

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works

Evaluation of the Concept of Dynamic Adaptive Streaming of Light Field Video

Peter A. Kara, *Member, IEEE*, Aron Cserkaszkzy, *Member, IEEE*, Maria G. Martini, *Senior Member, IEEE*, Attila Barsi, László Bokor, *Member, IEEE*, Tibor Balogh, *Member, IEEE*

Abstract—Light field visualization has progressed and developed significantly in the past years. At the time of this paper, light field displays are utilized in the industry and they are commercially available as well. Although their appearance on the consumer market is approaching, many potential applications of light field technology have not yet been addressed, such as video streaming. In this paper, we present our research on the dynamic adaptive streaming of light field video. In order to evaluate the presented concept of quality switching, we carried out a series of subjective tests, where test participants were shown light field videos containing stallings and switches in spatial and angular resolution.

Index Terms—Light field visualization, Quality of Experience, video streaming, spatial resolution, angular resolution, stalling.

I. INTRODUCTION

THE visual experience provided by light field displays enables a natural sense of 3D, as no special glasses or headgears are required to observe the visualized content. On the level of applications, its potential spans across numerous different fields. In the use case of gaming, one of the most notable opportunities enabled by light field systems is the visualization of direction-selective content, which can completely replace the currently utilized split-screen method. This means that two or more players can play simultaneously on the full spatial resolution of the screen, as different intervals of viewing angles in the field of view (FOV) can be allocated to different players. In fact, nothing prevents users from enjoying different types of entertainment on the same screen at the same time; e.g., while one user can watch a video stream from one position, another user can play a video game from a different position; the sound assigned to each content can be managed by earphones/headphones or even spatial sound. Similarly to conventional video conferencing applications, light field “telepresence” implementations (see, e.g., Figure 1) provide real-time audiovisual data to the members of the call, but due to the glasses-free 3D experience, the sense of presence can be achieved as well, especially in case of life-sized capture and display systems. For professional utilization, 3D design (for instance the structural design of a building) is what comes to mind first, but there are several other usages in which the capabilities of such systems can help. Light field visualization

Peter A. Kara and Maria G. Martini are with the Wireless Multimedia Networking Research Group of Kingston University, London, UK.

Aron Cserkaszkzy, Attila Barsi and Tibor Balogh are with Holografika, Budapest, Hungary.

László Bokor is with the Department of Networked Systems and Services of the Budapest University of Technology and Economics, Budapest, Hungary.

Manuscript received December 29, 2017.



Fig. 1. Experimental telepresence on a HoloVizio 721RC [5].

is also promising for the head-up displays of land and air vehicles, and flight safety and efficiency in air traffic controls can be further increased through the easily observable real spatial position of aircrafts. In the field of medical imaging, 3D visualization in general is already in use and it is expected to replace the conventional cross-sectional imaging. However, currently 3D medical data either requires special glasses for visualization [1] or is displayed on a 2D screen, which can show the investigated part of the human body from multiple directions, but only one at a time. Autostereoscopic displays in general have had limited application for medical imaging due to their performance limitations, although the usefulness of such displays was clearly highlighted [2] [3]. On a light field display, a medical expert can view medical data in real 3D, increasing the accuracy and efficiency of diagnostics. The accurate visualization of the planet’s surface is also important, especially for industrial needs, such as the exploration of gas and oil [4].

In this paper, we particularly focus on the use case of real-time video transmission, particularly light field video streaming. It is an under-investigated topic, as many find it too early to consider such an application on the level of research. Current efforts have a greater focus on still image visualization than video content transmission. For instance, if we look at the large collaborations, the JPEG Pleno framework [6] is already developing a light field format for standardization purposes. MPEG-I for immersive media has only recently started addressing light field. There are indeed certain milestones that need to be passed before implementing light field video streaming services – i.e., reductions in end-user device cost and in streaming data sizes – but the currently available technology already enables research in Quality of Experience (QoE) of light field visualization to be carried out, including real-time video transmission.

Dynamic Adaptive Streaming over HTTP (DASH) [7] [8] [9] [10] for conventional 2D visualization tackles the choice of the lesser evil in QoE: in case of lower available bandwidth, users tend to tolerate a temporary reduction in spatial resolution more than playback interruptions in the forms of stalling. Still, at the end of the day, both options result in the degradation of user experience, and the less frustrating one needs to be chosen. Today, it is a common practice that the user is actually given the choice of manually selecting the video resolution of a given content, overriding the DASH concept of lower-quality representations if necessary (e.g., if the video is a recording of an online game and the viewer would like to read information that is typically written with small font). In QoE research, similar dilemmas are frequent, such as being caught “between the devil and the deep blue sea” [11], which addresses initial waiting times and stalling events, or taking a foul-tasting medicine as “one spoonful or multiple drops” [12], referring to the distribution of total waiting times during video streaming. In the domain of DASH streaming, quality switching and stalling are the most critical impairments, and in the domain of light field visualization, video quality and video applications in general are currently under-investigated, yet hold great potential.

