Cserkaszky, Aron, Kara, Peter A., Barsi, Attila and Martini, Maria G. (2018) Expert evaluation of a novel light-field visualization format. In 2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON). Piscataway, U.S. : Institute of Electrical and Electronics Engineers, Inc. ISSN (online) 2161-203X ISBN 9781538661253

https://doi.org/10.1109/3DTV.2018.8478436

EXPERT EVALUATION OF A NOVEL LIGHT-FIELD VISUALIZATION FORMAT

*Aron Cserkaszky**, *Peter A. Kara*†, *Attila Barsi**, *Maria G. Martini*†

*Holografika
{a.cserkaszky, a.barsi}@holografika.com

†WMN Research Group, Kingston University
{p.kara, m.martini}@kingston.ac.uk

ABSTRACT

Light-field visualization is continuously emerging in industrial sectors, and the appearance on the consumer market is approaching. Yet this process is halted, or at least slowed down, by the lack of proper display-independent light-field formats. Such formats are necessary to enable the efficient interchange between light-field content creation and visualization, and thus support potential future use case scenarios of this technology. In this paper, we introduce the results of a perceived quality assessment research, performed on our own novel light-field visualization format. The subjective tests, which compared conventional linear camera array visualization to our format, were completed by experts only, thus quality assessment was an expert evaluation. We aim to use the findings gathered in this research to carry out a large-scale subjective test series in the future, with non-expert observers.

***Index Terms* —** Light-field, visualization format, perceived quality assessment

1. INTRODUCTION

Light-field is essentially a set of light rays that represents a scene in a "useful form". Useful in this context means that the data can be utilized to recreate the scene in a visualization use case scenario that is meaningful to the human observer. Generally, the goal of light-field representation is often to contain information that enables glasses-free, true 3D visual experience.

Technically speaking, when considering light-field visualization, we can state that each and every display in the world is a light-field display (and of course every camera is a light-field camera). However, in today's terminology, light-field displays visualize directional views (similarly to multi-view displays), and also provide continuous motion parallax, at least along the horizontal axis. These displays are referred to as horizontal-parallax-only (HPO) displays, and also can be found in the literature as super multi-view displays. In the future, the scientific community is aiming to develop full-parallax (FP) displays, which simultaneously support the horizontal and vertical parallax effect, similarly to what can be perceived in the real world.

A visualization format can be approached from various angles, but possibly the two most important characteristics are visual quality and data volume (or storage/transmission requirement). Visual quality is either measured by an objective metric or actual perceived quality through subjective tests, and data volume is quite straightforward; the less data, the better. The trade-off between characteristics is rather evident, and the goal in research and development is to find solutions that satisfy both criteria at the same time: to enable excellent visualization quality at a manageable data volume.

In case of light-field visualization, a third angle would be display independence. Current formats do not tackle this issue, yet its importance is growing parallel with the emergence of capture and display technologies. One core issue originates from the fact that such formats are initially created for densely and uniformly sampled content. However, insufficient input density on its own may result in severe degradations of visual quality, and the storage format's assumptions regarding the content can make it even worse. Also, while display-focused formats may be efficient for given displays, they suffer multiple penalties during the conversion on a different light-field display. First of all, such conversion procedures may have high computational demands, hindering real-time use cases. Furthermore, as conversion calculates the color of each and every light ray within the light-field, if this is combined with insufficient density, the resulting displayed light-field is prone to have inaccuracies such as color mismatches.

One of the first light-field format dates back to 1996, published by Levoy and Hanrahan [1]. The primary concept proposed in the work was to use slabs of bi-planar coordinate-based ray parametrization. The light-field format[1] defines rays by four coordinates: the point of intersection on one 2D plane (s and t) and on another one as well (u and v), thus four coordinates uniquely define a light ray. The main issue with this generic format is that it assumes uniform sampling, which is currently not supported by the existing light-field displays.

The *.lfp* format was specially designed by Lytro for their light-field cameras with microlense-based optics [2]. The primary consideration here is that the format was made to contain narrow-baseline light-field information, and thus such data is not compatible with wide-baseline light-field visualization technologies.

In practice, light-field content is frequently generated by normal pinhole cameras, also known as perspective cameras. Using a single camera to capture the entire light-field is possible, but it restricts the content to static scenes. In case of a camera array, no such restriction applies. Formats for the utilization of perspective

The work in this paper was funded from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreements No 676401, European Training Network on Full Parallax Imaging and No 643072, Network QoE-Net.

