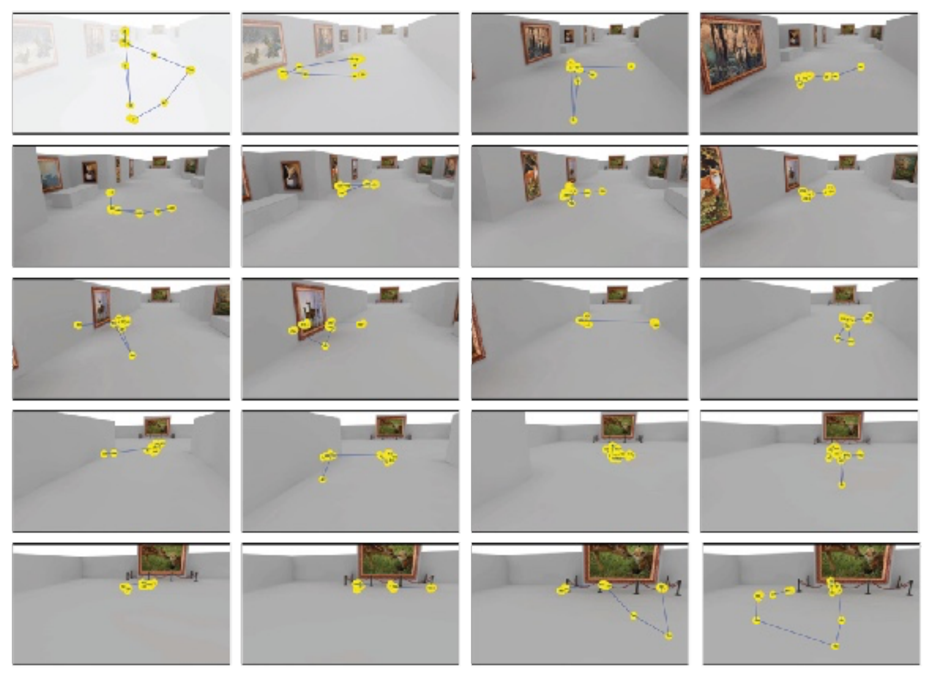


Figure 6. Gaze plots for Set 2: (d)PP1 without avatar

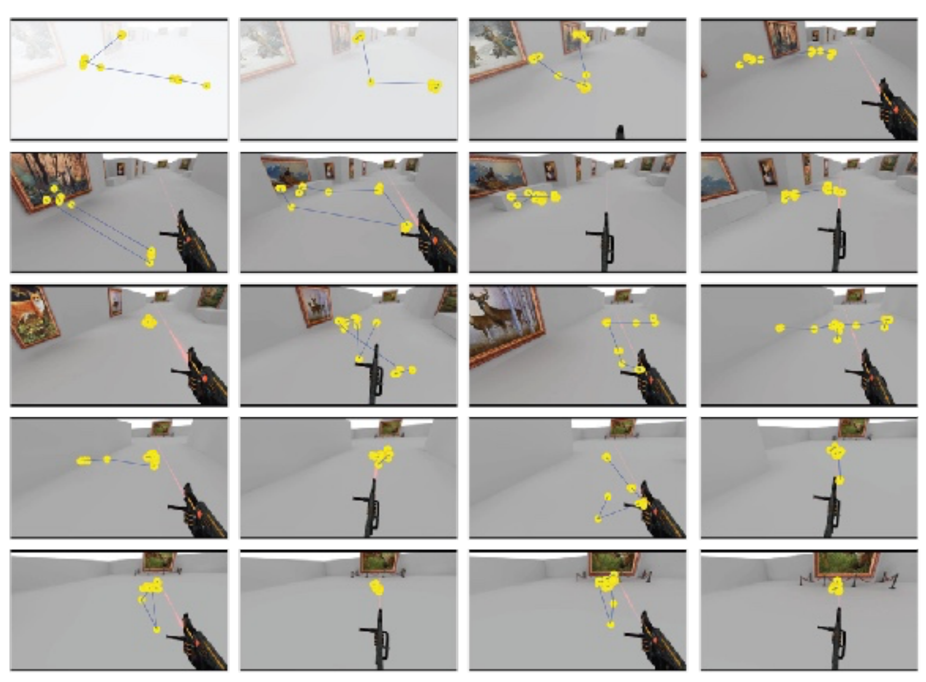


Figure 7. Gaze plots for Set 2: (e)PP1 with avatar

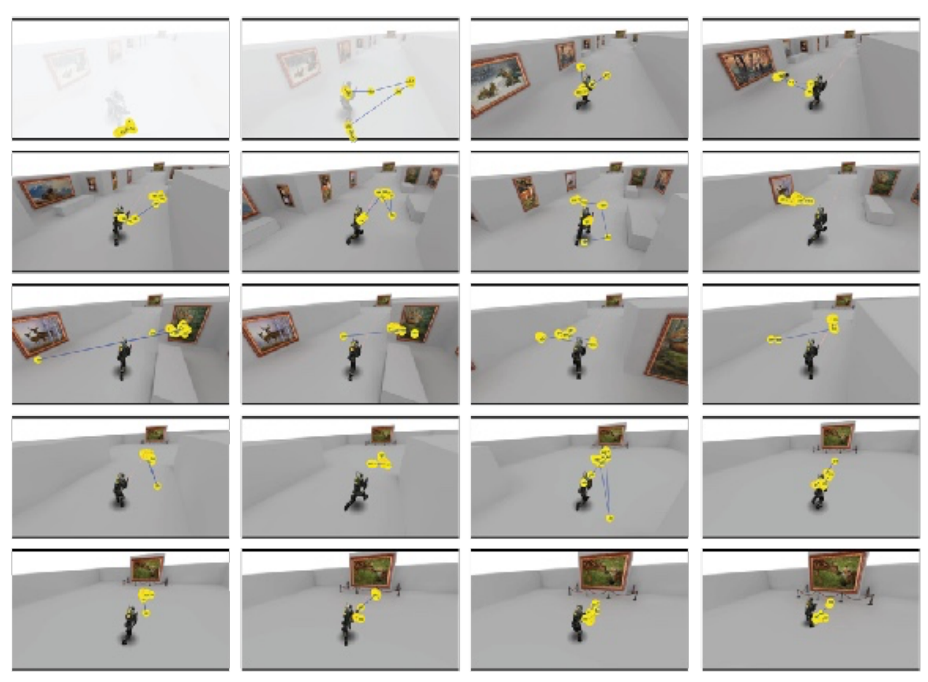


Figure 8. Gaze plots for Set 2: (f)PP3 human-far

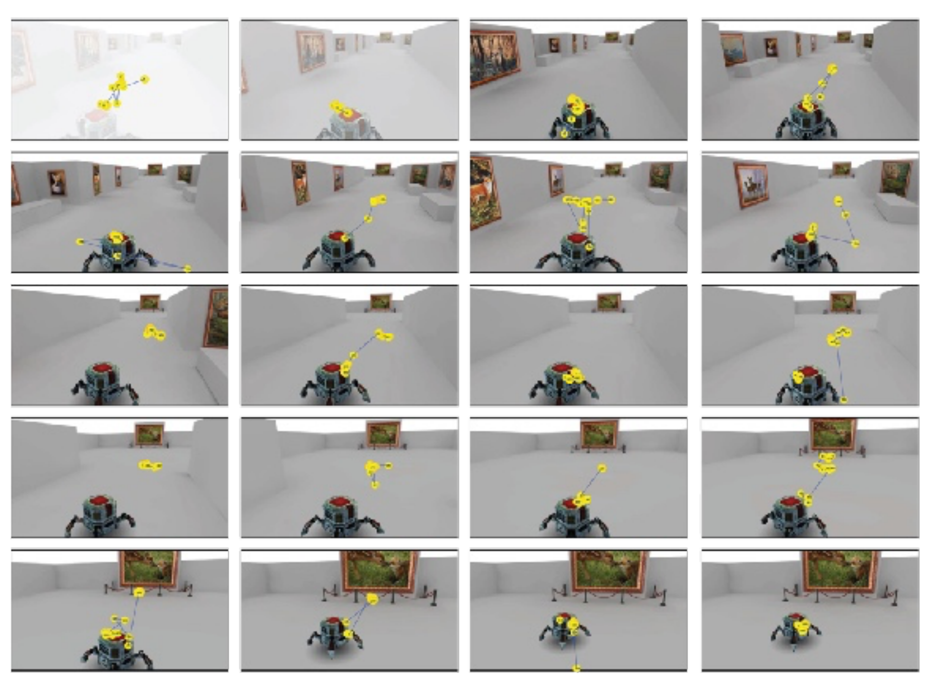