The aim of the research presented in this paper was to evaluate the concept of dynamic adaptive streaming of light field video. Spatial resolution applies to light field content as well, even though the concept of pixels does not apply to the visualized light field, as light rays hit irregular positions (e.g., on the holographic diffuser). However, angular resolution can be considered an even more important descriptor of a given video. It is the ratio of number of source views and the utilized FOV of the display, in case we consider a source that was recorded by a camera array, a moving camera system (e.g., in case of a static scene), or a single static camera (e.g., the turntable method). We thus involved angular resolution as a varying parameter of quality switching, as certain reductions in the density of source views may be tolerated by the users. This leads to the primary research questions of this paper: How tolerable is quality switching in case of light field video streaming? Is it more acceptable than stalling events during playback? Summa summarum, could the concept of dynamic adaptive streaming based on representations with different spatial and angular resolutions benefit light field video streaming?

To address these questions, we carried out two series of subjective tests on a light field display. We rendered light field videos with either quality switching (based on spatial resolution, angular resolution, or both) or stalling events with different durations. The test participants had to compare them in a full paired comparison test, using a 5-point discrete comparison scale. The concept that this paper addresses was initially proposed and briefly introduced in our prior publication [13], and several experiments involving subjective tests were performed to investigate the underlying perceptual thresholds and phenomena, such as the perceived angular resolution [14] [15] [16], spatial resolution [17] and the interdependence between them [18]. The tests in these experiments used still content for quality evaluation purposes. The primary novelty

introduced in this paper with respect to the current state of the art and our previous publications is that it utilizes the knowledge obtained from prior works to create a subjective test experiment that assesses the perceived quality switching in light field video, compared to playback interruption. Therefore, the paper reports a research that is the practical evaluation of our previously introduced theoretical concept. Also, this work directly addresses QoE variation over time, which was not investigated in prior experiments.

The remainder of the paper is structured as follows: Section II introduces the state-of-the-art research in the area of light field visualization quality, with particular focus on the perceived quality. Section III overviews the concept of dynamic adaptive streaming of light field video, and details the prior research, results of which are the pillars of the experimental setup, which is presented in Section IV. The results of the tests are analyzed in Section V, and the paper is concluded in Section VI, pointing out the potential continuations of the work.

II. PERCEIVED QUALITY OF LIGHT FIELD VISUALIZATION

The success of modern systems and services fundamentally depends on the user experience they provide. This makes QoE an essential factor, explains why manufacturers and service providers focus more and more on the satisfaction of users, and helps us understand the significant boost in QoE-related research in the past decades.

Light field technology has been present for over a hundred years: it was first introduced in 1908 under the term “integral imaging” [19]. The QoE of such systems has only been investigated in the past years, as they only recently became commercially available. Hence this is definitely a novel field, yet there are already notable research efforts towards the QoE of light field visualization. Light field acquisition is also vital, as the capturing quality of the content inherently affects its visualization quality, but it is not included in the scope of this paper.

As light field is gaining more attention among researchers and developers, the spelling of the term is slightly changing. Traditionally it is spelled as “light field”, as it is a field of light rays, but “light-field” and “lightfield” are appearing in recent disseminations of knowledge as well. In this paper, we use only “light field” as the spelling of this technical term.

The term light field can denote any set of light rays defined by their origin, direction, as well as other parameters. Even a conventional 2D monitor or a simple LED is projecting a light field, albeit a *very* simple one. In the scope of this article, we are focusing on light field displays that enable a glasses-free 3D visual experience with continuous motion parallax. This is achieved by using the same auto-stereoscopic principle that our eyes use in the natural world: the display recreates a “window of light” by rendering perspective correct images for hundreds of eyeballs, and the user(s) is/are able to move freely within the FOV of the display, receiving the correct view in each of their eyes, thereby perceiving the full 3D visual experience. The magnitude of the horizontal movement allowed within the FOV of the light field display is defined as the baseline.

