

A Rendering Solution to Display Light Field in Virtual Reality

Evgeniy Upenik, Irene Viola and Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Email: firstname.lastname@epfl.ch

Abstract—There is a need for affordable and easily deployable rendering and display solutions to take full advantage of light field imaging. In particular, the research community needs ways to better assess the impact of various light field imaging technologies based on different criteria. In this paper, we propose a solution to render light field images on head mounted virtual reality displays using off-the-shelf components. The proposed framework is based on WebGL and supports rendering of narrow baseline light field images with interactions that result in changes of perspective. The system can be used in subjective quality assessments of light field via crowd-sourcing. It also allows for tracking and recording of viewing interactions.

I. INTRODUCTION

Light field imaging is associated with augmented reality (AR) more often than it is with virtual reality (VR). Nonetheless, it is becoming a very desirable type of content for VR, along with omnidirectional imaging and point clouds which are naturally designed for this type of media.

When compared to traditional omnidirectional content, light field rendering allows for more realistic visualization of 3D spaces in a virtual or augmented reality scenario, thanks to the full parallax environment that can provide depth and focus cues. In recent years, several wearable light field display prototypes have been designed and proposed by both academics [1], [2], [3] and industry [4], [5]. The development of commercial devices such as the Avegant light field display headset¹ or Magic Leap Digital Lightfield² promises near-eye light field head-mounted displays (HMD) to be available to consumers in the near future. However, as those devices are currently either in prototype state or too expensive to be widely accessible to the public, off-the-shelf solutions are needed to perform quality assessment of light field content in a virtual reality scenario. In particular, if already available devices are used to perform quality assessment, crowd-sourcing can be employed to collect a large number of scores with reduced costs in terms of time and expenses [6].

Several studies of quality assessment for light field images can be found in the literature. Paudyal et al. investigate the impact of watermarking on visual quality of light fields using Absolute Category Rating (ACR), and in particular on the relationship between watermark strength and visual quality [7].

Darukumalli et al. and Kara et al. examine the quality of experience using light field displays, and their relationship with angular resolution and zooming levels [8], [9]. Viola et al. evaluate compression solution through subjective quality assessment using passive and interactive methodologies on conventional 2D displays [10]. They also perform a statistical comparison between the two methodologies to determine the impact of interaction on the results [11]. Konrad et al. evaluate the quality of experience related to several focus-tunable near-eye display modes, as well as the effect of the display mode on the user performance [12]. However, to the best of the authors' knowledge, no quality assessment of compression artifacts for light field images has been performed on HMD.

The remainder of this paper is organized as follows. Section II describes the software solution to render narrow baseline light field images in VR environment, its features and architecture. Section III describes a pilot subjective quality evaluation experiment conducted using the proposed rendering solution. Section IV presents the results of the pilot experiment and their analysis followed by a discussion. Finally, Section V concludes the paper.

II. RENDERING LIGHT FIELD IN VR

In this section we propose a software solution to render narrow baseline light field images in virtual reality implemented using WebGL. We describe its architecture, main components and features.

The proposed VR rendering solution allows visualizing light field images on mobile HMD platforms, such as Google Cardboard or Samsung GearVR, desktop computers, and head-mounted displays, such as HTC Vive and Oculus Rift. The portability is achieved by using a web-based platform. The types of implemented interactions include horizontal and vertical narrow baseline perspective changes. In the VR environment viewers interact with movements of a head, whilst on personal computer a mouse or a trackpad can be used.

The rendering can be performed in any web-browser which supports WebGL standard in its version 1.0, including all mobile devices supporting OpenGL ES 2.0. Source code is written in JavaScript language and requires Three.js 3D graphics library. Light field images are rendered as separate perspective views which are changed in real time according to the data from motion sensors of a device in the mobile and

