VISUALIZATION OF SPATIAL RELATIONS IN VIRTUAL ENVIRONMENTS WITH ARTIFICIAL DEPTH CUES

MATHIAS MÜLLER

Technische Universität Dresden, Dresden, Germany
mathias.mueller@tu-dresden.de

THOMAS GRÜNDER

Technische Universität Dresden, Dresden, Germany
thomas.gruender@tu-dresden.de

INGMAR S. FRANKE

Technische Universität Dresden, Dresden, Germany
ingmar.franke@tu-dresden.de

RAINER GROH

Technische Universität Dresden, Dresden, Germany
rainer.groh@tu-dresden.de

Keywords: Gesture-based Interaction, Depth Perception, Virtual Environment

Gesture-based interaction in virtual environments often suffers from inaccurate depth estimations by the user. In particular, the usability of selection and manipulation tasks highly depends on correct distance judgments. Since haptic feedback is not available, the visual representation and the unambiguous interpretation of spatial configurations are fundamental aspects that influence the usability of gesture-based interaction in virtual environments. Reasons for incorrect judgment of egocentric depth in three-dimensional spaces are still partially unclear. We propose a different approach to overcome this issue. We introduce additional artificial depth cues, which are designed to improve the user's understanding of spatial relations within virtual environments. Three visualizations of egocentric depth are introduced: a Depth-of-Field effect inspired by focus effects in photography, an effect called Color Perspective based on observations of the use of color in art, and the visualization of volumetric shadows using Crepuscular Rays.

1. INTRODUCTION

The use of gesture-based interaction in three-dimensional virtual environments offers several opportunities. Tracking the user's body is a promising approach to deliver sufficient degrees of freedom for spatial interaction. The use of gestures seems promising as gestural user interfaces offer the potential to reproduce real-world processes and procedures in virtual environments with natural human interaction methods (Seth, Vance and Oliver 2011, 6-7). However, the perception of spatial relations in virtual environments, especially the estimation of egocentric depth, is often inaccurate due to conflicting or missing depth cues. The term egocentric depth describes the perceived distance of an object from the observer's viewpoint. Tasks requiring high spatial precision are therefore particularly prone to errors, especially if only visual feedback is provided to the user. In these scenarios, lateral positions can be determined unambiguously, whereas the correct estimation of spatial distances along the viewing direction is difficult (Fig. 1). Although several studies demonstrated distance judgment issues in virtual environments under various conditions, the specific reasons have not been identified yet (Jones *et al.* 2008, 13), (Piryankova *et al.* 2013, 162-163). Therefore, the question arises, how to enhance the estimation of spatial relationships even if the source of the misperception is unclear and the available visual representation does not provide sufficient natural depth cues. In this paper, three techniques are introduced, which augment a virtual scene with artificial depth cues. The proposed cues were designed to improve the understanding of the spatial layout in a virtual environment. The chosen approach was developed in the context of selection tasks in three-dimensional environments using gestures, since this form of interaction depends heavily on the correct visual estimation of positions due to the lack of haptic feedback. Additionally, gesture-based interaction in 3D is not constrained regarding the available degrees of freedom, hence correct positioning requires a precise judgment of distances in all spatial directions. The overall goal was to convey a more appropriate and intuitive feeling about the current position of the user's hands in relation to the virtual environment.