[1] Stanford light-field file format
https://graphics.stanford.edu/software/lightpack/doc/file_format.html

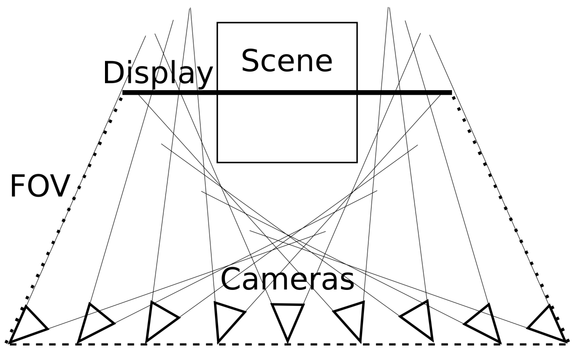


Figure 1: Arrangement of the linear camera array.

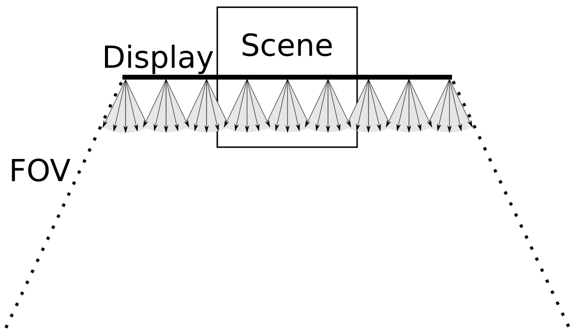


Figure 2: Arrangement of visual information in our new format.

cameras are already in use[2 3 4], and they have a good trade-off between visual quality and data volume. The perspective cameras are typically positioned along a straight line (linear camera array), but other solutions are possible as well, such as distribution along an arc (arc camera array). Although for capture and storage linear arrays of perspective cameras can be very cost-efficient, the main issue is the major difference between the capture and the display side. Due to the differences in sampling distribution, a highly resource-consuming conversion at the display side (involving light-field interpolation) is a must.

On the level of format standardization, at the time of this paper, the JPEG Pleno collaboration [3] is addressing light-field captured by camera systems using microlense-based optics and camera arrays. These systems inherently pose challenges, like the issue of narrow baseline in case of lenslet-based camera systems.

Summa summarum, there is a need for a display-independent light-field format, that can overcome the issues mentioned above. Recently, we introduced our own format for light-field visualization on a theoretical level [4], which we have successfully implemented for practical use. In this paper, we present the results of an expert evaluation, that subjectively compared the visual quality of the linear camera array visualization to our own novel format. We rendered different scenes (still objects) with various input parameters to test the performance in a paired comparison.

The reminder of this paper is structured as follows: Section 2 provides a brief introduction of our novel light-field format. Section 3 discusses the configuration and the results of the expert evaluation. The paper is concluded in Section 4.

[2]Disney Research light-field data-set https://www.disneyresearch.com/project/lightfields/

[3]Nagoya University sequences http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/

[4]Heidelberg Benchmark 4D light-field data-set http://hci-lightfield.iwr.uni-heidelberg.de/

Table 1: Investigated Input Types.

Input Type	**Views per degree**	**Number of views**
A	2	101
B	1	51
C	0.667	33
D	0.5	25

2. THE NOVEL LIGHT-FIELD FORMAT

Our novel light-format was designed to be an efficient mezzanine format between capture and display light-fields, while not using special properties of any particular light-field display. Capture light-fields are the rays that a given camera system records from a scene, usually with pinhole cameras, resulting in a light-field that is not efficient to display. Display light-fields are the rays that a display system can emit to the viewers and the structure of the light-field is normally unique to the type of the device.

The format only assumes that the screen of the display is approximately flat, and the rays it can emit have a symmetry in the angular domain. These are assumptions that all light-field displays currently on the market adhere to, according to the best knowledge of the authors. The format describes the 4D light-field with two spatial coordinates, that indicate the start positions of the rays, and two angular coordinates, that give the directions of the rays. The header of the format contains the properties of the light-field: the number of pixels and their physical size in each spatial dimension, the field of view and the number of angular views in each angular dimension. In our terminology, the dimensions are denoted with s, t, ϕ and θ. The coding of the format is being improved at the time of this paper, and we expect to publish it along with open-access data-sets at a later date. The practical difference between linear perspective camera array systems and our format is emphasized in Figure 1 and 2.