¹<https://www.avegant.com/>

²<https://www.magicleap.com/>

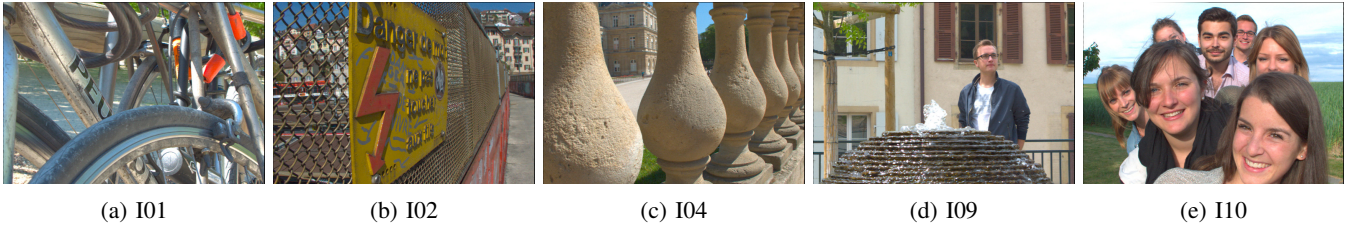


Fig. 1: Stimuli used in the pilot experiment

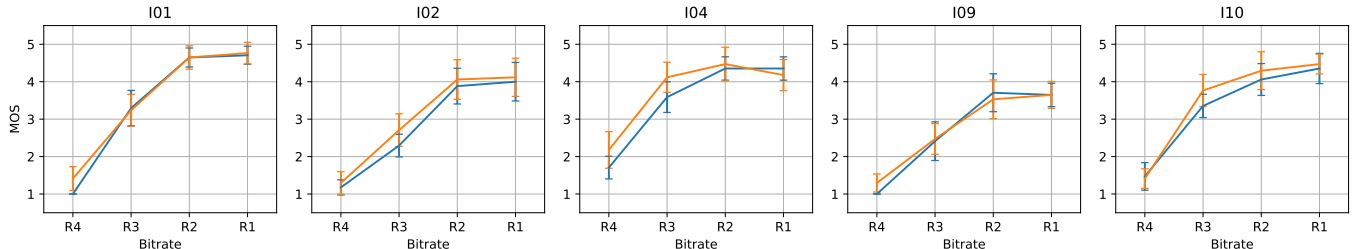


Fig. 2: Mean opinion scores (MOS) plotted against bit-rates and their 95% confidence intervals (CI). A blue solid line depicts HEVC and an orange solid line depicts VP9 encoded stimuli.

HMD case, and according to mouse or trackpad movements in the desktop computer case. In order to display a light field image, all its perspective views are downloaded from the server as texture files. Then those textures are loaded into GPU memory. The application tracks user movements and renders the texture corresponding to the current perspective view image.

Besides the rendering *per se*, the application provides additional features allowing one to use the software to conduct subjective quality evaluation experiments for narrow baseline light field assessment. This includes a storyboard implementing training and evaluation scenarios, an ability to assign a score to a light field image in an immersive way within VR, and store resulting assessment data on a server.

The storyboard currently includes a training session followed by an evaluation session. An absolute category rating (ACR) methodology [13] is implemented to collect subjects' votes and send them to the server after evaluation is completed. During the evaluation process subjects use a 3D menu for voting without leaving the immersive VR environment.

An important direction of research in immersive imaging and in particular in light field is the analysis of user interactions. The proposed software tracks how users interact with narrow baseline light field content. Every time a subject moves from one perspective view to another, it is recorded and sent to the server.

In order to deploy the developed software, one needs an HTTP-server supporting PHP server-side scripting. The latter is required for storing results on the server. Once the software is on-line it can be accessed by multiple users simultaneously over the Internet. Assuming the high availability of affordable consumer HMD, this can allow for large scale crowd-sourcing subjective quality evaluations with interaction analysis.

The source code of the developed software for rendering light field in VR and a demo are publicly available on-line at <https://mmspg.github.io/lightfieldvrtb/>.

III. PILOT EXPERIMENT

This section describes a pilot experiment on subjective quality evaluation in VR environment conducted in Multimedia Signal Processing Group laboratory at EPFL with the purpose to validate the solution proposed in Section II.

A. Population and environment

The experiment was performed with 17 subjects, of which 9 were males and 8 were females. The age of the subjects ranged from 18 to 37 years old, with the average equal 25.38 and the median equal 26.73. Prior to the experiment all the subjects were tested for their visual acuity and color vision.

B. Equipment

To render narrow baseline light field images in VR, experimenters used the software solution described in Section II. Subjects were wearing a Google Cardboard compatible HMD-mount for mobile devices (MergeVR³) with a Samsung Galaxy S7 Edge smartphone installed inside. The resolution of the device is 2560×1440 pixels or 1280×1440 pixels per eye. The pixel density is 534 pixels per inch. The field of view of this HMD-mount is 96 degrees. It has 42 mm lenses and allows for adjustment of interpupillary distance.