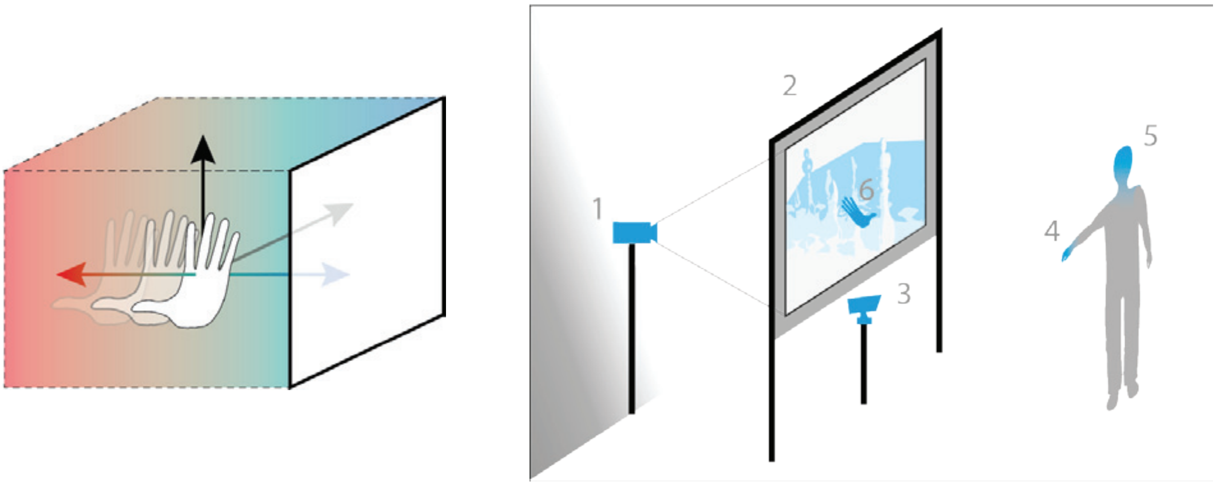


Fig. 1 Left: Spatial gestures performed in front of a 2D-projection. Movements parallel to the projection plane (black arrows) are mapped unambiguously onto the plane, whereas the perception of distances in depth (egocentric depth; colored arrows) is often incorrect in virtual environments. Right: System setup. Projector (1) displays virtual scene on a rear projection wall (2). The user is tracked by a Microsoft® Kinect™ (3). The user's head position relative to the display (5) is used to translate the virtual camera, the position of the hand (4) is projected into the virtual scene (6).

2. RELATED WORK

Interaction and the perception of space in virtual environments have been discussed in research for many years. Several publications in the field of psychology indicate that human depth perception in virtual spaces is less accurate than in reality (e.g. (Jones *et al.* 2008), (Thompson *et al.* 2004)). Although this phenomenon has been observed in numerous studies, the exact reasons are still subject to intense research. To identify the mismatches between virtual representation and the perception of three-dimensional spaces, it is important to understand how humans perceive their surroundings. The most common theory about depth perception assumes that there are several depth cues, which provide necessary information about spatial relations (Palmer 1999, 203-249), e.g. occlusion, perspective, shadow, motion parallax or binocular disparity. Due to the large number of different cues, which have been identified so far, experiments suggest that there is no single dominant cue responsible for depth perception (Wanger, Ferwerda and Greenberg 1992). In fact, humans perceive spatial layout by combining and rating different sources of information depending on the context. Additionally, the influence of specific cues is assumed to be varying depending on the distance of an object relative to the observer (Cutting and Vishton 1995). One possible reason for incorrect depth perception has been mentioned by Wann *et al.*: the conflict between ocular information (accommodation and convergence) and the stereoscopic representation provided by head-mounted displays (HMD) causes irritations, because the contradictory cues cannot be resolved properly (Wann, Rushton and Mon-Williams 1995). Watt *et al*. argued that this issue is not only present when using HMDs but in all stereoscopic representations (Watt *et al*. 2005). On the other hand, Willemsen *et al*. show that incorrect estimation of depth is not restricted to stereoscopic images, but also occurs when using monocular depth cues only (Willemsen *et al*. 2008). Blur is a depth cue, which is rarely represented in virtual environments. Although previous research shows that the distinction of small changes in blur appears to be difficult for the human visual system, blur can be seen as coarse depth cue (Mather and Smith 2002). According to Vishwanath and Blaser, blur has a significant impact on the perceived distance (Vishwanath and Blaser 2010). The authors mention the miniaturization effect used in photography, which changes the perception of the size of objects in images due to the amount of visible blur. Research also shows that, in conjunction with other pictorial depth cues, this miniaturization effect can be utilized for the estimation of absolute distances (Held *et al*. 2010). Another important finding related to virtual environments is, that the realism of the visualization seems to have only minor impact on the accuracy of distance estimations (Thompson *et al*. 2004). As the specific reasons for the inaccurate distance estimations in virtual environments have not yet been identified, the question for an alternative approach arises. Based on findings regarding depth perception, it becomes clear that even elemental cues, such as stereoscopy, are not absolutely essential for correct depth perception whereas subtle effects like blur have relative strong influence. As photorealism seems to have no significant impact on depth judgment, one possible approach could be to augment a virtual scene with subtle, yet unambiguous and easily assessable, artificial depth cues, which provide additional information about the spatial layout of a virtual environment.

