Perceptual preference for 3D interactions and realistic physical camera motions on light field displays

Mary Guindy[a,b], Peter A. Kara[c,d], Tibor Balogh[a], and Aniko Simon[e]

[a]Holografika, Budapest, Hungary
[b]Pazmany Peter Catholic University, Budapest, Hungary
[c]Budapest University of Technology and Economics, Budapest, Hungary
[d]Kingston University, London, UK
[e]Sigma Technology, Budapest, Hungary

ABSTRACT

The scientific literature on light field visualization has recently started addressing the 3D interaction techniques for light field displays, as well as the simulations for realistic physical camera motions. Both of them are highly relevant to civilian and military use cases. In case of the latter, 3D interactions are used for various purposes, including the control of surface and air vehicles, strategic and tactical planning, and real-time support of operations. The topic of realistic physical camera motion is particularly important for field surveillance and the situation awareness of dismounted operators, as well as for the civilian use case of cinematography. Yet thus far, such techniques have not been perceptually evaluated by non-expert observers. While the theoretical feasibility has already been investigated and expert reviews have initiated the first steps, data on actual perceptual preference is still lacking. The term "expert" refers to scientists, researchers and manufacturing professionals. However, at the end of the day, the efficiency of use cases is fundamentally determined by the observers'/operators' perceptual convenience of 3D visualization and the effectiveness of interactions. In this paper, we present our results on the empirical studies carried out on perceptual preference regarding the potential interaction techniques and the different types of realistic physical camera motions. The contents of the subjective tests were displayed on a large-scale light field cinema system. Multiple subjective quality metrics were used to decompose the visual experience of the observers, and additional attention was paid to the essential aspects of long-term usage, such as dizziness.

Keywords: Light field visualization, realistic camera simulation, physical simulation, 3D interaction, perceptual preference, cinematography

1. INTRODUCTION

With the scientific advancements of our era, light field (LF) technology has emerged as a means of representing the 3D world – to which it acts as a window – by light rays filling up the 3D space under representation.[1] Although LF displays (LFDs) have been around for roughly two decades now, much research is still required for many of its applications, such as medical imaging, prototype review, education, gaming, digital signage, cinematography, telepresence and many more.

The term "light field" was first proposed by Gershun in 1936.[2] For the purpose of capturing and displaying LFs, light field cameras and displays have been invented, respectively of course. Although LFDs provide the spectators an immersive 3D experience, they have many limitations due to their physical and optical restrictions. These limitations impose numerous challenges when working with LFs.[3] Comprehensively overcoming such challenges enables the creation of various use cases and applications of LF technology.

Further author information: (Send correspondence to Mary Guindy)
Mary Guindy: E-mail: m.guindy@holografika.com / guindy.mary.mohsen.messak@ppke.hu
Peter A. Kara: E-mail: kara@hit.bme.hu
Tibor Balogh: E-mail: t.balogh@holografika.com
Aniko Simon: E-mail: aniko.simon@sigmatechnology.se

Among the novel applications designed around LFDs are cinematographic and interactive ones as well, both of which hold immense potentials for future utilization scenarios. These have already been addressed in the scientific literature on the levels of fundamental feasibility and expert review; however, no study on user preference has been carried out so far. By addressing such research questions, we may acquire feedback data that is essential to the long-term development of such applications.

In this paper, we present the results for the different series of subjective tests carried out to address user preference with regard to 3D interaction techniques and realistic physical camera motion on LFDs. The latter is a particularly relevant aspect of cinematographic research efforts. Additionally, perceptual effects and personal tolerance are also investigated by our work, as they greatly contribute to the overall user experience.

The remainder of the paper is structured as follows: Section 2 reviews the relevant works of the scientific literature. The experimental configuration of the tests are detailed in Section 3. The results are presented and discussed in Section 4. Section 5 concludes our findings and highlights potential continuations of this work.