³<https://mergevr.com/goggles>

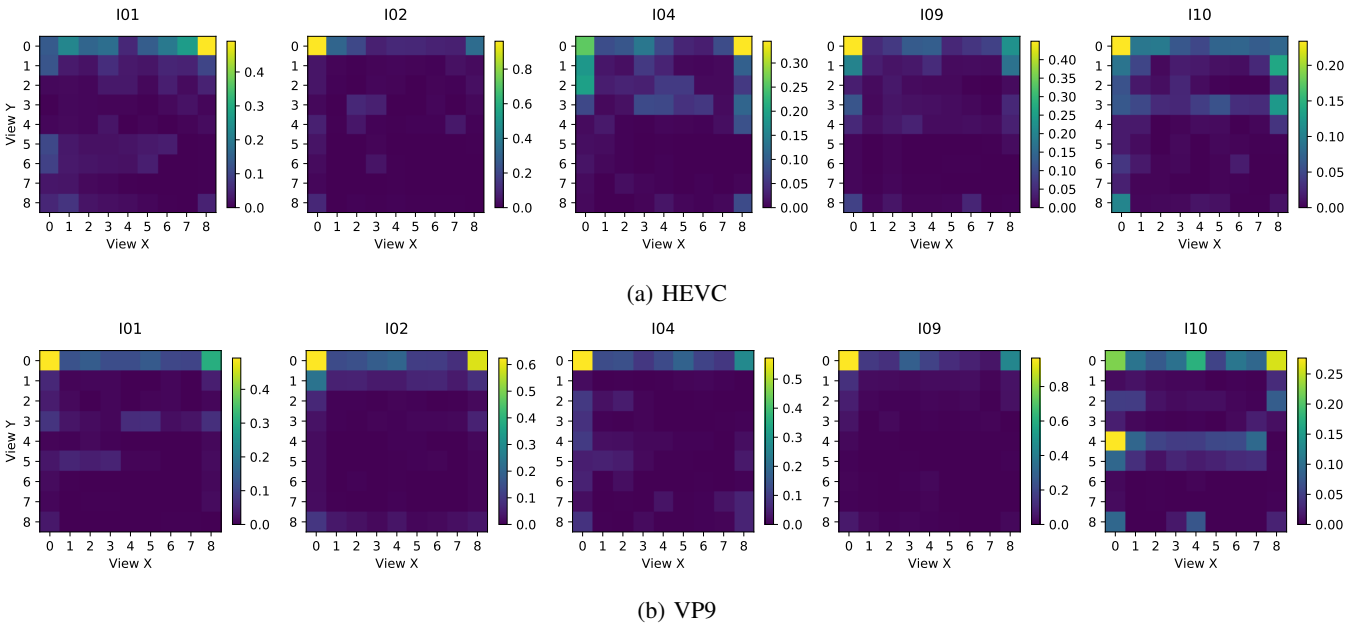


Fig. 3: Charts of time spent on a view in seconds

TABLE I: QP values selected to encode contents with HEVC.

Content	R1	R2	R3	R4
I01 (Bikes)	13	24	33	44
I02 (Danger de mort)	15	26	35	43
I04 (Stone pillars)	14	23	30	40
I09 (Fountain)	14	24	32	43
I10 (Friends)	12	21	29	40

TABLE II: Settings for VP9 coder.

```
--i422 --input-bit-depth=10 --profile=3 -w < Width > -h < Height >
--target-bitrate=< bitrate > --cq-level=0 --bit-depth=10 --codec=vp9
--fps=30000/1000 --best -o < Output > < Input >
```

C. Stimulus set

The set of stimuli used in the pilot subjective quality evaluation experiment is based on the light field image data set created by Rerabek et al. [14]. Five light field images have been selected to represent different categories. Figure 1 shows central-view all-in-focus thumbnails of the unimpaired stimuli in 2D representations.

Light field images have been compressed using two different codecs adapted in such a way that they process the perspective views as a pseudo-temporal sequence in a serpentine order. Before being fed to encoders, all the perspective images were padded with black pixels, color-space was converted to YUV and re-sampled to 422 with 10-bit depth.