3. SCENARIO

In order to identify possible approaches to enhance depth perception in a VR scenario, we used a simplistic setup for gesture-based interaction. The user interacts with a virtual scene presented on a large projection wall by using hand movements. The body tracking is implemented using a Microsoft® Kinect™, which is placed underneath the projection (Fig. 5). Virtual representations of the user's hands are projected in the virtual scene according to their position relative to the head position. The location of the user's head relative to the screen is used as position for the virtual camera. The virtual hand represents a cursor for selecting objects by virtually grasping them. Grasping is achieved by moving the hands to the position of the object. The user can move freely and his body movements trigger appropriate camera translations, in order to create the impression of an immersive interactive environment. The implemented tracking method only takes changes of the position of the user's head into account. The adaptation of the camera direction according to user's view direction was not implemented, due to missing capabilities of the tracking system. Because of the opportunity to move around and the focus on selection tasks within the virtual environment, the absence of this feature is negligible as scene navigation is not explicitly part of the scenario.

4. APPROACH

The design of the proposed artificial cues follows the idea of context-sensitive dynamic cues which should leverage the interactive dynamics of gesture-based interaction. Hence, the provided visualizations need to correlate with the position of the user's hand in the virtual scene and especially have to reflect changes of this position dynamically. Therefore we assumed, that during tasks, which require precise selection or manipulation, the user's attention lies in the area where he or she is currently interacting with his or her virtual hand. Therefore, the design of additional visual representations is focused on the dynamic visualization of the spatial relations between the virtual hand cursor and the objects nearby. The main goal was to support the perception of local depth relations in the area around the virtual hand as this is where the user needs to be supported with additional depth cues in order to achieve an efficient and precise interaction. The main idea behind the proposed effects is to provide additional visual indications, which convey a better impression about spatial relations. The user should be able to interpret these cues intuitively, in contrast to other techniques such as dynamic grids, which force the user to count rows in order estimate a correct depth relation. Therefore, the cues are based on familiar natural phenomena, which are easy to understand and to evaluate. Guiding the user's attention by manipulating the area next to the virtual hand can also reduce the distraction by other objects in the scene. The virtual representation of the user's hand becomes a tool, which is used to explore the scene in detail, without losing the capability to maintain a general overview of the whole environment. Choosing this approach, one design goal for the artificial depth cues was that they have to reflect the interactivity of the gesture-based user interface and therefore have to change their behavior dynamically according to the user's actions. Hence, the additional cues needed to be dynamic cues. Since previous research shows that dynamic depth cues are among the most powerful cues (Domini & Caudek 2003), it can be assumed that the proposed visualizations have noticeable impact on the user's understanding of the spatial layout of a virtual scene.

5. PROPOSED ARTIFICIAL DEPTH CUES

The first proposed visualization of spatial distances utilizes a Depth of Field effect. The area around the cursor is rendered normally, whereas objects farther away or closer to the observer are out of focus and increasingly blurred. As the distance between objects along the viewing direction increases, they appear more and more blurred. Therefore, a narrow area perpendicular to the user's viewing direction, called Depth of Field, is rendered sharply (Fig. 2). The basic idea of using this kind of depth visualization originates from photography, where the Depth of Field effect is used to guide the user's attentional focus to specific areas or objects. Although the different shades of blur over the scene do not serve as reliable depth cue alone, the change of blur during interaction and the corresponding movement of the focal area allow an iterative refinement of the target area. Moreover, the attribute of sharpness indicates clearly that the virtual hand and the object are at the same depth distance - contrary to occlusion alone, where the hand could be behind the object. To achieve this, the blurred region must be easily distinguishable from areas in focus, so that a high amount of blur was applied to the scene, comparable to a wide aperture in photography.