2. RELATED WORK

Regarding the approaches for interaction techniques for LFDs, previously published works include the FOX (Focus Sliding surface) metaphor of Marton *et al.*[4] and the theater presentation model of Guindy *et al.*[5] In the latter paper, the idea is to combine the three main tasks of the different types of 3D interaction into a unified presentation model. These tasks are: (i) navigation, (ii) selection and manipulation and (iii) application / system control.[6] Based on the category of the LFD – full-parallax or horizontal-only-parallax (HOP) display – various challenges arise for each of the proposed interaction tasks, with the full-parallax LFDs being more challenging. The potential techniques and ideas for the different interaction tasks on LFDs are suggested and investigated by the work of Guindy *et al.*[5] The publication specifies the line-up, carousel, 3D sphere, CAD / CAM, medical and theater presentation models. From these models, the theater model was chosen, as its nature of presentation is rather analogous to LFDs in general. Among the obvious similarities is the angular-selective attribute of the theater model and LFDs; both fundamentally depend on the position of the spectator. In order to comply with the interaction presentation model, the three main interaction tasks are defined for the theater model:

1. Navigation: Done by means of a static camera.
2. Selection and manipulation: A set of techniques is deployed. These include the alterations of spatial positions of the objects, the usage of carousels, and hiding / revealing elements.
3. Application / system control: Done via switching between two scenes. One is the theater model itself and the other one is the monitor room with feedback to the current state of the theater scene.[5]

Besides the need for interactions on LFDs for various applications, LFDs have great potentials for their usage in cinematography due to their immersive 3D experience without the need for additional 3D viewing gears.[7] Cinematography – also known as film-making – is the field defining a set of techniques and rules for the effective communication of actions. It encompasses ideas, words, motions, tones and so many more, and communicates them visually.[8] Some of the recent LFDs have been designed to support cinematographic purposes, such as the HoloVizio C80 cinema system[*]. Among these set of cinematographic rules are those for camera placement and path planning.[9] While adhering to such rules, displaying contents on LFDs requires taking into account their technology-specific and device-specific challenges and limitations. Accordingly, the produced results and motions are evidently different from those displayed on conventional 2D screens. The plausibility of the different camera motions was initially addressed by another work of Guindy *et al.*,[10] in which thorough insight is applied to test the applicability of realistic physical camera simulations on LFDs. Although the physical camera animations and interaction models were implemented and tested on LFDs, the matter of movement and visual plausibility was left unchecked.

[*]https://holografika.com/c80-glasses-free-3d-cinema/

3. EXPERIMENTAL SETUP

3.1 Apparatus and test scenario

Subjective tests were conducted on the HoloVizio C80 HOP LFD, which has a horizontal viewing angle of 40°, a screen size of 3 m × 1.8 m and brightness calibrated to 1500 cd/m^2. The tests were carried out in a laboratory environment, isolated from audiovisual distractions. The lighting conditions were approximately 20 lux. The test participants could freely observe the visualized contents within a well-defined valid viewing area (VVA). The viewing distance (i.e., the screen-perpendicular dimension of the VVA) ranged from 4 m to 8 m. The selection of the minimum value originated from the constraint that test participants should not be located between the screen and the optical engine array of the LFD, as it may risk invalid LF through ray occlusion. The maximum value was based on the findings of Kara *et al.*,[11] which define the threshold at which the visual experience becomes closer to 2D due to the lack of disparity between the rays addressing the eyes of the viewer, with respect to a single point on the screen. As for the distance for sideways mobility (i.e., the screen-parallel dimension of the VVA), the width of the screen was used. The top-down view of the test scenario is shown on Figure 1.

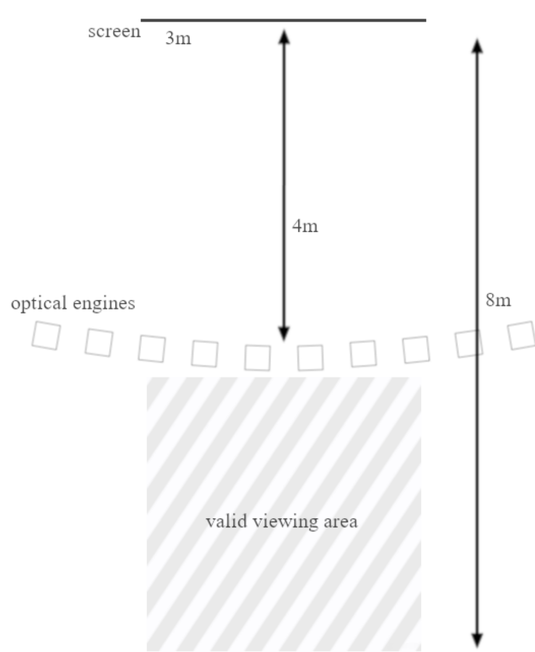


Figure 1: Top-down view of the test scenario.