The codec number one (P1) is HEVC Main10 profile. x265⁴ library was used to perform the compression. Quantization Parameters (QP) were set to match the preselected compression ratios. In the Table I one can find the exact values of different QP used in the test.

The codec number two (P2) is VP9⁵. The full command line used to produce compressed stimuli can be found in Table II.

The target bitrate was chosen to match the corresponding compression ratios as defined below.

The codecs were evaluated on four bitrates, namely $R1 = 0.75$ bpp, $R2 = 0.1$ bpp, $R3 = 0.02$ bpp, $R4 = 0.005$ bpp. The compression ratios are computed as ratios between the size of the uncompressed raw images in 10-bit precision ($5368 \times 7728 \times 10$ bits = 414839040 bits = 10 bpp) and the size of the compressed bitstream.

D. Methodology

The subjective quality evaluation experiment was designed to follow a single stimuli Absolute Category Rating (ACR) method [13]. ACR is a single stimulus evaluation where stimuli are presented subsequently to subjects, and voting is performed after each viewing. Images are assessed using five-grade quality scale with the following levels: 5 - Excellent, 4 - Good, 3 - Fair, 2 - Poor, and 1 - Bad.

Due to distortions naturally occurring in lenslet-based light field content, the border perspective views were deemed not suitable for visualization, since they would negatively bias subjects. Hence, only the central 9×9 perspective views out of the 15×15 views were selected for the test. The contents were converted from *PPM* file format in 10 bits to *PNG* file format in 8 bits, due to limitations of the display.

⁴<https://www.videolan.org/developers/x265.html>

⁵<https://www.webmproject.org/vp9/>

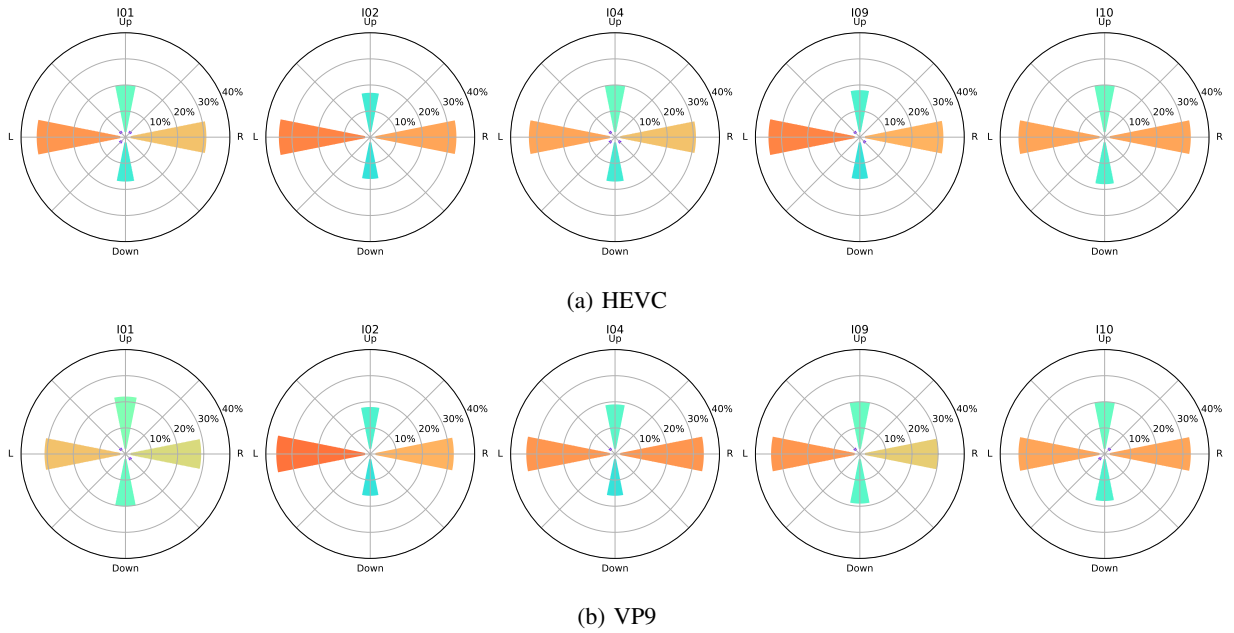


Fig. 4: Circular histograms of the directions of user interactions

IV. RESULTS AND DISCUSSION

This section describes the analysis of experimental data and presents its results including mean opinion scores and interaction vectors. A discussion follows the presentation of the results.

