https://doi.org/10.2352/EI.2023.35.2.SDA-383

Evaluating the angular resolution of a simulated light field display in regards to three-dimensionality, motion parallax and viewing experience

Sophie Kergaßner, Stuttgart Media University, Germany
Jan Fröhlich, Stuttgart Media University, Germany

Abstract

This paper investigates how the angular resolution of a simulated light field display affects the viewing experience of a 3D scene. Subjective tests are conducted in a user study, in which questions are asked to evaluate the effect of three-dimensionality, recognizability of shapes, and viewing pleasantness. A total of 32 results are recorded and evaluated. The quality metrics show a logarithmic trend, improving with an increasing number of views. A plateau at 0.25° per view can be found in two of the three evaluated questions. Moreover, the participants preferred the viewing experience of a regular 2D-Display over the simulated light field displays with more than 2° per view. Subjective ratings had a significant standard deviation, which lowers their validity/significance. It also indicates that users had difficulties quantifying their experience.

Introduction

For some visual applications, it is necessary to create a three-dimensional impression for multiple, freely moving people simultaneously. In these cases, conventional 2D, 3D, or even autostereoscopic 3D displays would be insufficient. However, they can be covered by using light field displays (LFDs).

Some information we naturally perceive in the real world cannot be represented by displays [1]. In a three-dimensional world, four phenomena arise: stereo parallax, motion/movement parallax, accommodation, and convergence [1]. To make 3D images look realistic, plenty of the four phenomena have to be recreated. All 3D displays show at least a stereo parallax [1]. Some are capable of creating a motion parallax.

The Importance of the Motion Parallax

In a three-dimensional world, the body subconsciously performs small movements to examine the volume properties, and surface structure of a viewed object more closely [2][3]. The resulting motion parallax is incorporated into the perception of the object [4]. Traditional stereo 3D displays show a single static source image per eye. Since these two views do not change when the head moves, the synthetic three-dimensional image appears less natural than the real-world surroundings, so the illusion decays [2]. Therefore, it is not sufficient to show only a single static source image per eye.

Multiview-Displays

A multiview display is characterized by the fact that it can show more than just two source images [5]. The different views (viewing angles of a scene) are cast into different directions in space (Fig. 1).

A different view is visible depending on the angle from which the viewer looks at the display. If the viewer moves, the individual views merge, and the observed image changes according to the viewer's position. In this way, the motion parallax can be simulated [6].

Some (auto)stereoscopic displays are capable of simulating a multiview display. A way to do that is by tracking the viewer's position and rendering a corresponding stereo image in real-time. A limiting factor for this method is the number of simultaneous users. Hence, these displays are often incapable of showing a correct view to multiple users simultaneously or require wearables.

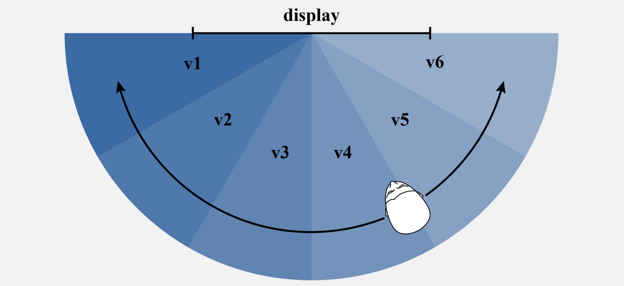

Figure 1: Schematic representation of a multiview display with 180° field of view and 6 views, therefore 30° per view [7].

Light Field Displays

LFDs, on the other hand, are both multiview and multi-user capable [3]. They reproduce the light field of a scene [8] in discrete regions of the viewing space. Fig. 2 shows an example of a person's view on a display. A different subpixel is perceived depending on the angle from which a pixel is looked at.

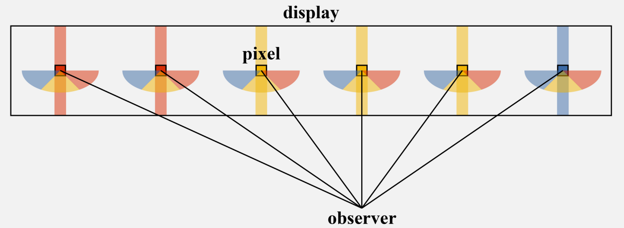

Figure 2: A person's view on the display: The colored vertical stripes indicate the perceived subpixel.