With this notion, we can separate light field displays with short baselines, in the scale of centimeters, and with wide baselines, in the order of meters. Short-baseline light field technologies have already entered the consumer market in the form of light field cameras by Lytro [20], and in the short-term future, they should reach consumers in the form of head-mounted displays [21]. However, wide baseline is essential to accommodate multiple freely moving users with no additional head-mounted devices or additional tracking technologies. This also means that at any given time only a portion of the FOV is being observed by the user(s) and mechanisms analogous to the tile-based Spatial Relationship Description (SRD) streaming [22] for navigable video and virtual reality could be used in the future to reduce the transmitted light field content, with the drawback that the user(s) would have to be tracked within the FOV.

Although the ITU defines a vast variety of recommendations for QoE assessment (e.g., the assessment of stereoscopic 3DTV systems [23] or the subjective assessment methods for 3D video quality [24]), the subjective test methodologies of light field visualization have not yet been standardized. Such standards are necessary to increase the efficiency of research and to make research results more comparable. There are already direct research efforts towards methodology itself, such as the works of Viola *et al.* [25] or Darukumalli *et al.* [26]. The measurement methodologies of visualization quality on a more objective level (i.e., angular resolution of a display system) are given by the International Display Measurement Standard [27]. The related publications of Kovács *et al.* [28] [29] [30] investigate both subjective and objective quality of light field visualization, with emphasis on spatial and angular resolution [31] [32] [33]. In this paper, angular resolution refers to the content itself, which is the ratio of the number of source views and the utilized FOV of the display, as stated in Section I. In contrast, the angular resolution of the display is defined as “the minimal angle of change that rays can reproduce with respect to a single point on the screen” [32]. In our research, the angular resolution of the display was not a variable, as we used a single light field display for the tests.

The works of Tamboli *et al.* [34] [35] [36] present subjective tests for the validation of their novel objective quality assessment metric (that is quite notable due to the angular component involved), and subjectively evaluate different extents of view synthesis. Both researches used the HoloVizio HV721RC light field display [5], which is a back-projection system with 72 optical modules. A light field display is a back-projection system if the optical modules are placed behind the screen (e.g., the television-like 80WLT [37]) and it is a front-projection system if the modules are on the same side of the screen as the observer (e.g., the C80 light field cinema system [38]).

The effect of view interpolation is also addressed in the work of Cserkaszkzy *et al.* [39], visualizing content on the C80 cinema system. The difference is that while Tamboli *et al.* investigated how the ratio of genuine to synthesized views affects the user experience (the more the views synthesized, the higher the degradation in QoE), Cserkaszkzy *et al.* targeted a potential benefits of synthesis and compared the performance

of interpolation methods. The findings of the latter indicate that the sweeping planes technique [40] performs better when the input is sparse, but interpolating with the disparity based method [41] can provide a similar visual quality if the input is sufficiently dense. If the content is captured by a real camera system (any of those listed in Section I: a camera array, or a single moving or static camera) interpolation for light field visualization can sometimes be an absolute must, if the content angular resolution is not high enough. In case of rendered content, there are less technical constraints and considerations of cost-effectiveness, even though having a higher view density increases both rendering time and data volume. Apart from the degraded visual quality of interpolated views – as interpolation is indeed an estimation and thus cannot provide perfectly synthesized views – one of the biggest current limitation is that it is an offline-only solution, due to the computational requirements. Real-time light field interpolation could be a major technological support to services such as video streaming, since it would be sufficient to transmit data over the network in a lower angular resolution, reducing the load on the network and enabling more efficient bandwidth usage. This means that the views that were not transmitted by the server would be interpolated real-time at the client’s side, restoring the original angular resolution. Although the extent of interpolation has to be kept within reasonable boundaries in order to prevent major visual degradations, the restoration of angular resolution can significantly contribute to the user experience [39].

While the previous works focus on still image quality, the work of Dricot *et al.* [42] addresses the feasibility of a light field video service. The video content was visualized on the C80 cinema system, and the authors used a linear camera setup to record the stimuli. As their camera system consisted of 80 cameras, and the C80 used 80 optical modules to visualize the content, practically one camera corresponded to one projector. They also used rendered content, with matching parameters (the same number of virtual cameras in the same layout). The light field display was calibrated to a FOV of 40 degrees, thus the angular resolution of the visualized content was 0.5 degree or 2 source views per degree. The videos were encoded in a way that certain views were skipped in the process, and they were replaced with synthesized ones, thus the content angular resolution was constant but the visual quality was degraded. The aim of the research was to compare different coding configurations, through subjective tests, and the test participants rated the stimuli on an impairment scale. In the scope of our research, we did not address multiple coding schemes to evaluate the concept of dynamic adaptive streaming; a lossless coding was used for all of the stimuli, in order to prevent coding-based visual artifacts from affecting the perceived quality.