Raw experimental data obtained from the evaluations consists of ACR scores given by each subject for each stimulus and iteration records for each subject grouped by stimulus. The latter contains time stamps for every change of a perspective view initiated by a subject.

Before proceeding to further analysis the data was screened for outlier subjects using the method described in [15]. Zero outliers were detected among the subjects.

A. Mean opinion scores

Figure 2 presents subjective mean opinion scores (MOS) for five contents compressed with two different codecs at four different bitrates selected as described in Section III. 95% confidence intervals were computed for each MOS assuming T-Student's distribution of subjective scores for a stimulus. The MOS values are plotted against the bitrates R4-R1 where R4 is the lowest bitrate and R1 is the highest bitrate. The exact values can be found in Subsection III-C.

B. Interaction analysis

In order to compute time spent on each view, the difference between the time stamps of two subsequently shown perspective views was taken. All the time stamps were grouped by stimulus. The first and the last time stamps in each group were dropped. Then the average time spent on a view was computed. Stimuli compressed with all the bitrates were counted for average. Figure 3 depicts maps of average time spent on a view for each content for two codecs.

Interaction vectors were computed in the following way. The time stamps for every change of a perspective view initiated by a user were grouped by stimulus and subject. The first and the last time stamps in a group were dropped. An interaction vector was computed per each time stamp record as a difference between corresponding x and y coordinate pairs of the subsequent view and the current view compensating the reverse Y-axis direction (i.e. from top to down) in perspective view coordinate system. Figure 4 shows circular histograms of interaction vectors per content per codec.

In order to analyze a typical path from each perspective view to a subsequent one, we have computed an average interaction vector for each view. This average was taken among all subjects and all bitrates grouped by content and by codec. One can find the vector field plots depicting average interaction vectors for each view in Figure 5.

C. Discussion

Mean opinion scores in Figure 2 show no statistically significant difference in visual quality for the two codecs except in bitrate R3 for content I04 where VP9 outperforms HEVC.

From average time spent on a view in Figure 3 one can see that subjects were systematically biased towards spending more time on the top row. This can be possibly explained by a wrong vertical position of light field images in the VR space, which was exactly in front of the camera. In order to compensate this bias in future experiments one should consider placing images higher than the camera. Furthermore, the average time spent on a view can be used in the future to compute weighted subjective scores for each view.

Circular histograms of the interaction vectors in Figure 4 show clearly that subjects tend to interact more horizontally