The Role of the Angular Resolution

The angular resolution of an LFD is determined by the ratio of its field of view (FOV) to its number of views [9]. It has a fundamental influence on the quality of the viewing experience [3][9]. The lower the angular resolution, the more views are rendered in a given FOV and the less noticeable the discrete presentation of the motion parallax becomes. If the angular resolution is too large, the images produce visible vertical slices that either transition abruptly or blend with double contours if their viewing ranges overlap [3] (Fig. 3).

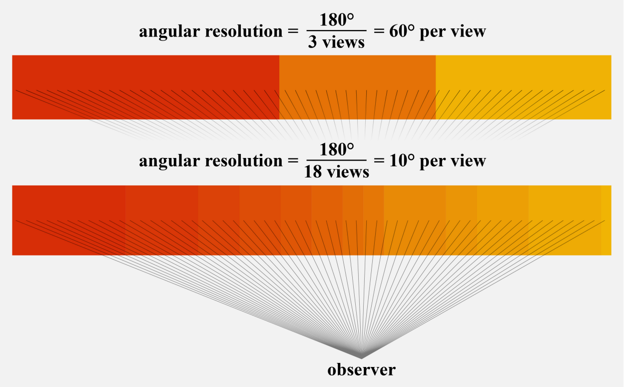

Figure 3: Light field displays with 3 and 18 views: The visual difference between the source images (indicated by color) decreases with a greater number of views.

The number of physical subpixels required to maintain a given image resolution grows linearly with the number of views, and the processing required to drive the display is affected similarly. Since the research and manufacturing of low angular resolution LFDs and the production of content are costly, it is worthwhile to weigh the benefit of a low angular resolution in advance.

The work presented in this paper aims to simulate an LFD in order to examine a number of angular resolutions in subjective tests. Factors, such as the perceived three-dimensionality, the recognizability of an object's shape, and the overall viewing pleasantness, are considered.

Related Work

In 2016, Kara et al. investigated the subjectively perceived quality of LFDs with different angular resolutions [10]. In the investigated range of 15–150 views at 40 degrees FOV (2.666°per view (pv) to 0.266°pv), the quality ratings unexpectedly reached a local maximum at a resolution of 0.533°pv (75 source images) [10]. The authors did not consider the results to be convincing since the evaluations showed strong inconsistencies and differed significantly from participant to participant [10].

One problem with new technologies is that a large proportion of the test subjects have often never been in contact with the corresponding medium (75% of the test subjects in [10]). Participants often find it difficult to assess these completely new visual experiences, and the metrics tend to be unreliable [10]. To get familiar with the technology, participants went through a training phase in advance [10]. The authors suspect that the study would have benefitted from a more elaborate training phase [10].

In 2017, Viola et al. evaluated diverse investigation techniques of LFDs [8]. They constructed four different real-world scenes in two sub-studies [8]. In the first study, subjects actively interacted with the image, while in the second study, they were shown a video sequence with pre-recorded motion [8]. Viola et al. concluded that it was desirable for the researchers to let the subjects move around freely and let them interact with the LFD rather than presenting a pre-recorded animation of the display [8]. Consequently, it is suggested to either regulate the movement in space (e.g. by a given path) or to track the subject's movement so that conclusions about their behavior can be drawn afterward [8].

In 2010, Hassaine et al. performed two studies on path-tracking tasks, in which they investigated a number of viewpoint densities (0–50 views/10cm at the eye) and scene depths (2cm and 10cm), with and without stereopsis present as a depth cue [11]. Even though Pastoor and Schenke (1989, as cited in [11]) suggest viewpoint densities as high as 100 views/10cm (translates to 0.029°pv at 2m viewing distance), the ratings determined by Hassaine et al. show a maximum at 8 views/10cm (0.358°pv) [11]. They state that the effect of the motion parallax was much weaker than the effect of stereopsis since it only affects the response latencies and not the response correctness [11]. The response latencies decreased significantly (by 21%) with increasing numbers of views (from 2 views/10cm (1.432°pv) to 8 views/10cm (0.358°pv)), but don't decrease further at 50 views/10cm (0.057°pv) [11].