In general, coding and compression will play one of the most important roles in the future of light field visualization. The immense data requirements need to be kept at bay, while preserving a desired level of perceived quality. The works of Viola *et al.* [43] [44] and Paudyal *et al.* [45] address this topic, from the angle of both objective and subjective quality assessment.

Adhikarla *et al.* [46] proposed a live capture system that consisted of 18 cameras. On the visualization side, they used the HV721RC, which had 72 optical modules. In their configuration, each camera was connected to a separate PC node, and each PC node was controlling 4 optical modules. Although an 18-camera capture system would have only provided a fairly low content angular resolution, it needs to be noted that the rendered output provided by the PC nodes was proven to be sufficient. Also, such work can bring the topic of video content angular resolution into spotlight as well, as, on a perceptual level, it can differ from the perception of static images. In future works, we aim to address the difference between the perception of angular resolution intervals of static scenes and videos.

Even if the type of the content (e.g., captured or rendered, static or video) does not differ, the characteristics of the content affect the perception of light field visualization on the display. Indeed, a good example could be the utilization of the depth budget; e.g., an insufficiently low content angular resolution tends to have a higher impact on the components of a scene that appear in front and behind the screen of a light field display. The perception of low content angular resolution in general is also fundamentally determined by the horizontal structural complexity of objects and scenes, if we consider a display that supports horizontal parallax only. The work of Paudyal *et al.* [47] emphasizes the content selection for subjective tests on light field systems.

Of course light field visualization is not limited to passive use cases, such as static scene visualization or video streaming. The work of Adhikarla *et al.* [48] details their experiment on the subjective evaluation of 3D interaction on a small projection-based light field display. The test participants had to use free-hand gestures to interact with the display – namely to touch given surfaces visualized within the FOV – and the movements were tracked by a Leap Motion Controller. The primary objective of the research was to compare the 3D task performance on the light field display with a 2D equivalent. While objective performance was mainly measured by indicators such as average task completion time, the subjective scores were collected via the User Experience Questionnaire (UEQ) [49]. Interactivity is also addressed by the work of Viola *et al.* [50], and it is without a doubt a significant potential for light field systems, but passive visualization use cases, such as video streaming, is expected to maintain a strong position in the future as well.

III. THE CONCEPT OF DYNAMIC ADAPTIVE STREAMING OF LIGHT FIELD VIDEO

Our proposed quality switching protocol (the idea of which was briefly anticipated in [13] and is elaborated and evaluated here) is based on multiple representations of the video content, with different spatial and angular resolution values. In order to conduct a research that evaluates this concept, we first had to explore the perceptual phenomena that affect the QoE of light field visualization. We used static objects and scenes in subjective tests, as they are less resistant to visual degradations, thus we could obtain boundaries of perceptual

thresholds for videos. It is enough to think about how exactly a participant in a subjective test observes the content. Based on the experimental setup of a test, one may have the freedom of movement, and thus can explore the content from multiple angles, which enables a more accurate perception of the smoothness of the motion parallax.

Current displays are typically horizontal parallax only (HPO), as full-parallax light field displays are yet to be developed. Also, since users can achieve more change in positioning horizontally than vertically in the majority of potential use cases, we consider the smoothness of the horizontal motion parallax as an indicator of perceived quality. The higher the angular resolution, the higher level of smoothness that can be achieved. The failure to provide a sufficiently high angular resolution does not only reduce the 3D continuity, but also results in visual degradations such as the crosstalk effect (see Figure 2) or discrete image borders between adjacent source views (also known as jumps). Again, a light field display projects a continuous field of light rays, based on the content, and not a series of 2D images, even though the input itself is indeed commonly a series of 2D images. However, if the input is not dense enough, then the adjacent discrete views of the content overlap each other and are seen simultaneously. This phenomenon is known as the crosstalk effect, which is highly degrading to the 3D experience. As the content angular resolution gets even lower, it can pass a point after which discrete image borders are constantly present. The visualization begins to feel like an actual series of discrete views, as the content jumps between a small set of views as the observers moves horizontally.

In our works regarding these visual phenomena [14] [16], we focused on their effects on QoE. Our initial tests called the need for extensive training phases, as naïve test participants without any prior experience with light field visualization can have a difficult time discriminating the different degrees of angular resolution, causing inconsistencies in the experimental results. The studies involved extensive evaluations of content angular resolution, ranging from 0.33 source view per degree to 3.3 source views per degree, or from 3 degrees to 0.3 degree. The findings indicate a high level of sensitivity towards the aforementioned visual phenomena, resulting in significant degradations of perceived quality at low numbers of source views. In general, we found that users tend to be satisfied with the smoothness of the horizontal motion parallax if the content angular resolution is 2 source views per degree or more, and that visual degradations can heavily penalize the perceived quality below 1 source view per degree; however, depending on the content itself, 1 source view per degree can still provide a sufficient level of quality. With the dynamic adaptive streaming of light field video content, we target lower angular resolutions, that can be adequate, but have reduced requirements in transmission rates.