Fig. 2 Left: Depth of Field used in photography to guide the viewer's attention and minimize distracting influences background objects. Right: Implementation of the effect.

The second depth visualization is inspired by painters, who used the effect of color perspective in their pictures to enhance the impression of depth (Gombrich 2002). One part of this technique is called aerial perspective and is one of the pictorial depth cues used for estimation of large distances. Due to dirt particles in the air, the light is scattered differently depending on its wavelength. As long-wave light is scattered to a higher degree, objects far away from the observer have a bluish color (Goldstein 2010, 232). Color perspective uses this observation and its inversion for closer distances by additionally applying warm colors for objects nearby the observer to intensify the spatial impression. This idea is applied to the virtual environment by dividing the scene into depth segments, which are overlaid with different color shades. Before the color overlay is applied, the saturation of the scene is adapted to the current depth position. Therefore, proximal parts are gradually less saturated and get a reddish color overlay. Objects near the current cursor position keep their saturation value and have only a slight color overlay, whereas objects far away from the virtual hand have a reduced saturation and are drawn with a blue shade. Using this effect, we expect the user to be able to intuitively judge depth. Similar to the Depth of Field cue, colors in the area of interest nearby the virtual hand preserve their original color, whereas the effect of the color change increases with the distance to the current hand position.

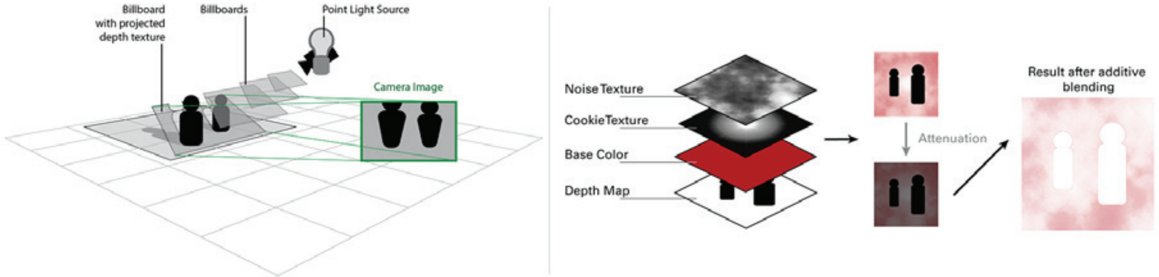
The concept for the third proposed visual effect originates from a natural phenomenon, which can be observed especially in nebular environments. Sunlight, which is scattered by dust particles in the air, produces so called crepuscular rays. This effect is adapted to the visualization of the hand's spatial position: Light shafts are displayed around the position of the hand in all directions. Objects inside these light shafts produce visible shadow volumes (Fig. 3). A torch illustrates this effect, which illuminates its surroundings, whereas the shadows provide the depth cue relative to the torch. The idea behind the effect is to utilize the power of dynamic depth cues. The shadow volumes change their size and shape with the movement of the virtual hand and therefore provide a highly dynamic cue for position relative to nearby objects. Despite being an artificial cue, the resulting shadows are intuitively interpreted and provide a natural hint for the underlying spatial layout, in contrast to “synthetic” cues such as auxiliary grids.

Fig. 3 Left: Implementation of Color Perspective, Middle: Crepuscular rays in natural environments. Right: Crepuscular Rays as artificial depth cue.