Methodology

The following research is based on a simulated, idealized, autostereoscopic LFD that is built using a regular single-view 2D display. The content is created by tracking the subject's position and rendering the corresponding image in real-time. In the following, this study setup and the creation of the stimuli will be described.

Dividing the FOV among the Subpixels

In this research, views are only generated along the horizontal axis. On the one hand, most of a person's movement in space occurs on a plane at a constant height – only a change in height would result in vertical movement. On the other hand, the required number of views grows quadratically rather than linearly if two axes are considered. This increases the number of subpixels and the runtime complexity of rendering the image.

Several subpixels are assumed to form each regular pixel (Fig. 4). Each subpixel has a defined angular emission range and an ideal emission pattern (Fig. 5).

The field of view can be divided among the subpixels in various ways. If the subpixel emission pattern is irregular (every pixel has a unique distribution pattern), it is possible to generate artifact-free and perspective-correct views in certain parts of the viewing environment (Fig. 6). However, the total area of the viable viewing zones decreases with an increasing number of views. The source images are perspective renderings.

While moving within a viewing zone, no motion parallax occurs. The transition between the viewing zones is abrupt, especially with a low number of views. This also applies to the second distribution variant.

If the FOV is divided regularly and identical per pixel, a perspective-correct image can be reproduced at any point in space, assuming an infinite number of views. This means that there are no sweet spots, and the viewer is not restricted in their movement.

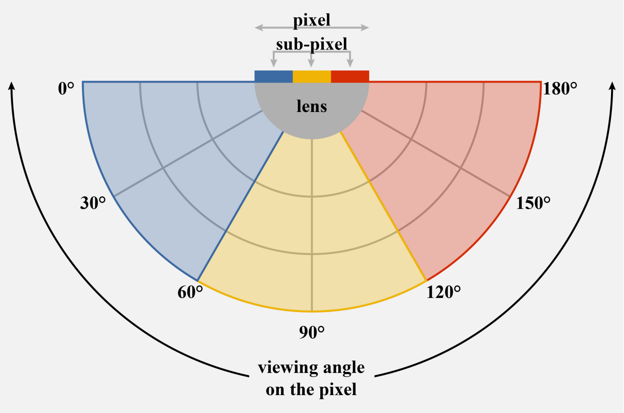

Figure 4: Top-down view on a pixel with 3 horizontally aligned subpixels. The shaded areas indicate which subpixel will be seen regarding the viewing angle.

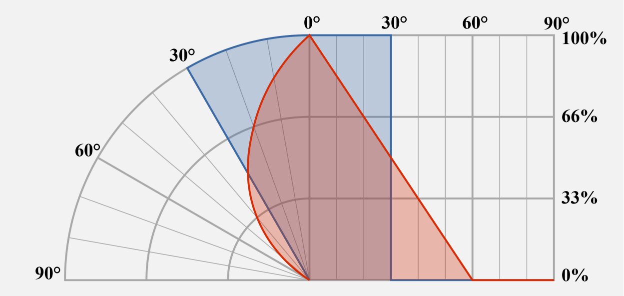

Figure 5: The diagram shows the emitted light intensity corresponding to the viewing angle. Blue shows an emission pattern without decay, red shows a linear decay. The angle is shown relative to the angle bisector of the emission range. Both halves show the same pattern using a curvilinear grid on the left and a rectilinear grid on the right.

The observer sees multiple source images simultaneously next to each other on the display (Fig. 2, 3, 7). In this distribution variant, the source images are parallel renderings rather than perspective renderings.

The lower the angular resolution, the more source images are visible simultaneously and, consequently, the smaller the width of the visible stripes gets (Fig. 3). Additionally, two source images, whose perspective angles differ slightly, are visually similar. The change between adjacent stripes becomes less noticeable, and the impression of the image becomes more uniform. Hypothetically, a display with an infinite number of views would render a perspective-correct image from every viewing position.

The simulated emission pattern of the subpixels (Fig. 5) ensures no luminance loss within a viewing area. This assumption is made to simplify the simulation, as an ideal LFD would have these characteristics.