3.2 Subjective tests on 3D interactions

The experimental setup included two sets of LF videos, and each set consisted of three scenarios. The first set was intended for the evaluation of the different 3D interaction techniques on LFDs. In this set, three theater models – along with their interactions – were shown on the LFD (see Figure 2). In all of these scenarios, a static camera was used as a means of navigation. However, a different selection / manipulation model was implemented for each scenario, including: (i) moving stage, (ii) single elliptical carousel and (iii) multiple carousels. Finally, for the application / system control, the two views – the theater model and the control room – were swapped.

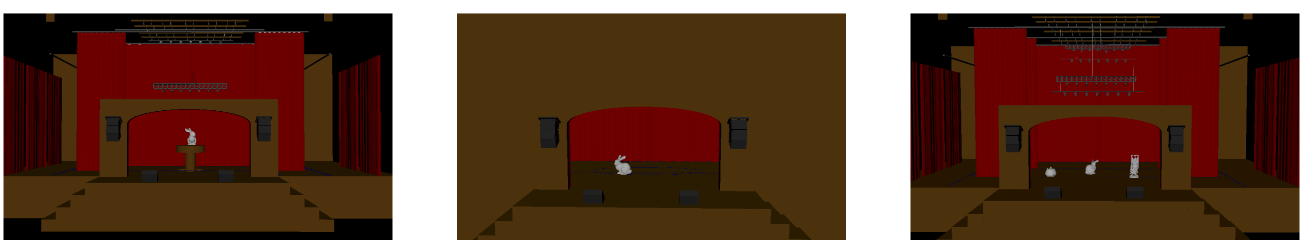
Test participants were asked to rate the test stimuli in terms of visual plausibility. Regarding the navigation task, users were asked whether they prefer the static camera or the moving camera on a 3-point scale ("static", "moving" or "unable to decide"). As for the second task, participants were asked to choose the preferred manipulation and selection model from their personal visual perspective ("moving stage", "single elliptical carousel" or "multiple carousels"). In addition to the selection of the model, users were asked to rate the motion of the curtains and the flying system on a 5-point Absolute Category Rating (ACR) scale ("bad", "poor", "fair", "good" and "excellent")[†]. Lastly, regarding the application / system control task, users were asked whether they prefer the swapping process between the two scenes over the GUI buttons displayed over the main scene ("yes", "no" or "unable to decide"). For every investigated scenario, participants were asked whether they prefer standing still or walking around ("moving", "standing" or "equal"), less or more interactions displayed ("less interactions", "more interactions" or "equal") and whether or not they personally deem it better to visualize the content on the LFD compared to conventional 2D displays ("better", "worse" or "same").

3.3 Subjective tests on realistic physical camera motions

In the second set of videos, we considered some of the physical camera movements typically used in cinematography (see Figure 3). These include: (i) collision camera, (ii) falling camera and (iii) suspension camera. Regarding the collision camera, a scene was created where a car moves towards a set of columns. Thereafter, the car accelerates and collides with one of the columns, on top of which the camera is placed. Hence, the name "collision camera" was given. For the second scene, the camera was designed to fall towards the ground which contains numerous objects, mainly boxes and cylinders. Finally, in the third scene, the camera is placed on a suspension object harmonically oscillating in a physical manner. Physical simulation tests were carried out in order to test the plausibility of these camera movements on the LFDs.