6. IMPLEMENTATION

The depth visualizations have been realized as OpenGL shaders. The Depth of Field effect was implemented by computing a depth map and applying a circular blur, with the radius increasing over the distance relative to the focal plane. This approach is similar to the approach described in (Scheuermann 2004), except that no stochastic sampling was chosen, but a uniform circular blur on several rings with increasing radius. This approach has the disadvantage that more samples need to be processed for each segment, but produces less visible blurring artifacts. The number of samples and the count of circle rings along which the blur is applied can be specified to increase the quality of the effect. One issue of this technique is an effect called edge bleeding, where fragments from sharp regions are blurred into regions farther away from the focal plane. In order to prevent this behavior, only fragments with a higher relative depth value than the processed fragment are blurred. Highlights within the scene are additionally intensified by multiplication of a luminance gain value with the resulting color. The Color Perspective effect works in a similar way by manipulating the saturation according to the rendered depth map and adding a color overlay afterwards. The implementation of the Crepuscular Rays followed the proposal of Mitchell (Mitchell 2005). Hence, six cameras in each spatial direction are attached to the virtual hand. Subsequently, a depth map is rendered from each of these perspectives. Additionally several billboard planes are aligned to the camera frustums. In the composition pass these planes are rendered with a semitransparent material, and depth values higher than the distance of the plane from the virtual hand are rejected (Fig. 4).

Fig. 4 Left: Rendering of Crepuscular Rays using projective texture mapping. Right: Composition of noise texture and depth texture to achieve natural look.



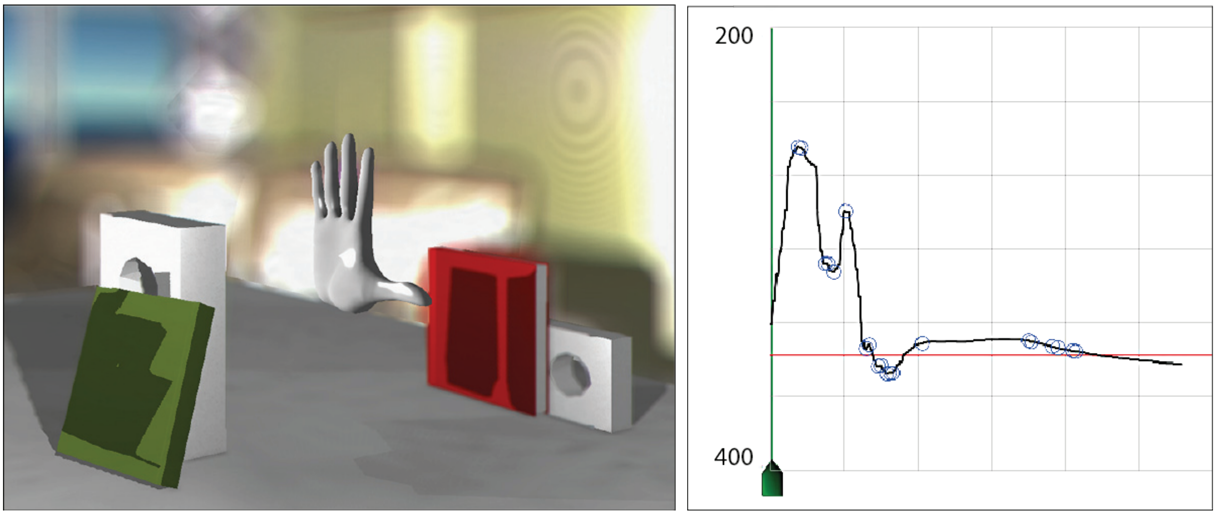
7. EXPERIMENTAL PROCEDURE

In order to acquire feedback regarding the design and spotting weaknesses of the current implementation, a preliminary survey with 24 (15 male) participants between 23 and 36 years (mean: 28,92 years, SD: 3,38) was conducted, using the setup described in Section 3 (Fig. 1). The participants were instructed to grasp different objects in the scene. Before they started, the purpose of the survey and the implemented effects were described shortly. Additionally, participants had 5 test runs for each setting, including test cases with deactivated additional depth cues, in order to become accustomed to the visualizations. Afterwards, they executed 15 trials in which they had to grasp objects with different difficulty, meaning that the objects were partially occluded by other objects or had been placed not directly in front of the user and therefore required more precise actions for selection. After completing the trials, the participants were asked to rate the usefulness of the visualizations and rank them according to their usability and interference factor regarding the visibility of the rest of the virtual scene. Moreover, they could comment on their impressions regarding the additional depth cues, their effectiveness and suggest optimizations, which they could imagine to enhance the usefulness of the system.