Creation of the Stimuli

The stimuli simulate seven different LFDs with varying angular resolutions, ranging from 180°pv to 0.125°pv. The display simulation is implemented in TouchDesigner. A Kinect (Xbox One) tracks the head position (`head:tx`, `head:tz`) of the test person in the room, from which the angle of view δ on the display pixels is computed. Corresponding to the viewing angle δ, alpha maps are generated that define whether a specific source image is seen at the pixel position (Fig. 8).

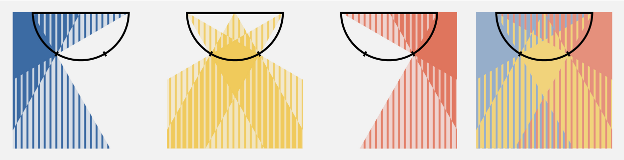

Figure 6: Irregular distribution of the FOV: Colored areas indicate which source images are seen. If the area is shaded solidly, a perspective-correct image can be seen.

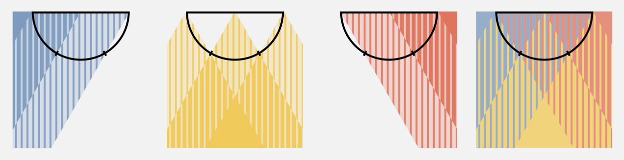

Figure 7: Regular distribution of the FOV: Colored areas indicate which source images are seen. If the area is shaded solidly, an orthographic image is seen.

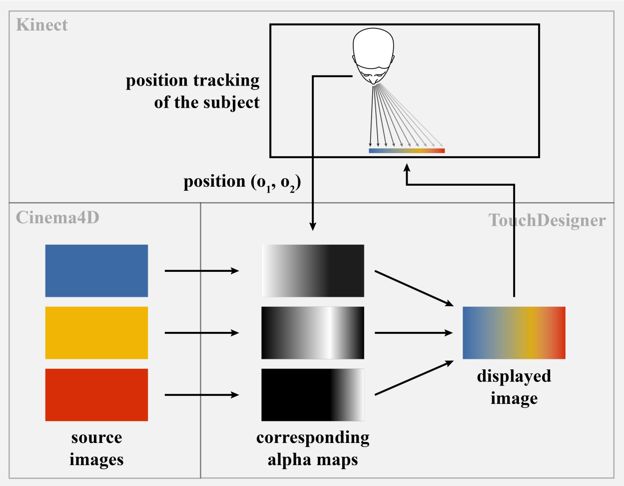

Figure 8: Schematic showing how software and hardware components interface with one another.

Calculation of the Alpha Maps

The calculation of the viewing angle δ is subject to the following formula:

$$\begin{aligned}\delta &= \begin{cases}\beta & \text{if } o_x \leq p_x\\ 180-\beta & \text{if } o_x > p_x\end{cases}\\ \beta &= \frac{180}{\pi}\arccos\left(\frac{|o_x-p_x|}{o_z}\right)\end{aligned}\tag{1}$$

where $\vec{o} = [o_x, o_z]$ is the position of the observer and $\vec{p} = [p_x, p_y]$ is the position of the observed pixel (Fig. 9).

The FOV is quantized into n discrete, equidistant steps, where n is the number of views. Therefore, each section has a width of $2b$ degrees, with $b = \frac{FOV}{2n}$. A pre-rendered source image $I^k_{src}(\vec{p})$ is assigned to all subpixels that emit in the same direction. For the decision, which subpixel source image $I^k_{src}(\vec{p})$ is visible

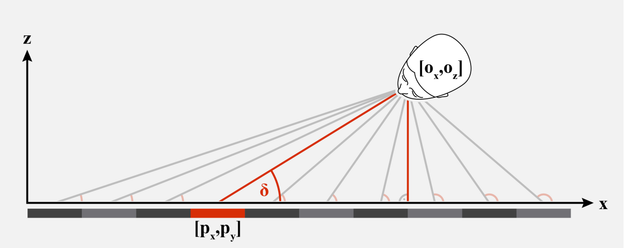

Figure 9: Computing the viewing angle δ. Top-down view.

