How People Use Visual Landmarks for Locomotion Distance Estimation: The Case Study of Eye Tracking

Huiting Zhang and Kan Zhang

Abstract Research has been focusing on how people navigate in the virtual space since the technology of virtual reality was developed. However, not enough has been known about the process of the virtual space cognition. During locomotion, distance could be visually accessed by integrating motion cues, such as optic flow, or by the self-displacement process in which people compare the change of their self-position relative to individual identifiable objects (i.e. landmarks) in the environment along the movement. In this study, we attempted to demonstrate the effect of the later mechanism by separating the static visual scenes from the motion cues in a simulated self-movement using a static-frame paradigm. In addition, we compared the eye tracking pattern in the static scene condition (without motion cues) with the eye tracking pattern in the full visual cue condition (with motion cues). The results suggested that when only static visual scenes were available during the simulated self-movement, people were able to reproduce the traveled distance. The eye tracking results also revealed there were two different perceptual processes for locomotion distance estimation and it was suggested that locomotion distance could be estimated not only by optic flow as we already knew, but also by the self-displacement process from the visual static scenes.

Keywords Landmarks · Locomotion distance estimation · Eye tracking

H. Zhang (✉) · K. Zhang
Institute of Psychology, Chinese Academy of Sciences, Beijing, China
e-mail: zhanghuiting@psych.ac.cn

K. Zhang
e-mail: zhangk@psych.ac.cn

H. Zhang
University of Chinese Academy of Sciences, Beijing, China

F. Sun et al. (eds.), *Foundations and Applications of Intelligent Systems*,
Advances in Intelligent Systems and Computing 213,
DOI: 10.1007/978-3-642-37829-4_8, © Springer-Verlag Berlin Heidelberg 2014

1 Introduction

Visual perception has long being a critical subject in the spatial cognition and virtual reality research. Estimation of the locomotion distance is important for navigation and spatial learning. When people are moving in the real world, distance estimation process happens naturally throughout the trip that the information from vision [1], proprioception, motor commands [2, 3] and vestibular sense [4] is received and integrated together. However, in the virtual reality, not only is the visual information different from the real world, but the body sense is also largely reduced. In previous studies, humans were found to be able to judge the distance in a virtual environment or estimate the traveled distance regardless of a certain amount of bias via visual information as the main perceptual cues. In this study, we further divided the visual information during a simulated movement into motion cues and static scenes, and attempted to prove the existence of two different perceptual processes based on these two kinds of visual information.

2 Related Works

Humans were found to be able to estimate their movements and traveled distance depending only on the visual information [5–7]. In a visually full-cue environment, there are two basic mechanisms that visual information can be used to localize oneself in an environment. The estimation of locomotion distance could be done by integrating the visual motion cues about the direction and speed of one's movements over time [8, 9]. For example, in order to judge the distance traveled on the basis of optic flow, the observer can first estimate the velocity and duration of the self-motion [1, 10]. And then the velocity could to be integrated over time to determine the distance that the observer has traveled. Estimation of distance traveled based on optic flow, such as in a texture environment, has been demonstrated with various tasks, such as distance discrimination [11], distance adjustment [12, 13] and distance reproduction [14]. In all cases a linear relationship between the perceived and the actual distance was observed, but with a consistent undershooting of absolute magnitude.

Vision can also be used to directly determine one's position relative to individual identifiable objects in the environment [15] based on a place/scene recognition process. For example, landmarks, defined as visible, audible, or otherwise perceivable objects which are distinct, stationary, and salient [14], can be treated as reference about one's position within the perceptual range, and can support self-localization process. It is known that people have the ability to judge the egocentric distance between oneself and an object in different static environments [16, 17]. Much research has been done to examine how humans perceive egocentric distance when viewing a target from a fixed viewpoint [18]. Perceived distance, for the most part, is also linearly correlated to physical distance though consistently underestimated as the simulated distances increased [13, 19]. Therefore, traveled distance can also be assessed by subtracting the egocentric distance from the observer to a certain perceivable landmark between the static scenes along the trip, between the end of the trip and the start of the trip for instance.

Very few studies to our knowledge investigated the effect of only static visual cues on the estimation of the distance traveled during locomotion. Lappe et al. [15] studied distance judgment involving static scenes using a virtual environment of a hallway with randomly colored panels on both walls. Participants were first visually moved a certain distance and then asked to adjust a target in a static scene for the same distance or first shown a target in a static scene and then asked to translate the same distance with active or passive simulated movement. However, in this experiment, the static visual scenes were only used for estimating the egocentric distance instead of the locomotion distance. Thus it did not address the question on whether locomotion distance estimation is possible with static visual information alone.

