Investigating Incoherent Depth Perception Features in Virtual Reality using Stereoscopic Impostor-Based Rendering

Kristoffer Waldow *
TH Köln, Computer Graphics Group
FAU Erlangen-Nürnberg

Lukas Decker†
TH Köln, Computer Graphics Group

Martin Mišiak ‡
University Würzburg, HCI Group

Arnulph Fuhrmann *
TH Köln, Computer Graphics Group

Daniel Roth§
Technische Universität München
Klinikum rechts der Isar
Orthopedics and Sports Orthopedics

Marc Erich Latoschik ¶
University Würzburg, HCI Group

ABSTRACT

Depth perception is essential for our daily experiences, aiding in orientation and interaction with our surroundings. Virtual Reality allows us to decouple such depth cues mainly represented through binocular disparity and motion parallax. Dealing with fully-mesh-based rendering methods these cues are not problematic as they originate from the object's underlying geometry. However, manipulating motion parallax, as seen in stereoscopic imposter-based rendering, raises multiple perceptual questions. Therefore, we conducted a user experiment to investigate how varying object sizes affect such visual errors and perceived 3-dimensionality, revealing an interestingly significant negative correlation and new assumptions about visual quality.

Index Terms: Human-centered computing—Visualization—Emipirical sutdies in visualization; Applied computing—Law, social and behavioral sciences—Psychology;

1 INTRODUCTION

Depth perception is an important part of our everyday experiences. It plays a crucial role in orientation and interaction with our surroundings. Within the field of virtual reality (VR), it is essential for VR systems to match the quality of depth perception experienced in the real world to create the most immersive experience possible and increase presence [1, 4].

The two most prominent depth cues, which are also considered by VR systems, are binocular disparity and motion parallax. While both of these cues are inherently linked together in the real world, VR offers us the possibility of decoupling them to gain more insight into the workings of human depth perception [2]. One example for the decoupling of binocular disparity and motion parallax are stereoscopic impostors [3]. This image-based rendering technique provides binocular disparity cues by showing two pre-generated images of an object to the corresponding eyes of the viewer. Correct motion parallax cues are provided by raymarching a heightfield, which is associated with the depth map of a pre-generated image. Modifying the scale of the depth map controls the amount of motion parallax that is generated without affecting binocular disparity.

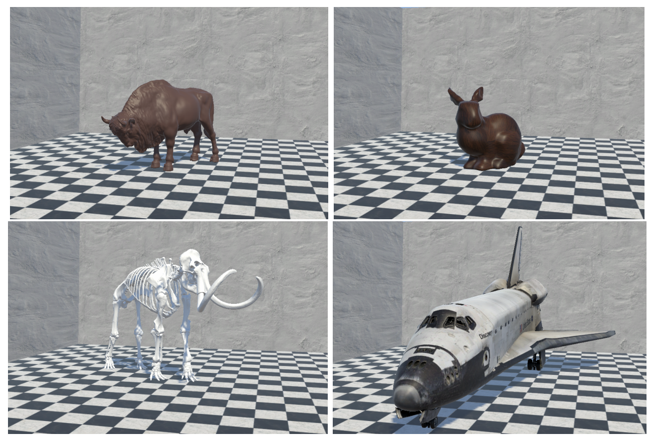

Figure 1: Overview of our different models used in our experiment: Bison (*top-left*, only for introduction), bunny (**B**), mammoth (**M**) and space shuttle (**S**).

As most image-based techniques, the impostor algorithm is prone to disocclusions in areas which were previously not visible in the pre-generated image. While small disocclusions can be mitigated via inpainting, disocclusions spanning a large depth extent are commonly discarded. Reducing the scale of the depth map can help lower the disocclusion probability. However, it will also reduce the amount of generated motion parallax, creating a cue inconsistency with binocular disparity.

This paper investigates the effects of reduced motion parallax on the visual quality of the stereoscopic impostors technique. To this end, a user study is conducted where participants rate the perceived 3-dimensionality, as well as the visual error of different objects rendered using a reduced amount of motion parallax.

2 EXPERIMENT

Our experiment ($n = 16$) used a within-subject design and took place in a large room with $40\,m^2$ where the subjects could walk freely while wearing a Meta Quest 2 head mounted display. Three different objects with varying topological structure are presented to the subjects. Each object impostor was rendered with full binocular disparity cues, but using 11 different depth map scale values, which translate to motion parallax gains ranging from 0 to 1.0 in 0.1 steps.

After each iteration, the subjects were asked to rate the visual quality of the object representation. In the context of our experiment, we defined visual quality as an orthogonal combination of

*e-mail: {prename}.{surname}@th-koeln.de
†e-mail: ldeckerluss@gmail.com
‡e-mail: martin.misiak.89@gmail.com
§e-mail: daniel.roth@tum.de
¶e-mail: marc.latoschik@uni-wuerzburg.de

3-dimensionality and visualization error. A 7-point Likert scale (1=strongly disagree, 7=strongly agree) was used to answer the two questions: "*The 3-dimensionality of the object convinced me*" and "*The object had strong visualization errors*".

The term *3-dimensionality* refers to the subjective depth of the object and is simplified so that test subjects without prior knowledge of computer graphics can describe it. The term *visualization error* should refer to the overall image quality of the displayed object and accounts for various forms of systematic distortions, which are discriminable in a no-reference quality assessment.