from the viewer's position $\vec{o}$ on each display pixel $\vec{p}$, n alpha maps are generated. The values of the alpha maps are computed from:

$$I_a^k(\vec{o},\vec{p}) \;=\; \begin{cases} 1-\left|\frac{(2k-1)b-\delta}{2b}\right| & \text{if the outcome is} > 0 \\ 0 & \text{else} \end{cases}, \qquad (2)$$
$$k \in \mathbb{N}, 1 \leq k \leq n$$

The sum of all alpha maps $I_a^k(\vec{o},\vec{p})$ at a given pixel $\vec{p}$ is 1. The value increases linearly from $(2k-3)b°$, the direction (meaning the angle bisector of the emission range) of the neighbor subpixel, to $(2k+1)b°$, the exact direction in which the subpixel itself is pointing and decreases again similarly towards the next neighbor subpixel. This mirrors the emission pattern shown in Fig. 5.

The displayed output image $I(\vec{o},\vec{p})$ is the sum of all source images $I_{src}^k(\vec{p})$, weighted by their corresponding alpha map $I_a^k(\vec{o},\vec{p})$, and is generated as follows:

$$I(\vec{o},\vec{p}) = \sum_{k=1}^{N} I_{src}^k(\vec{p}) \cdot I_a^k(\vec{o},\vec{p}) \;, \in \mathbb{R}^3 \qquad (3)$$

Rendering of the Source Image Sequence

The source images are rendered in Cinema4D. To generate them efficiently, a camera with parallel ray incidence moves around the object semicircularly while facing it. The length of the render sequence (frames) corresponds with the number of views.

Due to viewing angles $\neq 90°$, a horizontal compression occurs, which is corrected in-camera via distortion maps (Fig. 10). The distortion maps are automatically generated as an EXR-sequence by a python-script.

Figure 10: Object rendered without distortion map (left) and with distortion map (right).

Content Decisions for the Stimuli

The stimuli show a minimalist object consisting of basic geometric shapes (Fig. 11). Occlusion within the object occurs due to its interweaved design. Most of the surfaces are solid blue with 50% roughness, the spheres and a thin plate at the front of the object have a mirror texture. The object floats in a minimal space that extends to the back, where luminous segments emphasize the edges. The scene is lit by an interior HDRI and three area-lights.

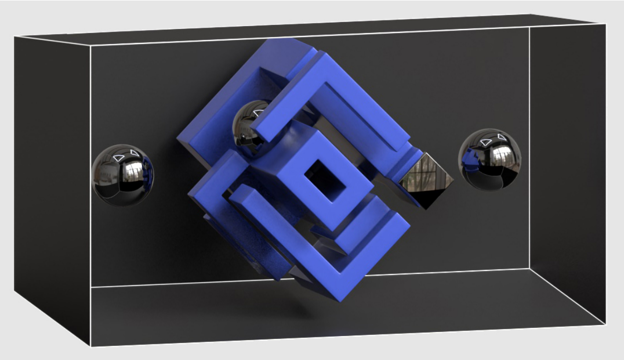

Figure 11: Rendering of the Cinema4D scene, showing the room-extension and the floating stimulus.

User Study

The study was conducted from 31.05.2022–02.06.2022 at the Stuttgart Media University, Germany. A total of 32 subjects with an average age of 24.9 years participated. 14 participants were female and 18 were male. A quantitative survey was performed to determine a mean opinion score (MOS) (1: bad, 2: poor, 3: fair, 4: good, 5: excellent) for the perception of several aspects of a simulated LFD at varying angular resolutions.

The subjects enter a room with a masked area in which they can move freely, but are instructed to move from corner to corner at least once per stimulus. The simulated LFD is installed at eye level and has a size of 31.5" (LG 32BN67U-B). In the beginning, there is a short training phase in which the 45 and 1440 views stimuli are shown to guide as a reference. Afterward, all seven stimuli are presented in random order. The evaluated numbers of views are 1, 45, 90, 180, 360, 720, and 1440.