Test participants were asked to view these scenarios on the LFD and evaluate the visualized contents through a series of questions. However, before displaying the videos on the LFD, participants were asked about their preferred simulation camera type ("first person (FP) camera", "third person (TP) camera" or "equal"). The same question was then re-asked at the end of the experiment – after watching the videos – based on what was visualized on the LFD. Regarding each scenario – where a different camera type is shown on the LFD – participants were asked several questions to evaluate the visualized contents based on the camera type. The questions (i.e., assessment tasks) were the following:

- Rate the dizziness and the loss of focus resulting from the camera motion (5-point ACR scale).
- Choose your preference between standing still and moving around the LFD while observing the visualized content ("moving", "standing" or "equal").
- Rate occlusions with respect to the camera, the blurriness of objects and the camera collisions (5-point ACR scale).
- Indicate your personal preference whether or not it is better to visualize the contents on the LFD rather than on a 2D display ("better", "worse" or "same").



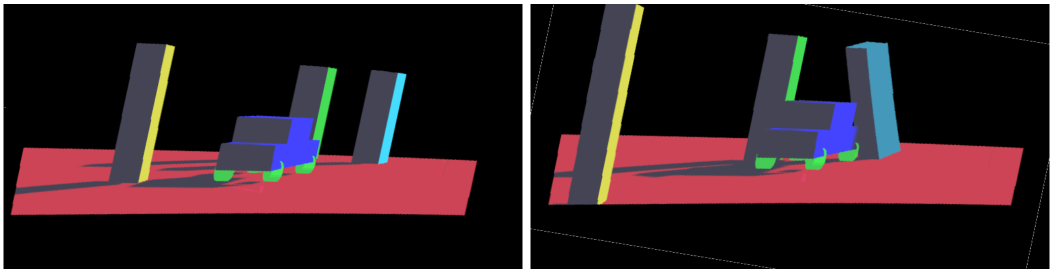
(a) Theater model 1 (b) Theater model 2 (c) Theater model 3

Figure 2: The different videos for the theater model.

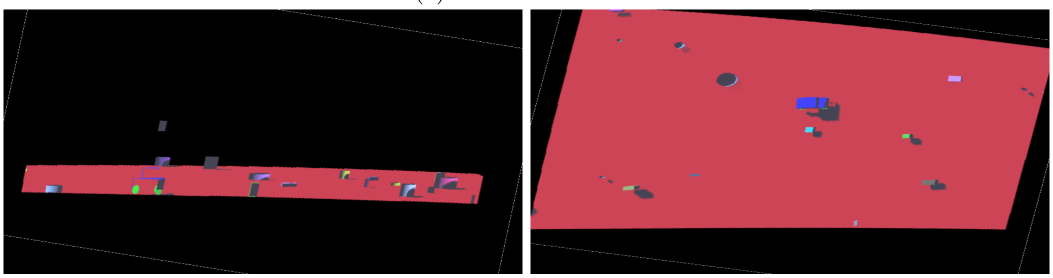
3.4 Test participants

The subjective tests were completed by 21 participants. 9 (42.85%) of the test participants were female and 12 (57.15%) were male. The test participants were pooled from an age range between 20 and 65. The average age was 29.

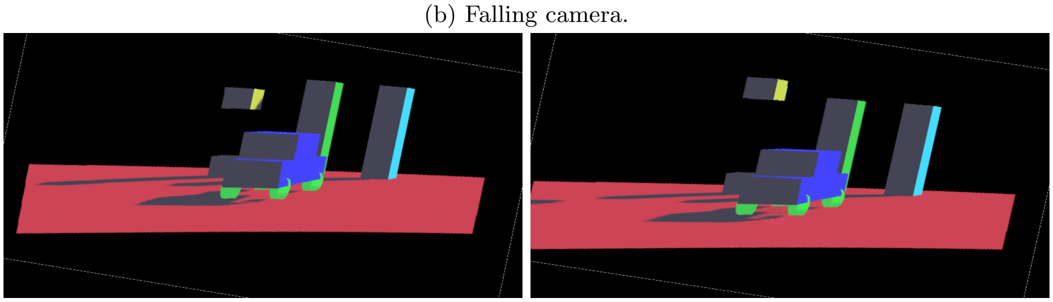
[†]ITU-T Rec. P.910 : Subjective video quality assessment methods for multimedia applications



(a) Collision camera.