Besides, studies on eye movements during locomotion understand well about how the visuomotor and vestibule-motor systems function and interact [20]. However, there was little research focusing on the eye tracking with the visual cues, specifically on landmarks, in the field of locomotion distance estimation either in the real world or in virtual realities, partly because the technological constrains and the challenges for recording and the data analysis of the eye movement on dynamic stimuli in 3D space.

In this study, we sought to compare the effect of static visual information (without motion cues) on locomotion distance estimation with the full cue visual condition (with motion cues), in other word, to investigate of the effectiveness of static-scene mechanism. The virtual environment built in this experiment was a tunnel with several distinctive, identifiable landmarks on the walls, floor and ceiling. The perceptual landmarks can first provide unique patterns of optic flow during locomotion and thus can be used for distance estimation using the motion-based mechanism. At the same time, distinctive perceptual landmarks can also function as self-localization cues to determine one's position at a given moment, and therefore supporting a mechanism based on the static scene. To separate the static scenes from the motion cues, a static-frame paradigm was first developed and used. The detailed description would be given in the method part. Furthermore, to validate whether people actually use different mechanisms when different cues were available, we also recorded their eye movement throughout the experiment. In our hypothesis, when full visual information is available during the locomotion, people follow the position of the perceptual landmarks throughout the trip to access the optic flow and integrate the traveled distance and their gaze points should be moving smoothly and highly close to the position of the landmarks. However, when the motion cues are eliminated and only discrete static scenes are provided, their gaze points should be moving inconsecutively or jumpily according to the discontinuously position change of the landmark between the static scenes.

3 Method

3.1 Subjects

Sixteen undergraduate and graduate students (9 males, and 7 females) participated in this experiment. All participants had normal or corrected-to-normal vision and signed the consent form and were paid for their participation.

3.2 Apparatus and Virtual Environment

The experiment was conducted on a PC, running a C++ program using open GL. Participants were seated in a dimly illuminated room 60 cm from a 17-in display monitor, rendered at 72 Hz refresh rate, a resolution of 1,024 × 768 pixels, and a graphical field of view of 40° × 30°. The eye movement was recorded by Tobii T120 Eye Tracker whose sampling rate is 120 Hz.

The virtual environment was a hallway-like tunnel, 2 m wide and 3.2 m high (the simulated eye height was 1.6 m). The floor, ceiling and walls of the tunnels were set with different visual cues and there were different visual environments for learning phase and distance reproduction phase in each trail. In the learning phase (see Fig. 1a), all four walls were the same solid gray color, with four objects in different shapes, one on each wall, in the order of a blue rectangle on the left wall, a red circle on the floor, a green triangle on the right wall, and a yellow star on the ceiling. The positions of the four shapes were fixed for all trials. They were placed in the middle of the wall about 0.625, 6.25, 11.875 and 17.5 m from the starting point, respectively. While in the reproduction phase, another environment was used to prevent people from using simple scene-match strategy. Four walls of the new tunnel were painted with a green and dense-patterned texture. In addition, 12 yellow diamond shapes were used as perceptual landmarks in this reproduction environment, with fixed locations different from each of the four objects used in the learning phase, three on each wall (see Fig. 1b). The order and position of these shapes were randomly chosen and kept the same for all trials.

3.3 Design and Procedure

In each trial, participants were asked to first watch a simulated self-movement along the center line of the tunnel for a certain distance (the learning phase) and then to reproduce the distance with another simulated movement in a different speed (the reproduction phase). Each simulated self-motion was initiated by pressing the "SPACE" key on the keyboard by the participants, followed by a fixation screen for 500 ms before the test stimuli appeared. In the learning phase,

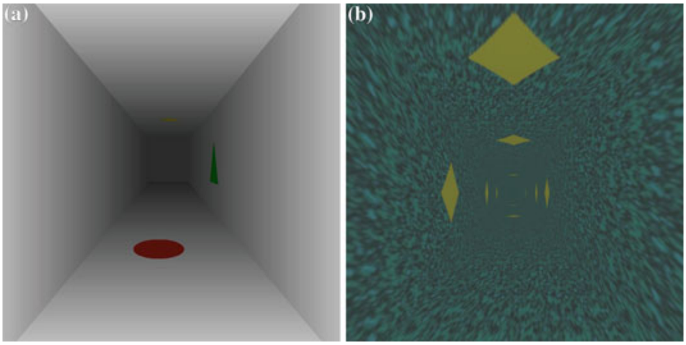

Fig. 1 The illustration of the tunnel in **a** the learning phase and in **b** the reproduction phase