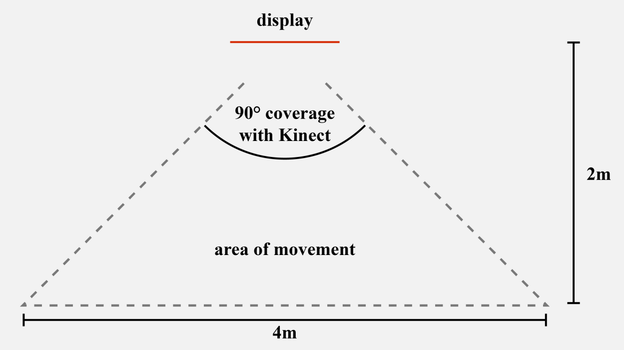

Figure 12: Top-down view of the study setup.

Results and Discussion

The following graphs show the evaluations of the stimuli per number of views. The dots in the scatter chart represent the average rating with its associated standard deviation (SD). The area of the dots in the bubble chart is proportional to the number of people, that gave the corresponding MOS rating. The FOV has a width of 180° horizontally and the numbers of views are 1 (180°pv), 45 (4°pv), 90 (2°pv), 180 (1°pv), 360 (0.5°pv), 720 (0.25°pv) and 1440 (0.125°pv).

How three-dimensional does the object appear?

This question examines the subjective rating on how real the presented object behaves in comparison to the real-world surroundings. It is neither about the quality of rendering, nor the effectiveness of the simulated motion-parallax, but on the overall appearance of the object in the setup.

Fig. 13 shows that the ratings align mostly in a logarithmic trend, with an average rise of 0.42 points (p) per step, improving with an increasing number of views. The stimuli that show more than 1 view result in a higher MOS than the 1 view stimulus. The largest increase in MOS can be found between 180 and 360 views (0.67p), while the smallest increase is located between 720 and 1440 views (0.07p). At the same time, these two stimuli provocate the second-largest and lowest SD, with 0.84p at 720 views and 0.49p at 1440 views. This indicates that the probands were more confident in rating 1440 views with a good score, compared to the evaluation of 720 views.

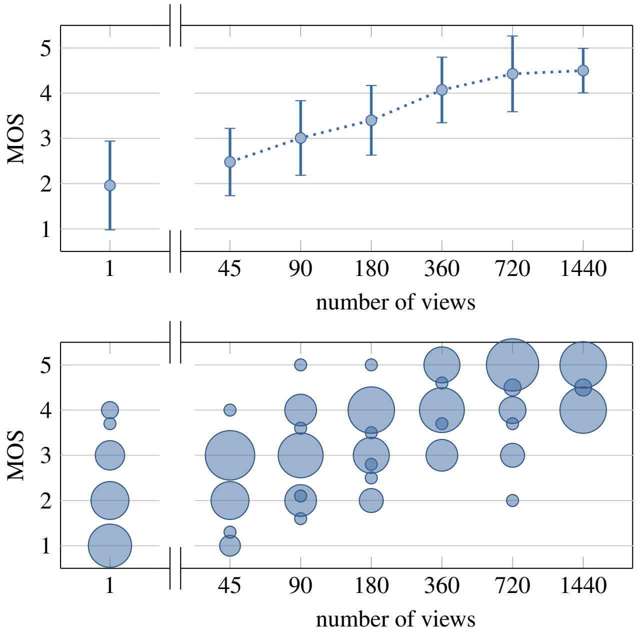

Figure 13: MOS per number of views determined for the question ”How three-dimensional does the object appear?”.

Does the position-dependent appearance of the object help to recognize its shape?

For this question, the subjects should focus on the motion parallax and evaluate by standing still vs. moving, if they gain more information about the object.

Again, the ratings show nearly a linear rise in MOS on the log scale between 45 and 720 views with 0.44p increase per step on average (Fig. 14). A plateau forms between 720 and 1440 views, there is even a slight decline from 4.53p to 4.51p. This indicates, that the participants might not have recognized a difference between these two stimuli, since objectively the higher number of views offers fewer artifacts and therefore more information about the shape. The three stimuli with the lowest angular resolution (360–1440 views) have an average SD of 0.64p, the stimuli from 45 to 180 views have an average SD of 0.97p.