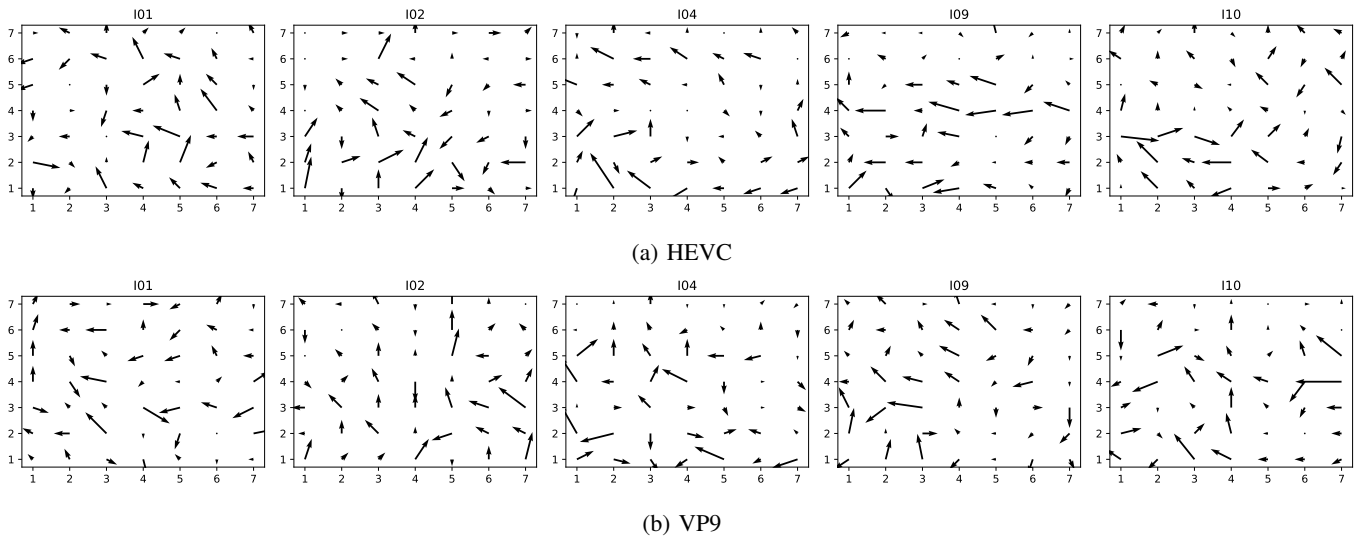


Fig. 5: Average interaction vectors for each perspective view per content per codec

than vertically. One can also notice correlation between the histograms for the same content with respect to different codecs.

Average interaction vectors for each perspective view presented in Figure 5 can be used as ground truth data for estimation of the most probable subsequent view. This is required to develop efficient compression algorithms providing fast random access to perspective views of a light field image.

V. CONCLUSION

In this paper we propose a solution for rendering narrow baseline light field images in a VR environment which allows interactions with their perspectives. The developed software includes features to perform subjective evaluations and tracks users interactions.

A pilot subjective quality evaluation experiment for light field in VR was conducted with 17 subjects participating in the assessments.

The results of the pilot experiment have been presented, including MOS and interaction analysis for 5 light field images compressed with two different codecs.

ACKNOWLEDGMENT

This work has been conducted in the framework of ImmersiaTV project under the European Unions Horizon 2020 research and innovation program (grant agreement no. 688619) and funded by Swiss State Secretariat for Education, Research and Innovation SERI; and also in the framework of the Swiss National Foundation for Scientific Research (FN 200021_159575) project Light field Image and Video coding and Evaluation (LIVE).

REFERENCES

- [1] F.-C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 60, 2015.
- [2] A. Maimone and H. Fuchs, "Computational augmented reality eyeglasses," in *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pp. 29–38, IEEE, 2013.
- [3] M. Liu, C. Lu, H. Li, and X. Liu, "Near eye light field display based on human visual features," *Opt. Express*, vol. 25, pp. 9886–9900, May 2017.
- [4] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 220, 2013.
- [5] F.-C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar, "Eyeglasses-free display: towards correcting visual aberrations with computational light field displays," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, p. 59, 2014.
- [6] D. Saupé, F. Hahn, V. Hosu, I. Zingman, M. Rana, and S. Li, "Crowd workers proven useful: A comparative study of subjective video quality assessment," in *QoMEX 2016: 8th International Conference on Quality of Multimedia Experience*, 2016.
- [7] P. Paudyal, F. Battisti, A. Neri, and M. Carli, "A study of the impact of light fields watermarking on the perceived quality of the refocused data," in *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2015*, pp. 1–4, IEEE, 2015.
- [8] S. Darukumalli, P. A. Kara, A. Barsi, M. G. Martini, and T. Balogh, "Subjective quality assessment of zooming levels and image reconstructions based on region of interest for light field displays," in *2016 International Conference on 3D Imaging (IC3D)*, 2016.
- [9] P. A. Kara, M. G. Martini, P. Kovacs, S. Imre, A. Barsi, K. Lackner, T. Balogh, *et al.*, "Perceived quality of angular resolution for light field displays and the validity of subjective assessment," in *2016 International Conference on 3D Imaging (IC3D)*, 2016.
- [10] I. Viola, M. Reřábek, and T. Ebrahimi, "Comparison and evaluation of light field coding approaches," *IEEE Journal of selected topics in signal processing*, 2017.
- [11] I. Viola, M. Reřábek, and T. Ebrahimi, "Impact of interactivity on the assessment of quality of experience for light field content," in *9th International Conference on Quality of Multimedia Experience (QoMEX)*, 2017.
- [12] R. Konrad, E. A. Cooper, and G. Wetzstein, "Novel optical configurations for virtual reality: evaluating user preference and performance with focus-tunable and monovision near-eye displays," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1211–1220, ACM, 2016.
- [13] "ITU-REC-P.910: Subjective video quality assessment methods for multimedia applications," 2008.
- [14] M. Reřábek and T. Ebrahimi, "New light field image dataset," in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.
- [15] "ITU-R BT.500-13: Methodology for the subjective assessment of the quality of television pictures," 2012.