the movement stopped automatically. In the reproduction phase, the participants needed to decide when/where to stop the movement by pressing the "SPACE" key again. Participants were told the locomotion speed was randomly chosen and were different in the learning and reproduction phase to prevent them from counting the time. However, for the analysis of the eye tracking data, actually we kept the speed constant as 1.25 m/s in the learning phase, and 1.5 m/s in the reproduction phase. The corresponding learning duration was 4.8–9.6 s, and the response duration was determined by the participants.

A static-frame paradigm was used to manipulate the availability of the motion cues' during the movement. In the learning phase, while the simulated movement was continuous, only a sequence of static scenes at given instances were provided to the participants, mimicking a walk in the dark tunnel where a strobe light periodically illuminated the surroundings. Every static frame was presented for 100 ms, and participants could see where they were in the tunnel at that moment. The interval between every static scene was completely dark, and the durations lasted for 0 or 1,000 ms. When the interval is 0, though the physical stimuli were a series of static frames, the apparent motion existed and continuous visual scenes would be perceived and the 0 ms interval condition served as a full-cue condition (with both motion cue and static scenes). When the interval was 1,000 ms, the blank was long enough for participants to notice, and the discrete static scenes would be perceived. During the reproduction phase, both the movement and the visual scene were continuous (like the tunnel was continuously illuminated).

A two-factor within-subject design was used, with 3 (distance: 6, 9 or 12 m) × 2 (interval durations: 0 or 1,000 ms) and each condition was tested for three times in three blocks. All trials were randomized within a block. Participants were not given any feedback about their performance. The whole experiment lasted about 20 min.

4 Results

Using dynamic stimuli in the eye movement study was challenging from the technical aspect for the eye movement would be more complicated than the eye movement on an image or a sentence. It was reasonable that the eye tracker could not record as complete data as it does when it is used for pictorial material or web pages. In this case, we used data that had more than 50 % sampling results (data of one participant was eliminated from the following analysis). Since the original sampling rate is 120 Hz, we believed the 50 % sampling was acceptable.

Participants' eye movements were recorded for the whole experiment. In this study, since the stimuli were changing continuously and rapidly, we currently focused on the gaze data in the learning phase, to investigate which position on the screen they looked at to perceive the distance. The distance reproduction performance and eye tracking data were analyzed and reported separately.

4.1 Distance Reproduction

One data point was excluded from the analysis whose reproduction distance was 0.225 m. For the shortest distance to be estimate was 6 m, we believed this extreme data was from accidentally pressing the SPACE key too quickly in the reproduction phase.

A repeated measure ANOVA (3 correct distances × 2 intervals) was conducted on the reproduced distance. To better understand the results, we referred 0 ms interval condition as the continuous condition and the 1,000 ms interval condition as the static scene condition. Only the main effect of the correct distance was significant ($F_{(2,28)} = 104.338$, $p < 0.001$, see Fig. 2). Participants tended to overestimate the distance of 6 m ($M = 6.303$, SE = 0.352) and underestimate the distance of 9 m ($M = 8.355$, SE = 0.428) and 12 m ($M = 10.537$, SE = 0.591). Neither the effect of the interval ($F_{(1,14)} = 0.912$, $p = 0.356$) nor the interaction was significant ($F_{(2,28)} = 0.120$, $p = 0.888$).

In addition, if the participant's estimation is accurate, the reproduction distance should be highly correlated with the correct distance, and the slope of the reproduction distance in the function of the correct distance should be close to 1. Therefore, we ran linear regressions for the reproduction distance in relationship with the correct distance for each participant. And the t test on the slope between two interval conditions ($M\_{\text{static}} = 0.692$, SE = 0.058; $M\_{\text{continuous}} = 0.719$, SE = 0.093) failed to reveal any significant differences ($t_{(14)} = -0.294$, $p = 0.774$).