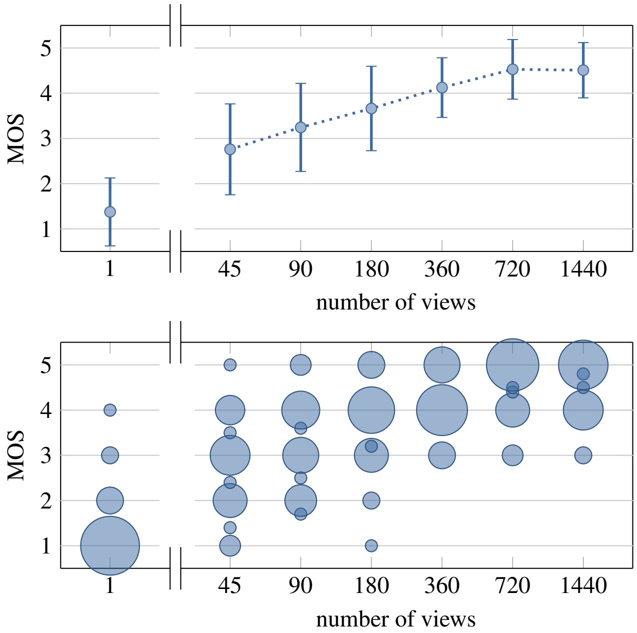

Figure 14: MOS per number of views determined for the question ”Does the position-dependent appearance of the object help to recognize its shape?”.

How pleasant is the viewing experience?

This question is subjective and emotional as well since even an artifact-free image is worthless when it’s distracting to look at it.

The 1 view rating stands out since the participants agreed to rate it decently, with 3.16p on average, which is greater than the rating of 45 views (1.51p), 90 views (2.02p), and 180 views (2.4p) (Fig. 15). Nonetheless, participants’ ratings fluctuated a lot, which can be seen at the high SD (1.17p).

A large increase in MOS by 1.2p is located between 180 (2.4p) and 360 views (3.6p). The 1440 views stimulus got the highest MOS with 4.66p and also the smallest SD with 0.46p. When comparing this observation to the ratings of the other examined factors, questions raise. In terms of the viewing experience, the subjects prefer the lowest angular resolution (1440 views) over 720 views, even though they seem to recognize little to no improvements in motion parallax (Fig. 14) and three-dimensionality (Fig. 13).

Limitations

Overall, the metrics are noisy. The average standard deviations range from 0.74p to 0.8p. This could be due to several reasons, some of them being: the small sample size of 32 participants, the length and setup of the training phase, the varying familiarity with the technology between participants, and the different movement patterns in space of the subjects. Moreover, it would be interesting to examine even lower angular resolutions and to lower the step size between the examined resolutions. Biases due to the shape of the shown object, the choice of the

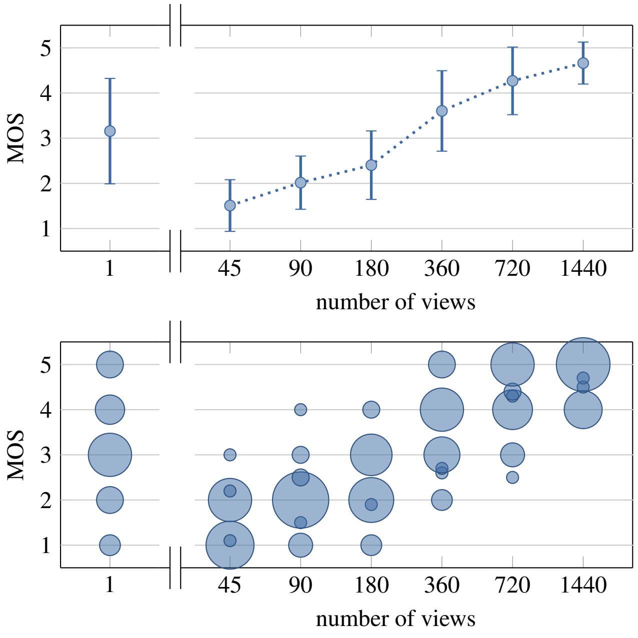

Figure 15: MOS per number of views determined for the question "How pleasant is the viewing experience?".

emission pattern (linear decline), and a learning effect in terms of the object's shape can't be verified, nor ruled out.