(b) Falling camera.



(c) Suspension camera.

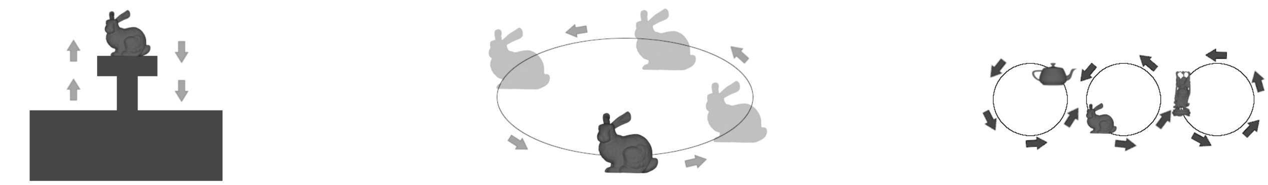
Figure 3: The different scenarios for the realistic camera physical simulation.

4. RESULTS AND DISCUSSION

As mentioned in Section 3, two sets of scenarios were visualized on the LFDs for assessing the plausibility of the viewed content. In the first set, the 3D interaction tasks (navigation, selection and manipulation, application / system control) for LFDs were primarily evaluated. In efforts to imitate the theater environment, a static camera was used for visualizing the theater stage viewed by various spectators. When asking participants about their camera preference, 61.9% agreed that the static camera was the best since LFDs already create a perceptual 3D effect. Hence, in case a boost is needed in the achieved 3D effect, walking around the screen will suffice. Accordingly, a moving camera may be visually disturbing due to the extra added motion, leading to dizziness and loss of focus.

A variety of selection and manipulation models were implemented on the LFD (see Figure 4). They include the upward / downward stage movement, the rotation of the main model on an elliptical carousel and the rotation of different models on separate carousels.

Participants were asked to choose their personally preferred selection / manipulation model. As shown on Figure 5, almost half of the users preferred the rotation of different models on separate carousels. The reason is that the rotation of models incorporates more interactions on the LFDs, hence, enhancing the perceived 3D effect. This is followed by the upward / downward stage movement selection model with more than a quarter of the votes. Finally, 19% of the test participants preferred the rotation of a single model on an elliptical carousel, as it includes the least amount of motion on the LFD. In addition to applying animation to the main element on the stage – as a means of selection and manipulation – animations were applied as well to the curtains and the theater's flying system. For curtains, side-to-side movements were carried out, in addition to the conventional



(a) Stage movement (upward / downward).

(b) Rotating the main character on an ellipse.

(c) Rotation of different models on separate carousels.

Figure 4: Selection and manipulation for the theater model.

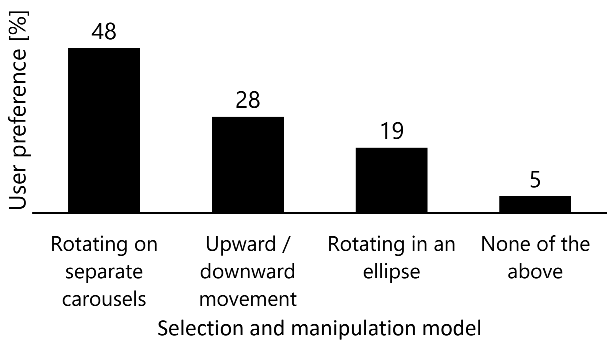


Figure 5: Users' preferences for the different selection and manipulation techniques used.