8. RESULTS

Due to the preliminary character of the study, the participants were asked to focus on the accurate selection instead of performing the task as fast as possible in order to have the opportunity to understand the visual effects in more detail. As a result, task completion times varied significantly and did not produce a reliable result regarding efficiency of any of the cues. Due to occlusion as very powerful depth cue in this scenario, the evaluation of accuracy alone did not provide better results either. During the trials, tracking data of the participants was recorded, but as assumed before, the tracking-system didn't produce results robust enough for an in-depth comparison (Fig. 5). In addition to the inaccurate tracking method, the simple task and missing constraints for the participants are reasons for very high variance in the results. However, the qualitative analysis of the subjective rating by the test subjects and their comments about the concept revealed valuable insights, hints for possible enhancements, and also several issues and directions for further improvements. The results and comments indicate that the provided cues have some positive impact on depth judgment. 21 Participants (87.5%) rated at least one effect as enhancement compared to the standard 3D representation. The overall design of the effects was appreciated by the users. Although some were irritated due to the overemphasizing of the visualization in the current implementation, only 6 (25%) participants preferred the standard projection without additional depth cues regarding the overall visual impression. Tweaking some parameters therefore is likely to enhance the usability and acceptance of the proposed effects. The ranking regarding the interference factor of the effects shows that users were less distracted by Crepuscular Rays and Color Perspective effects as by the Depth-of-Field effect, which was rated as less distracting than other effects by only 3 participants (12.5%). The Depth-of-Field effect and the Crepuscular Rays were rated to be more useful than the Color Perspective effect and absence of any additional visualization. The participants also mentioned the use of occlusion as the main cue while being in close proximity to the object. As soon as the virtual hand disappeared behind it, they knew that the correct depth had been reached. It was also stated, that the usefulness of the effects might be increased by extending Depth-of-Field and Color Perspective to three dimensions instead of limiting them to the depth dimension only.

Fig. 5 Left: Scene used in the survey. Right: Example trajectory of distance relative to the target.



9. LESSONS LEARNED AND FUTURE RESEARCH

The observations in the preliminary survey reveal that the used tracking technology is one of the main points for optimization. More accurate tracking methods are necessary both for a reliable evaluation of the effects and for the proposed techniques to leverage their entire potential since they are target precise selection and manipulation tasks. Another weak point was represented by the observation that simple grasping of whole objects with the virtual hand does not represent an optimal evaluation strategy as this could be easily achieved by using occlusion as main depth cue without critical impact on the overall performance. Therefore, a more sophisticated task is intended to be used for future evaluation. Possible tasks include the manipulation of small parts or vertices of an object's surface in order to align the surface to given control points or exact positioning tasks in which neither occlusion or collision cues could influence the performance. Additionally, with accurate tracking methods and a more substantial training for the participants, the evaluation task performance regarding time needed for selection could be used to quantify the influence of the proposed visualizations. Currently, the developed prototype is being adapted for usage with the Leap Motion (Leap Motion 2013), which promises to deliver far better tracking accuracy but also requires a very different system setup. Regarding the visualization, mainly two issues are subject to further improvements. One of the most prominent complaints about the visualizations, especially the Depth-of-Field and Color Perspective, was related to the limited possibilities for the user to orient themselves in the scene. Although the proposed techniques were used to support selection tasks, they interfere with orientation and navigation tasks. When entering the virtual environment, users first need to orient themselves. In this phase, the proposed artificial depth cues did receive negative feedback because users were not able to immediately gain an overview of the complete scene. Therefore, the possibility to activate and deactivate the visualizations depending on the current task or interaction context is necessary. The most promising solution is to restrain the Depth-of-Field and Color Perspective effects to the area nearby the virtual hand only, similar to the Crepuscular Rays, which only affects the direct surroundings of the user's hand position. The second issue is related to the adaptation of the design of the proposed artificial depth cues. Subject to further optimizations is the reduction of the intensity of the visualizations in order to produce more subtle effects. Future research directions also include the exploration of additional options for depth visualization and variations of the proposed concepts.