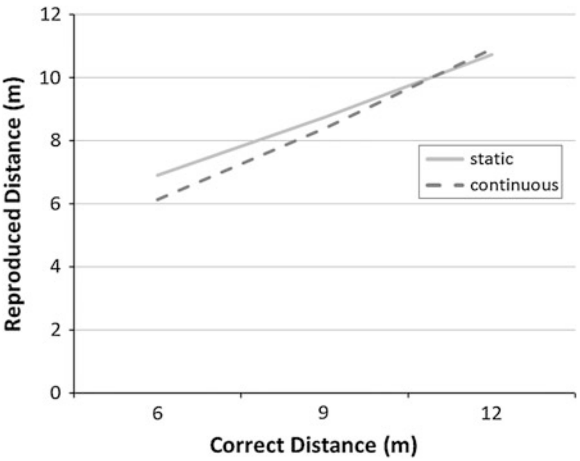

Fig. 2 Reproduction distance as the function of the correct distance in different interval conditions

4.2 Gaze

The raw gaze data was recorded in form of the x–y coordinates of the screen. In order to decide whether the participants were looking at the specific landmark, we need to compare the gaze points of the participants with the location of the landmarks along the simulated movements. First, the simulated self-movements in the interval-0 condition were divided into several 100 ms-long segments, and the first frame of these 100 ms-long segments were used as the key frames. Next, the coordinates of the center of the landmarks in the key frames were recorded and set as the target points. Then the distance (referred to "the deviation distance" hereafter) from participants' gaze point to the simultaneous target point was calculated as the dependent variable. The main consideration of this transformation was that both the location and the size of the landmarks were changing continuously and it would be a time-consuming process to circle the area-of-interest frame by frame for each participant. Compared to the effort, the benefit from this finest coding would be trivial since the pixel change of the location and size of the landmarks between two sequent frames was really small. Therefore, we used the deviation distance for the following analysis.

We ran a repeated measure ANOVA (3 correct distances × 2 intervals) on the deviation distance. First, as the same as the result of the reproduced distance, the main effect of distance was significant ($F_{(2,28)} = 18.537, p < 0.001$, see Fig. 3). It was suggested that participants better followed the location of the landmarks when the distance was short ($M\_{6} = 56.290$, SE = 1.547; $M\_{9} = 65.006$, SE = 2.409; $M\_{12} = 68.656$, SE = 1.962). Besides, the effect of the interval was also significant ($F_{(1,14)} = 4.566$, $p = 0.051$) that when the interval was 0 ($M = 66.239$, SE = 2.395), participants were more likely to follow the location of the landmark than they did when the interval was 1,000 ms ($M = 59.729$, SE = 2.032). No interaction effect of these two factor was significant ($F_{(2,28)} = 0.207, p = 0.814$). In addition, we did the same analysis on the standard deviation of the deviation distance

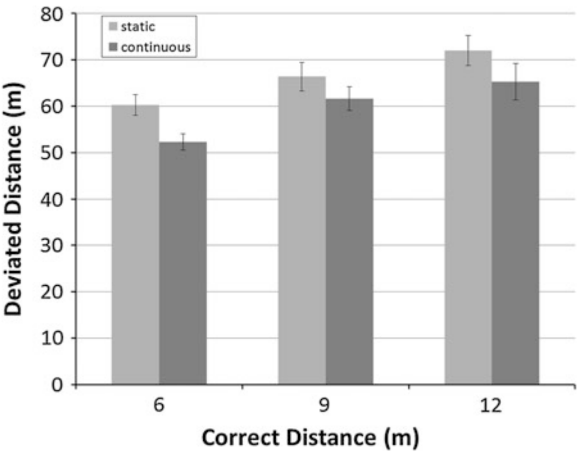

Fig. 3 Deviation from the center of the landmark as the function of the correct distance in different interval conditions

to investigate the constancy of the eye following behavior on the landmarks in different interval conditions. However, the only significant result found was the effect of the correct distance ($F_{\_\text{distance (2,28)}} = 3.962$, $p < 0.05$; $F_{\_\text{interval(1,14)}} = 0.105$, $p = 0.750$; $F_{\_\text{interaction(2,28)}} = 2.146$, $p = 0.136$).

5 Discussion

Two questions were addressed in this experiment: whether people were able to reproduce the distance during simulated locomotion when the motion cue was removed, as in the interval-1,000 condition; and how they used the perceptual landmarks when they perceived the distance.

In a distance reproduction task, participants first visually traveled a given distance in an environment with only fixed landmarks, and then they were asked to reproduce this distance in a different environment. Two alternative strategies were eliminated or restricted in our experiment. People could not reproduce the distance by just visually matching the visual scene at the end of each movement from the learning environment to the reproduction environment. And they could not reproduce the distance by just counting the time they spent in the learning phase and reproducing the duration in the reproduction phase either, otherwise, their reproduced distance would be shorter than the correct distance in each distance condition because the reproduction speed was always faster than the learning speed.