up / down curtain motion. However, for the flying system, conventional vertical motion was achieved while alternating between the different light bars. In addition to the aforementioned motion, rotations of the different traditional and ellipsoidal spotlights were performed. Test participants were asked to rate the movement of the curtains and the flying system on a scale from 1 to 5, where higher scores indicate higher user satisfaction. This resulted in scores of 2.76 and 3.81, respectively, further emphasizing the additional 3D effect achieved by increasing the visualized interactions. As shown on Figures 4a and 4c, the theater scene is viewed along with its backstage, unlike on Figure 4b. Participants were asked to indicate their visual preference between these two alternatives. Most users (76.2%) decided upon the theater scenes showing the backstage as they included more interactions, hence, increasing the perceived 3D effect. In addition to increasing the 3D effect by means of interactions, walking around the LFD adds to the achieved sense of immersion. Accordingly, 57.14% of the participants preferred walking around compared to remaining still. Regarding the last task of 3D interactions, application / system control illustrated some differences in the participants' visual preference. Users were asked to choose between two scenarios: (i) control buttons displayed on the main theater scene and (ii) switching between the theater scene and a monitor room with the control buttons while providing a visual feedback for the current theater state. A total of 57.14% preferred to have the control buttons on the same screen as the main scene, while 33.33% favored the second scenario. Although, at first sight it seems that incorporating the control buttons on the main scene is the best option, the problems of breaking the 3D immersion on the LFDs may arise. The reason is that it is quite obvious on the LFD that the GUI is a 2D overlay visualized on a 3D scene. It remains an ongoing research question how the feedback for 3D scenes should be conveyed to the users of LFDs. Overall, 76.19% preferred the interaction techniques on the LFDs compared to the conventional 2D displays.

Regarding the second set of experiments, participants were asked to choose the type of simulation camera they preferred – FP versus TP – with respect to the LFD. Most participants (76.2%) preferred the TP camera for visualizing the contents on LFDs, compared to conventional 2D displays. In general, for such conventional 2D displays, participants had a variety of choice between the FP and TP cameras, however, for LFDs, almost all of them considered using the TP camera as the best option. This is expected, considering that the FP camera has a closer look at the object of interest, leading to an increase in the blurriness of objects due to the optical limitations of the LFDs. Therefore, using TP cameras on LFDs or considering a "visual hack" to imitate the

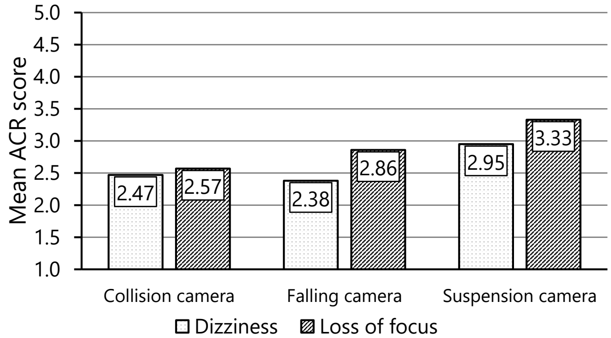


Figure 6: Diagram to illustrate the ratings for dizziness and loss of focus for the different camera types.

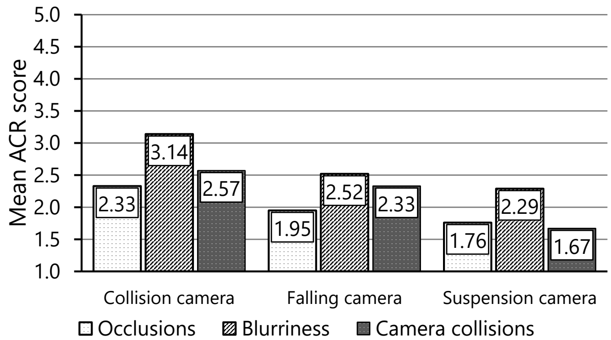


Figure 7: Diagram illustrating the ratings for the metrics proposed for the camera simulation types.

FP camera effect is preferable. As a means of doing so is to increase the scaling of the region of interest (ROI) with respect to the main element. The ROI is a box-shaped region within which all objects are visible. That is to say, for LFDs, the ROI is a virtual region within which all elements are visible on the LFD.[12]