2.1 Conditions & Procedure

For our experiment, we chose four different objects varying in size, shape and depth complexity. We chose known shapes to give subjects a reference and positioned and rotated them in a way to have different dimension in the z-axis from the subject's point of view (z_I, z_M, z_B, z_S). The selected objects are the bison (**I**, $z_I = 2.71$), mammoth (**M**, $z_M = 3.13$), bunny (**B**, $z_B = 1.81$) and the space shuttle (**S**, $z_S = 9.05$) (see Fig. 1). The bunny object represents the best behaved model, as it has the lowest depth extent and its geometry has also a low disocclusion potential. The skeleton structure of the mammoth model ensures a very high disocclusion potential, however, due to its lower depth extent, most visible disocclusions are covered by inpainting. The space shuttle model has the highest depth extent out of the selected test objects. Due to this, a majority of disoccluded areas are discarded instead of inpainted.

Our experiment started with a training session and the bison model to create a baseline and to have an example of clear occlusion artifacts on the legs. To trigger motion parallax cues, the subjects were motivated to move on a pre-defined path while viewing the impostor. Subjects were guided by an unintrusive visual indicator in their lower visual field. The main session involved the rating of the bunny, mammoth and space shuttle objects. Once an object was randomly selected, all of its ten motion parallax levels were presented in random order.

3 Results

We conducted multiple Spearman's correlation tests to evaluate the relationship between motion parallax, perceived 3-dimensionality and visual error. The results reveal a significant negative relationship between perceived 3-dimensionality and visual error, $r_s(528) = -.506, p =< .001$ but no significant correlation of motion parallax and visual error ($r_s(528) = .012, p = .786$). However, a significant relation is found for motion parallax and 3-dimensionality, $r_s(528) = -.309, p =< .001$. Split up by the objects, a Friedmann significance test with related samples shows a significant difference between perceived 3-dimensionality and visual error, $\chi^2(2) = 52.494, p < .001$. The mammoth and bunny show a negative correlation between both measured aspects of visual quality - $r_{s_M}(11) = -.865, p_M =< .001$, $r_{s_B}(11) = -.610, p_B =< .001$. The space shuttle, on the other hand, has a positive correlation between 3-dimensionality and visual error, $r_{s_S}(11) = .761, p_S = .006$. For the bunny model, a final correlation test is performed between the parallax values of the motion greater than 0.5 and the visual error. No correlation was found, $r_{s_B}(288) = .01, p_B = .868$.

3.1 Discussion

Against our expectations, the visual error did not increase with motion parallax for all objects. Only the space shuttle model showed higher visual error for stronger motion parallax scales. The bunny and mammoth models behaved inverted in terms of visual error (see Fig. 2). 3-dimensionality correlated positively with motion parallax for all tested objects, although their relationships seems to be nonlinear. For example, the bunny model reached diminishing returns at a parallax value of 0.5. In contrast to our initial assumption, 3-dimensionality and visual error are not orthogonal aspects of visual quality. They are strongly correlated, and the direction of the relationship depends heavily on the object type. A possible explanation could revolve around the use of inpainting to cover disoccluded fragments. When most fragments are discarded, as seen on the space shuttle model, the disocclusions are interpreted as a visual error. With most fragments inpainted however, visual error is interpreted differently as no visible "gaps" are observable. In addition to this, it is also possible that the term "visualization error" covers a wide semantic range, that could be considered too vast. As a result, it was sometimes unclear to the subjects what kind of error to look for.

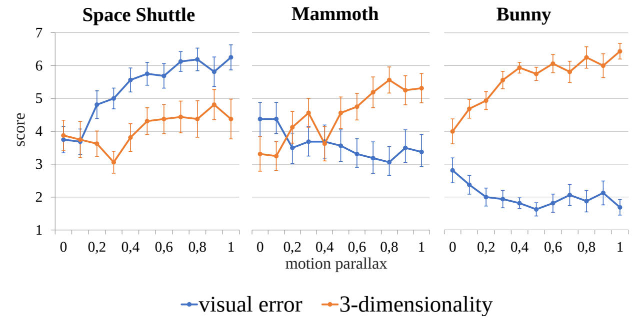

Figure 2: Overview our mean results (±SEM) of motion parallax split up by the objects. Negative correlations were found for mammoth and bunny and positive one for the space shuttle. (1= strong disagree, 7 = strong agree)

4 Conclusion & Future Work

In conclusion, a user experiment was performed to evaluate the influence of inconsistent depth cues on the visual quality of a VR application. The stereoscopic impostors technique was used to render objects with full binocular disparity but with reduced motion parallax. While the *3-dimensionality* of objects increased with higher levels of motion parallax, the correlation direction of visual error and motion parallax was strongly dependent on the observed object. Our experiment highlights the need for a unified definition of visual quality in VR, as well as a corresponding set of questions that consider classical image quality aspects, as well as VR specific aspects of depth perception.

Acknowledgments

This work was partially funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia under grant number 005-2105-0046 as part of the project KoViTReK.

References

[1] W. IJsselsteijn, H. De Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman. Perceived depth and the feeling of presence in 3dtv. *Displays*, 18(4):207–214, 1998.

[2] B. Julesz. Binocular depth perception without familiarity cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis. *Science*, 145(3630):356–362, 1964.

[3] M. Misiak, A. Fuhrmann, and M. E. Latoschik. Impostor-based rendering acceleration for virtual, augmented, and mixed reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, VRST '21. ACM, New York, NY, USA, 2021.

[4] K. Waldow, A. Fuhrmann, and S. M. Grünvogel. Investigating the effect of embodied visualization in remote collaborative augmented reality. In *International Conference on Virtual Reality and Augmented Reality*, pp. 246–262. Springer, 2019.