The advantages of the format include backwards compatibility with plain 2D images by setting both angular dimensions to 1. FP displays likewise can easily show HPO light-fields in this novel format. Furthermore, by having a defined yet flexible structure for the rays, both the capture and the display side can use precomputed look-up-tables for fast conversion of the light-field, instead of the computationally expensive direct conversion between arbitrary camera images and specific display light-fields. This makes the format especially attractive for real-time light-field use cases.

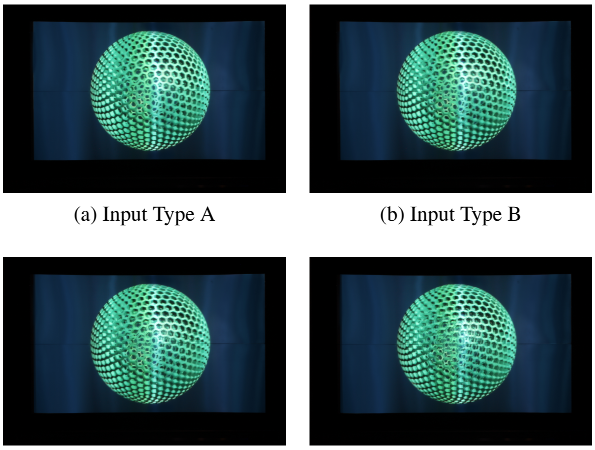
Due to the HPO nature of currently existing, commercially available light-field displays, we refer our novel format as the "s-t-phi format" or as the "angularly continuous light-field format".

3. EXPERIMENTAL SETUP AND OBTAINED RESULTS

As stated earlier, the purpose of the research was an initial assessment of visual quality via an expert evaluation. We compared the outputs of a conventional linear camera array technique with our implementation of the novel s-t-phi.

For visual stimuli, we used the 972-faced polyhedron and the structure of 120 regular dodecahedra[5]. In this paper, we refer to them as stimulus A and B, respectively. In numerous past researches [5] [6] [7], we have already involved these complex mathematical bodies in our tests, due to their sensitivity towards degradations.

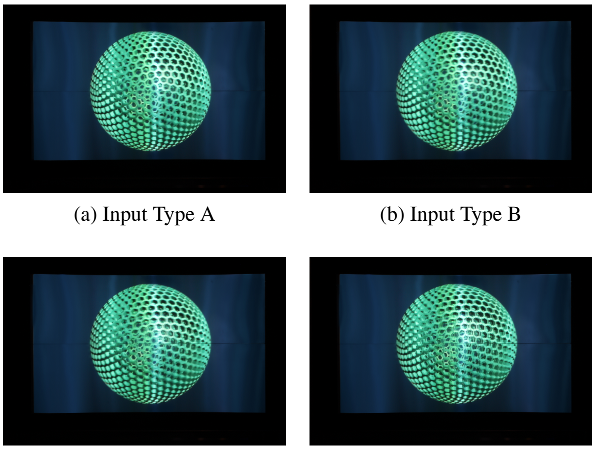
[5]George W. Hart's Rapid Prototyping Web Page www.georgehart.com/rp/rp.html



(a) Input Type A (b) Input Type B

(c) Input Type C (d) Input Type D

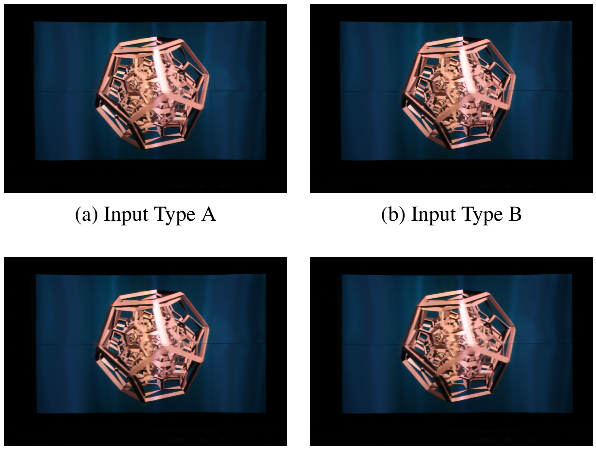
Figure 3: DSLR images of stimulus A for perspective cameras.