Unlike the case of 3D interactions, for all physical scenarios, most of the test participants (66.6%) preferred standing in front of the screen instead of walking around. Particularly, the preference of the participants for moving or walking depended on the type of physical camera used. Most users preferred standing for the collision camera (76.2%), followed by the falling camera (71.4%), whereas the suspension camera (52.4%) had the least amount of votes. This is due to the fact that for the collision camera, extreme motions were visualized on the LFD because of the camera collision in the scene. From both sets of experiments, it could be inferred that the movement of participants and the movements / interactions on the LFD are indirectly proportional. Test participants were also asked to rate the dizziness and loss of focus (higher values indicate higher levels of dizziness and loss of focus). Participants rated the dizziness and loss of focus with an average of 2.60 and 2.92, respectively. Details for the mean scores for the different camera types are illustrated on Figure 6. Poor ratings for both the dizziness and loss of focus emphasize that irritating visuals may arise when using FP cameras on LFDs. Consequently, 71.4% of the participants preferred conventional 2D displays to LFDs when watching physical camera simulations. It is noted that among the different camera types used, suspension camera had the lowest percentage (38.1%) for promoting the usage of conventional 2D displays compared to LFDs. This is due to the fact that the suspension camera had the least amount of movement in the scene.

In the paper of Guindy *et al.*,[10] three metrics were suggested to evaluate the physical camera simulations. Although these metrics are objective – since they mostly count the number of elements in the scene that are occluded, collided or blurred – they could still be used in a way for subjective testing as well. In other words, spectators were asked to rate these metrics from their perspective instead of objectively applying these metrics. Figure 7 summarizes the results for this task. Starting with the occlusions with respect to the camera, test participants were asked to rate them from 1 to 5 (higher values indicate more perceived occlusions). Averages of 2.33, 1.95 and 1.76 were given to the collision, falling and suspension camera, respectively. For the blurriness of objects, the same rating scale was applied (higher values indicate a higher extent of perceived blur). For the

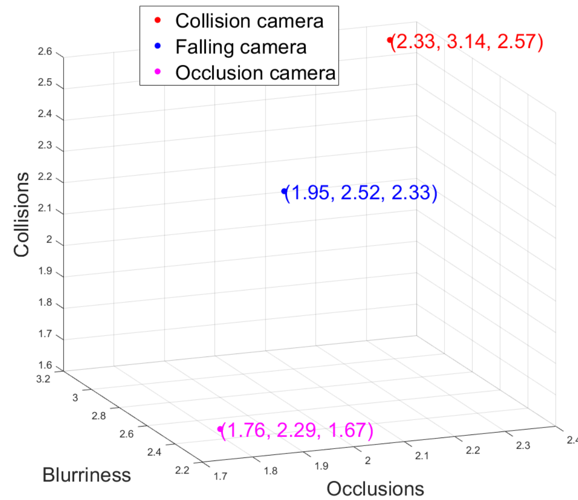


Figure 8: Graph showing the values of the metrics with respect to each camera type.

collision, falling and suspension cameras, average ratings of 3.14, 2.52 and 2.29 were given, respectively. Finally, the camera collisions were also rated from 1 to 5 (higher values indicate more collisions) with averages of 2.57, 2.33 and 1.67 for collision, falling and suspension camera, respectively.

It is quite evident how the intensity of camera movements affects the metric measures. Figure 8 depicts the relation between camera motion and the different metrics (occlusions, blurriness and collisions). From the figure, the directly proportional relation between the different metrics measures and the intensity of camera motion is deduced. In other words, as the camera motion increases – becomes more "vigorous" – occlusions, blurriness and collisions increase as well. This, however, leads to poor perceived visual quality, hence the initiative to consider slight camera motions for LFDs.

From all the scenarios for both experiments, the following key inferences are made:

- The preference of TP to FP cameras in almost all scenarios – due to the LFD optical limitations – lead to the blurriness of objects when using FP cameras. Hence, the simulation of the FP camera effect on LFDs remains open to further research.
- More interactions in scenes is preferred by most users as it adds to the perceived 3D sense created by LFDs.
- For LFDs, cameras with slight motions are preferred compared to those with vigorous motions. In other words, the less camera motion, the more plausible visualization for the contents on LFDs.
- The less interactions and movements on the LFD, the better for users to walk around, and vice versa.