There was a strong distance effect, that people slightly overestimated the shorter distance and increasingly underestimated the longer distances. This result was consistent with previous finding in the distance estimation studies with blind walking tasks and distance estimation tasks in the virtual reality [11, 18, 21]. The range of the underestimation and overestimation was also comparable to the results from these previous studies.

However, the results from the reproduction distance revealed no difference under the different interval conditions. As we expected, when the interval was 0 ms, participants reported to sense a smooth locomotion in debriefing, while when the interval was 1,000 ms, they could clearly tell there were blanks from time to time and perceived a discontinuous, jump-like movement. The change of the environments and the moving speed from learning to testing required the participants to actually estimate the distance without using other strategies. Therefore, it was suggested that participants could estimate the locomotion distance when they motion cues were reduced, and their estimation was as good as the estimation in the full visual cue condition.

On the basis of the reproduction performance, we further explored the gaze data of the participants when they perceived the distance in the learning phase. The hypothesis was, when the interval was 1,000 ms and there was an apparent black between static frames/scenes, if people still can use the motion-based mechanism, for example to estimate the velocity, their gaze point should be closely and smoothly follow the position of the landmarks, that would be the actual position of the landmarks when the static frame was presented or the imaginary position of the landmarks during the blank. If they use the static mechanism, they should look at the position of the landmarks only when the static scene is presented and might use the saccadic amplitudes between each frame to integrate the overall distance. Key frames were selected, and the center position of the landmarks was used to compare with the gaze point of the participants matched by time point. The distance from participants' gaze point to the center position of the landmarks (deviation distance) was calculated and used as an indicator for the usage pattern of the perceptual landmarks along the simulated locomotion.

On one hand, a strong effect of distance was also found in the deviation distance that participants showed to better follow the location of the landmarks when the distance was short. This result should be treated with caution. There might be due to accuracy loss and sampling data loss for longer recording duration with dynamic stimuli. In our case, we were not able to rule out the influence from the device. On the other hand, we found the effect of interval condition that the deviation distance under the continuous condition was significantly shorter than the one under the static scene condition. In other words, when there were apparently blank between each static scene/frame, participants either did not try to imagine and follow the possible position of the landmarks, or they were not able to do so, which suggested that they did not used the motion-based mechanism, However, the reproduction performance showed participants were still able to reproduce the distance from only several static scenes with comparable accuracy as they did in the full visual cue condition. We could preliminarily believe that the distance estimation based on static mechanism functioned as well as the motion mechanism within the distance range used in the current experiment.

6 Implications and Future Work

Visual information is so far one of the most important senses in virtual reality, though we believe other senses have been incorporating gradually into the virtual world as technology progresses. Knowing how people perceive the space in a virtual world help to discover the cognitive process in real world and on the other hand, help to improve the sense of embodiment in the virtual world. It is well known that the space in virtual world is perceived compressed or less accurately compared to its counterpart of the real world [22, 23] when asking people to make a distance judgment. However, during simulated locomotion people were able to integrate the accurate distance from only a series of discrete static scenes at least in a certain range of distances. It implied that the estimation of the near space (within the action space in our case) is accurate and the integration process is well done too and the compression is probably caused by the perception of comparable farther space. Furthermore, it was suggested that the contribution of visual information in a full visual condition should not only be attributed to the effect of optic flow, especially when there were silent landmarks in the environment.

Our work made potential contributions for the future research in two aspects. First, we developed a paradigm to separate the static scenes from the optic flow in a simulated movement and made it possible to investigate the effect of only static scenes in locomotive spatial learning. Second, though we used a coarse analysis process for the eye movement data, it is suggested the possibility to use eye tracking technology on the studies of the dynamic stimuli or 3D space. Further work is needed to elaborate the data processing method so that we can fully use the visual information presented in the virtual reality and also match it more accurately with the eye movement data from the participants.

One limitation of our study is only the desktop virtual reality was used, and the results should be taken carefully because of the restricted field of view [24]. Further exploration with immersive virtual reality technology combined with eye tracking is needed to validate the findings in this study. In addition, in the current study, we attempted to explore the people's eye tracking behavior on the landmarks during the simulated movements, and we selected key frames from the dynamic stimuli to analyze where people were looking at along a simulated trip. However, only the gaze data from the learning phase were analyzed here. In the future we could also compare the eye tracking data from the reproduction phase and incorporate more index, saccades for example, to further unveil the integration process of this static mechanism.