(a) Input Type A (b) Input Type B

(c) Input Type C (d) Input Type D

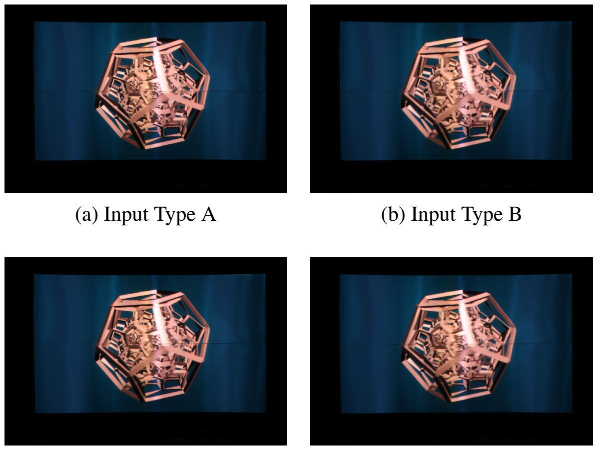
Figure 4: DSLR images of stimulus A for s-t-phi.



(a) Input Type A (b) Input Type B

(c) Input Type C (d) Input Type D

Figure 5: DSLR images of stimulus B for perspective cameras.



(a) Input Type A (b) Input Type B

(c) Input Type C (d) Input Type D

Figure 6: DSLR images of stimulus B for s-t-phi.

While the spatial resolution was constant (1440×1080), we created the stimuli with different extents of angular resolution. The four input types are reported in Table 1. For s-t-phi, the input is interpreted in views per degree, and for the perspective camera format, in the corresponding number of views.

From these values, it is evident that the field of view of the datasets were 50 degrees. Figure 1 shows the arrangement of the virtual linear camera array, that was used to render images of the scene. The lens properties of the virtual cameras were set to match the rendered area of the s-t-phi case. These images were converted directly to the specific light-field of the display system, creating the best image quality achievable with perspective cameras. The s-t-phi arrangement is shown on Figure 2, indicating that only the directions within the field of view were stored in it, making it fully comparable to the perspective case.

The display we showed the visual stimuli on was the HoloVizio C80 light-field cinema system[6], calibrated for 50 degrees. The tests were carried out in an isolated laboratory without any audio-visual distractions, and the lighting condition of the environment was approximately 20 lx. The initial viewing distance was 2.5H, which corresponded to 4.6 meters. However, in this research, the viewing distance and the viewing angle was arbitrarily changeable by the test participants, in order to precisely examine the entire static object. The major restrictions in these aspects were defined by the valid field of view, and also by the display type: as the C80 is a front-projection light-field display, observing the screen from an insufficiently small distance can result in the occlusion of a portion of the projected light rays. The stimuli visualized on the C80 using the two formats and the four inputs are shown in Figure 3, 4, 5 and 6.

The subjective quality assessment task of the experts was to determine whether the stimuli in the pairs were distinguishable, and if they were, which one was better. Originally, we considered to use the ITU-R Rec. BT-500.13 seven-point (*"Much Worse"*, *"Worse"*, *"Slightly worse"*, *"Same"*, *"Slightly better"*, *"Better"*, *"Much better"*) scale. However, instead of scoring on a fine-grained scale, we used a three-point scale and added an option for detailed written feedback regarding the differences, as this option provided us more valuable data from the experts.

The stimuli were shown in pairs, and the subjective rating was directly after each pair. There were no separation screens within the pairs, only between the pairs (5-second blank screen). Also, the test participant could request multiple switchings between the stimuli in a pair, in order to enhance visual accuracy of the experts (many requested the switches from different angles and distances).

[6] HoloVizio C80 light-field cinema system
http://holografika.com/c80-glasses-free-3d-cinema/

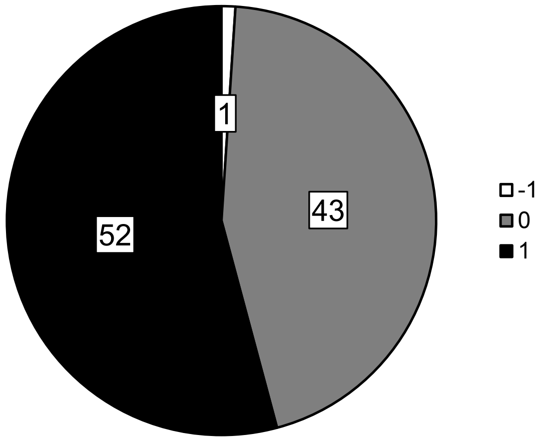


Figure 7: Overall scoring distribution of the tests.