5. CONCLUSION AND FUTURE WORK

Throughout the years, LFs have emerged as a means of describing the 3D world via light rays. LFDs were then invented as a means of visualizing the captured LFs. Since then, LFDs were used in many applications due to their ability to display the 3D world efficiently without the additional need of visual gears. In this paper, we presented empirical studies on the user preference regarding the different interaction techniques, as well as the numerous realistic physical camera motions. As a result of these studies, some notable inferences can be deduced. Generally, test participants agreed that LFDs provide a sense of 3D immersion, hence they are better than conventional 2D displays for visualizing specific contents. They were pleased with the 3D interactions and they even voted for more interactions on the LFDs. As for the physical simulations, controversial opinions were given as the camera motions involving lots of oscillations and collisions resulted in loss of focus on the LFD.

Addressing camera motions on LFDs remains an open research issue, including the choice of the best camera motions suitable for the LFDs and the means of implementing the other controversial movements resulting in less visual issues. Overall, more investigation is needed when it comes to the usage of FP camera on LFDs. A number of alternatives can be proposed to imitate the FP camera on LFDs without artefacts arising due to the limitations of the LFDs. As for interactions, different approaches for an application / system control feedback on LFDs can be still investigated in order to come up with an efficient solution to display the GUI while giving feedback to the current scene, without breaking the 3D immersion.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813170. Also received funding by 2018-2.1.3-EUREKA-2018-00007 and KFI 16-1-2017-0015, NRDI Fund, Hungary. The research reported in this paper is part of project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme.

REFERENCES

[1] Balram, N. and Tošić, I., "Light-field imaging and display systems," *Information Display* **32**(4), 6–13 (2016).

[2] Gershun, A., "The light field," *Journal of Mathematics and Physics* **18**(1-4), 51–151 (1939).

[3] Cserkaszky, A., Kara, P. A., Tamboli, R. R., Barsi, A., Martini, M. G., and Balogh, T., "Light-field capture and display systems: limitations, challenges, and potentials," in [*Novel Optical Systems Design and Optimization XXI*], International Society for Optics and Photonics (2018).

[4] Marton, F., Agus, M., Gobbetti, E., Pintore, G., and Rodriguez, M. B., "Natural exploration of 3D massive models on large-scale light field displays using the fox proximal navigation technique," *Computers & Graphics* **36**(8), 893–903 (2012).

[5] Guindy, M., Barsi, A., Kara, P. A., Balogh, T., and Simon, A., "Interaction methods for light field displays by means of a theater model environment," in [*Holography: Advances and Modern Trends VII*], International Society for Optics and Photonics (2021).

[6] Jankowski, J. and Hachet, M., "A survey of interaction techniques for interactive 3D environments," in [*Eurographics 2013-STAR*], (2013).

[7] Kara, P. A., Martini, M. G., Nagy, Z., and Barsi, A., "Cinema as large as life: Large-scale light field cinema system," in [*International Conference on 3D Immersion (IC3D)*], IEEE (2017).

[8] Brown, B., [*Cinematography: theory and practice: image making for cinematographers and directors*], Taylor & Francis (2016).

[9] He, L.-w., Cohen, M. F., and Salesin, D. H., "The virtual cinematographer: A paradigm for automatic real-time camera control and directing," in [*Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*], *SIGGRAPH '96*, 217–224, Association for Computing Machinery, New York, NY, USA (1996).

[10] Guindy, M., Barsi, A., Kara, P. A., Balogh, T., and Simon, A., "Realistic physical camera motion for light field visualization," in [*Holography: Advances and Modern Trends VII*], International Society for Optics and Photonics (2021).

[11] Kara, P. A., Tamboli, R. R., Cserkaszky, A., Barsi, A., Simon, A., Kusz, A., Bokor, L., and Martini, M. G., "Objective and subjective assessment of binocular disparity for projection-based light field displays," in [*International Conference on 3D Immersion (IC3D)*], IEEE (2019).

[12] Darukumalli, S., Kara, P. A., Barsi, A., Martini, M. G., and Balogh, T., "Subjective quality assessment of zooming levels and image reconstructions based on region of interest for light field displays," in [*International Conference on 3D Imaging (IC3D)*], IEEE (2016).