Figure 9. Gaze plots for Set 2: (g)PP3 robot-near

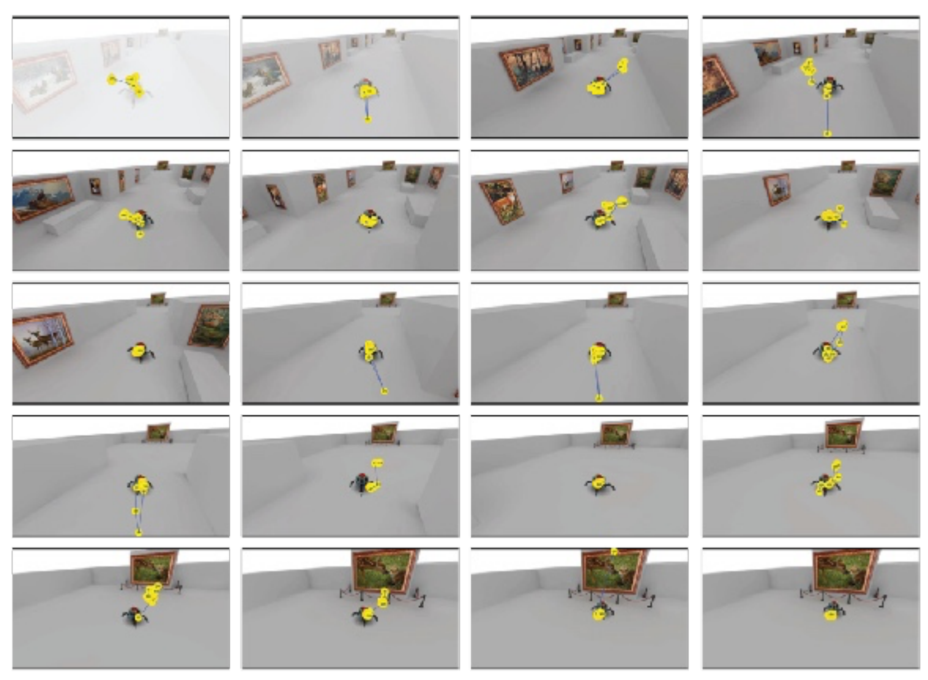


Figure 10. Gaze plots for Set 2: (h)PP3 robot-far

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the "IT Consilience Creative Program" (IITP-2015-R0346-15-1008) supervised by the IITP (Institute for Information & Communications Technology Promotion).