ACKNOWLEDGEMENTS

We thank all the participants of the survey. This research has been funded by the European Social Fund and the Free State of Saxony (project no. 100076040).

REFERENCES

Cutting, James E., and Peter M. Vishton. "Perceiving Layout and Knowing Distances." In *Perception of Space and Motion*. Edited by William Epstein and Sheena Rogers, pp. 69–117. Elsevier, 1995.

Domini, Fulvio, and Corrado Caudek. "3-D structure perceived from dynamic information: a new theory." *Trends in Cognitive Sciences 7*, nº. 10 (2003): pp. 444-449.

Goldstein, E. B. *Sensation and perception*. 8th ed. Belmont, CA: Wadsworth, Cengage Learning, 2010.

Gombrich, E. H. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. London: Phaidon Press, 2002.

Held, Robert T., Emily A. Cooper, James F. O'Brien, and Martin S. Banks. "Using blur to affect perceived distance and size." *ACM Transactions on Graphics* 29, nº. 2 (2010): pp. 1–16.

Jones, J. A., J. E. Swan, Gurjot Singh, Eric Kolstad, and Stephen R. Ellis. "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception." In *Proc. APGV '08*, pp. 9–14. ACM Press, 2008.

Leap Motion, Inc. "LEAP Motion.". https://www.leapmotion.com/ (2013).

Mather, George, and David R. R. Smith. "Blur discrimination and its relation to blur-mediated depth perception." *Perception 31*, nº. 10 (2002): pp. 1211–1219.

Mitchell, Jason L. "Light Shaft Rendering." In *ShaderX3: Advanced rendering with DirectX and OpenGL*. Edited by Wolfgang F. Engel, pp. 573–88. Hingham, Mass: Charles River Media, 2005.

Palmer, Stephen E. *Vision science: Photons to Phenomenology*. Cambridge, Mass: MIT Press, 1999.

Piryankova, Ivelina V., Stephan de La Rosa, Uwe Kloos, Heinrich H. Bülthoff, and Betty J. Mohler. "Egocentric distance perception in large screen immersive displays." *Displays* (2013).

Scheuermann, Thorsten. "Advanced depth of field." GDC 2004 (2004).

Seth, Abhishek, Judy M. Vance, and James H. Oliver. "Virtual reality for assembly methods prototyping: a review." *Virtual Reality 15*, nº. 1 (2011): pp. 5–20.

Thompson, William B., Peter Willemsen, Amy A. Gooch, Sarah H. Creem-Regehr, Jack M. Loomis, and Andrew C. Beall. "Does the Quality of the Computer Graphics Matter when Judging Distances in Visually Immersive Environments?" *Presence: Teleoperators and Virtual Environments 13*, nº. 5 (2004): pp. 560–571.

Vishwanath, D., and E. Blaser. "Retinal blur and the perception of egocentric distance." *Journal of Vision 10*, nº. 10 (2010).

Wanger, L.R, J.A Ferwerda, and D.P Greenberg. "Perceiving spatial relationships in computer-generated images." *IEEE Computer Graphics and Applications* 12, nº. 3 (1992): pp. 44–58.

Wann, J., S. Rushton, and M. Mon-Williams. "Natural problems for stereoscopic depth perception in virtual environments." *Vision Research 35*, nº. 19 (1995): pp. 2731–2736.

Watt, S. J., K. Akeley, M. O. Ernst, and M. S. Banks. "Focus cues affect perceived depth." *Journal of Vision 5*, nº. 10 (2005): 7.

Willemsen, Peter, Amy A. Gooch, William B. Thompson, and Sarah H. Creem-Regehr. "Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments." *Presence: Teleoperators and Virtual Environments 17*, nº. 1 (2008): pp. 91–101.