Acknowledgments We would like to thank Dr. Frances Wang from University of Illinois at Urbana-Champaign and Dr. Wang Ying from Institute of Psychology, CAS for their instructions and constructive suggestions on our study. And we thank Dr. Yao Lin from Institute of Psychology, CAS for his help on data analysis. Special thanks are given to our participants for their patience and valuable time.

References

1. Warren WH (1995) Self-motion: visual perception and visual control. In: Epstein W, Rogers S (eds) Perception of space and motion handbook of perception and cognition. Academic Press, San Diego, pp 263–325
2. Klatzky RL, Loomis JM, Golledge RG, Cicinelli JG, Doherty S, Pellegrino JW (1990) Acquisition of route and survey knowledge in the absence of vision. J Mot Behav 22:19–43
3. Klatzky RL, Loomis JM, Golledge RG (1997) Encoding spatial representations through nonvisually guided locomotion: tests of human path integration. In: Medin D The psychology of learning and motivation, Academic Press, San Diego, pp 41–84
4. Berthoz A, IsraIël L, Georges-François P, Grasso R, Tsuzuku T (1995) Spatial memory of body linear displacement: what is being stored? Science 269:95–98
5. Lappe M, Frenz H, Bührmann T, Kolesnik M (2005) Virtual odometry from visual flow. Proc SPIE 5666:493–502
6. Redlick PF, Jenkin M, Harris RL (2001) Humans can use optic flow to estimate distance of travel. Vision Res 41:213–219
7. Wan X, Wang RF, Crowell JA (2012) The effect of landmarks in human path integration. Acta Psychologica 140(1):7–12
8. Ellmore TM, McNaughton BL (2004) Human path integration by optic flow. Spat Cogn 4(3):255–272
9. Gibson JJ (1950) Perception of the visual world. Houghton Mifflin, Boston
10. Kearns MJ, Warren WH, Duchon AP, Tarr MJ (2002) Path integration from optic flow and body senses in a homing task. Perception 31:349–374
11. Bremmer F, Lappe M (1999) The use of optical velocities for distance discrimination and reproduction during visually simulated self-motion. Exp Brain Res 127:33–42
12. Frenz H, Lappe M (2005) Absolute travel distance from optic flow. Vision Res 45:1679–1692
13. Frenz H, Lappe M, Kolesnik M, Bührmann T (2007) Estimation of travel distance from visual motion in virtual environments. ACM Trans Appl Percept 4(1):1–18
14. Riecke BE, van Veen HACH, Bülthoff HH (2002) Visual homing is possible without landmarks: a path integration study in virtual reality. Presence 11(5):443–473
15. Lappe M, Jenkin M, Harris LR (2007) Travel distance estimation from visual motion by leaky path integration. Exp Brain Res 180:35–48
16. Loomis JM, Da Silva JA, Philbeck JW, Fukusima SS (1996) Visual perception of location and distance. Curr Dir Psychol Sci 5:72–77
17. Witt JK, Stefanucci JK, Riener CR, Proffitt DR (2007) Seeing beyond the target: environmental context affects distance perception. Perception 36:1752–1768
18. Sun H, Campos JL, Young M, Chan GSW (2004) The contributions of static visual cues, nonvisual cues, and optic flow in distance estimation. Perception 33:49–65
19. Frenz H, Lappe M (2006) Visual distance estimation in static compared to moving virtual scenes. Span J Psychol 9(2):321–331
20. Angelaki DE, Hess BJM (2005) Self-motion-induced eye movements: effects on visual acuity and navigation. Nat Rev Neurosci 6:966–976
21. Loomis JM, Klatzky RL, Golledge RG, Cicinelli JG, Pellegrino JW, Fry PA (1993) Nonvisual navigation by blind and sighted: assessment of path integration ability. J Exp Psychol Gen 122(1):73–91
22. Loomis JM, Knapp JM (2003) Visual perception of egocentric distance in real and virtual environments. In: Hettinger LJ, Haas MW (eds) Virtual and adaptive environments. Erlbaum, Mahwah, pp 21–46
23. Thompson WB, Willemsen P, Gooch AA, Creem-Regehr SH, Loomis JM, Beall AC (2004) Does the quality of the computer graphics matter when judging distances in visually immersive environments? Presence 13:560–571
24. Riecke BE, Schulte-Pelkum J, Bülthoff HH (2005) Perceiving simulated ego-motions in virtual reality—comparing large screen displays with HMDs. Proc SPIE 5666:344–355