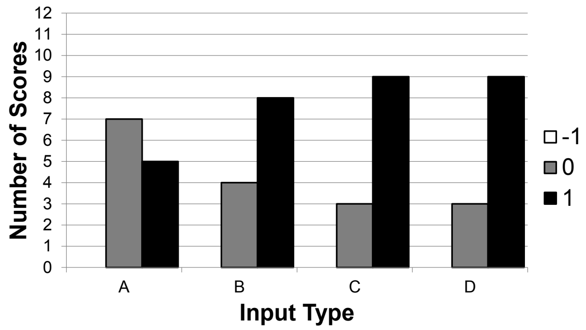


Figure 8: Scoring distribution for stimulus A.

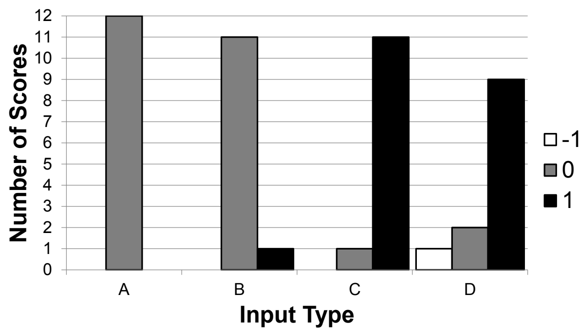


Figure 9: Scoring distribution for stimulus B.

A total of 12 experts completed the test, 11 males and 1 female, with an average age of 38, within the age interval from 27 to 58. Before the experiment, the test participants were subject to screenings based on the Snellen charts and Ishihara plates.

The obtained numerical results were either -1, 0 or 1. -1 was given as a score if the test participant deemed the perspective camera representation to be better in general, 0 if the two stimuli could not be distinguished and 1 if s-t-phi provided the better visualization quality. As there were 12 test participants, 2 source stimuli and 4 input types, a total of 96 scores were collected.

The overall scoring distribution, reported in Figure 7, show that more than half of the scores favored the s-t-phi format. In a similarly large number, the two stimuli were indistinguishable. There was only one single score favoring the perspective camera format.

Figure 8 and 9 depict the scoring distribution for the two stimuli separately. We can see that in case of both stimuli, when insufficient inputs (C and D) were used, visualization with the s-t-phi format was clearly preferred. For dense inputs (A and B), stimulus B was indistinguishable, while the feedback for the stimulus A was mixed 0 and 1.

The comments provided by the experts revealed that the major issues at low-density inputs for the perspective camera format were the crosstalk effect and the occasional blur, both mostly at the front of the objects, which utilized the most of the depth budget. Although the s-t-phi format suffered degradations as well, but endured more due to its angularly continuous nature.

4. CONCLUSION

In this paper we presented an expert evaluation of our previously proposed s-t-phi light-field format. Our results shows that with the same amount information stored, experts rated the resulting image quality of the s-t-phi to be better than the currently commonly used camera-image-based light-field formats. We will continue this research with finding efficient coding schemes for the s-t-phi format and we will conduct subjective evaluation studies to further validate its advantages. Such subjective experiments on perceived quality will mainly target test participants without prior experience with light-field visualization, as even though expert evaluations provide many valuable insights, it is indeed the average user that needs to be visually satisfied at the end of the day.

5. REFERENCES

[1] Marc Levoy and Pat Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[2] Todor Georgiev, Zhan Yu, Andrew Lumsdaine, and Sergio Goma, "Lytro camera technology: theory, algorithms, performance analysis," in *Proc. SPIE*, 2013, vol. 8667, p. 86671J.

[3] Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, and Peter Schelkens, "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.

[4] Aron Cserkaszky, Attila Barsi, Peter A Kara, and Maria G Martini, "Towards display-independent light-field formats," in *3D Immersion (IC3D), 2017 International Conference on*. IEEE, 2017, pp. 1–7.

[5] Aron Cserkaszky, Attila Barsi, Peter A Kara, and Maria G Martini, "To interpolate or not to interpolate: Subjective assessment of interpolation performance on a light field display," in *Multimedia & Expo Workshops (ICMEW), 2017 IEEE*. IEEE, 2017, pp. 55–60.

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

[7] Peter A. Kara, Aron Cserkaszky, Subbareddy Darukumalli, Attila Barsi, and Maria G. Martini, "On the Edge of the Seat: Reduced Angular Resolution of a Light Field Cinema with Fixed Observer Positions," in *9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, 2017.