Conclusion

The quality metrics of the subjects mostly show a logarithmic trend, improving with an increasing number of views. A plateau at 720 views (0.25°pv) is noticeable in the ratings of the appearance of three-dimensionality and the ability to recognize shape, where the ratings even decline from 720 to 1440 views (0.125°pv). Regarding the viewing pleasantness, there is no such plateau as a rise between 720 and 1440 views is still noticeable.

The stimuli with more than 1 view were consistently rated higher in MOS than the frozen frame with 1 view in the appearance of three-dimensionality and the ability to recognize the shape. Nevertheless, in terms of viewing pleasantness, only 360 views (0.5°pv) and above were rated higher than the frozen frame.

The ratings show a significant standard deviation, reaching from 0.74p to 0.8p on average between the questions. This could be due to several reasons, specified in the limitations section, but they can't be verified, nor ruled out.

Acknowledgments

Thanks to the participating subjects and Prof. Dr. Simon Wiest, Laurent Gudemann, Prof. Dr. Martin Fuchs, Leonie Kergaßner, Achim Johann, Steffen Mühlhöfer, and Peter Ruhrmann.

References

[1] Neil A. Dodgson, *Autostereoscopic 3D Displays*, Institute of Electrical and Electronics Engineers (IEEE). (2005).

[2] Péter Kovács, Tibor Balogh, *3D Visual Experience*, Signals and Communication Technology. (2010).

[3] Peter A. Kara, Roopak R. Tamboli, Aron Cserkaszky, Maria G. Martini, Attila Barsi, Laszlo Bokor, *The Viewing Conditions of Light-Field Video for Subjective Quality Assessment*, International Conference on 3D Immersion (IC3D). (2018).

[4] W. Vanduffel, D. Fize, H. Peuskens, K. Denys, S. Sunaert, J. T. Todd, G. A. Orban, *Extracting 3D from Motion: Differences in Human and Monkey Intraparietal Cortex*, American Association for the Advancement of Science (AAAS). (2002).

[5] Oscar H. Willemsen, Siebe T. de Zwart, Martin G. H. Hiddink, Dick K. G. de Boer, Marcel P. C. M. Krijn, *Multi-view 3D Displays*, SID Symposium Digest of Technical Papers. (2007).

[6] Felipe M. L. Ribeiro, José F. L. de Oliveira, Alexandre G. Ciancio, Eduardo A. B. da Silva, Cássius R. D. Estrada, Luiz G. C. Tavares, Jonathan N. Gois, Amir Said, Marcela C. Martelotte, *Quality of Experience in a Stereoscopic Multiview Environment*, IEEE Transactions on Multimedia. (2018).

[7] Peter A. Kara, Aron Cserkaszky, Subbareddy Darukumalli, Attila Barsi, Maria G. Martini, *On the edge of the seat: Reduced angular resolution of a light field cinema with fixed observer positions*, Ninth International Conference on Quality of Multimedia Experience (QoMEX). (2017).

[8] Irene Viola, Martin Řeřábek and Touradj Ebrahimi, *Impact of interactivity on the assessment of quality of experience for light field content*, Ninth International Conference on Quality of Multimedia Experience (QoMEX). (2017).

[9] Peter A. Kara, Aron Cserkaszky, Attila Barsi, Tamas Papp, Maria G. Martini, László Bokor, *The interdependence of spatial and angular resolution in the quality of experience of light field visualization*, International Conference on 3D Immersion (IC3D). (2017).

[10] Peter A. Kara, Maria G. Martini, Peter T. Kovacs, Sandor Imre, Attila Barsi, Kristof Lackner, Tibor Balogh, *Perceived quality of angular resolution for light field displays and the validy of subjective assessment*, International Conference on 3D Imaging (IC3D). (2016).

[11] Djamel Hassaine, Nicolas S. Holliman, Simon P. Liversedge, *Investigating the Performance of Path-Searching Tasks in Depth on Multiview Displays*, ACM Transactions on Applied Perception. (2010).

Author Biography

Sophie Kergaßner received her B.Eng. in Audiovisual Media from the Stuttgart Media University in august 2021 and is currently studying a consecutive master's program (Audiovisual Media M.Eng.) with a focus on computer graphics. Her research interest includes computer vision, stereoscopic and light field displays, and perception.