We also studied how having static observers may increase the tolerance towards angular resolution reduction [15]. Currently there is no consent within the research community whether test participants should be static during the tests and view light field visualization from a fixed viewing angle and distance, or observe the content from multiple discrete viewing

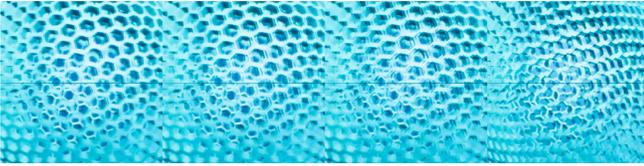


Fig. 2. Content angular resolution reduction, as seen by a single pinhole camera. The polyhedron was visualized with angular resolutions of 0.56, 0.75, 1 and 2.25 degrees (left to right).

positions, or even include actual user movement during assessment. Of course the horizontal motion parallax effect is perceived in all of these cases, even if the observer is static. First of all, there is always at least a given unintentional horizontal movement of the head, plus the horizontal separation of the eyes, and also the changes in the orientation of the eyes. That means that if an observer had an actually physically stabilized head and was viewing the light field visualization with one single eye, the sole movement of that eye would be sufficient to support the cognitive processes regarding the parallax effect. However, the way observers can perceive degradations in visual quality is affected by the viewing conditions. We found that there is a measurable difference between the QoE of a static-user and a moving-user scenario; subjective rating penalization due to low angular resolution was less critical in case of static users. As an example, the “jumps” caused by discrete image borders cannot be perceived in the same frequency and number when the content is watched from a given viewing angle and distance as when users move from the left to the right and vice versa. Static viewing conditions are rather relevant, especially when the user cannot carry out horizontal movements during the visualization of the content, e.g., in case of a light field cinema [51] [52].

As content angular resolution is the ratio of number of views and the FOV they are visualized in, we addressed the FOV of light field displays as well [53]. In our tests, we considered the use case scenario of moving observers, which was also necessary due to the selected test methodology; test participants moved in a given area to discover the FOV value of the test conditions. We used the HoloVizio 80WLT [37] to display the content, which supports FOV up to 180 degrees. In order to achieve different FOV values for the test conditions, we replaced views symmetrically with the background color of the visualized objects. This means that if the user moved outside the selected FOV of the test case, the object was no longer visible. Our findings showed that FOV values above 135 degrees did not benefit the users in this scenario. However, there are indeed use cases where the high number of simultaneous viewers necessitate larger FOV values, e.g., public exhibitions. The relevance to streaming services is that utilizing a smaller FOV can reduce the required transmission rate, as only a given portion of the available data needs to be transmitted. As an example, streaming for only 135 degrees of FOV on a full-horizontal 180-degree light field display can reduce the total data to be sent by 25%. In the protocol presented in this paper, we did not consider variations in FOV, only in spatial and angular resolution.

In current schemes of dynamic adaptive streaming for conventional 2D displays, spatial resolution is one of the key quality indicators that differ between representations. Due to the properties of light field visualization, low content spatial resolution appears differently to the end user; there is no pixelation that is uniform across the entire scene in the plane of the display, instead, the content becomes blurred. We investigated the magnitude of QoE degradation originating from the insufficient content spatial resolution in a paired comparison [17], using a 5-point Degradation Category Rating (DCR) scale [54]. The ratings indicated whether there was any perceivable difference between the resolutions or not, and also the extent of annoyance in case the differences were perceivable. The results indicate that the highest selected resolutions could not be distinguished, and that even the lowest resolutions can be sufficient and adequate, as in comparison to the highest available resolutions, they were commonly rated “slightly annoying”.

In one of our most recent works [18], we investigated how the reductions in spatial resolution affect the perception of angular resolution. Our hypothesis was that the blur caused by the low spatial resolution could actually lessen the visual degradations caused by an insufficient content angular resolution. We conducted a series of subjective tests on numerous static objects and scenes with different content spatial and angular resolutions. In these paired comparisons, the stimuli pairs always had a switch in spatial resolution, while angular resolution remained the same, thus we basically studied spatial resolution reductions at given degrees of angular resolution. We found that in case of high angular resolution (3 source views per degree), reducing the spatial resolution cannot benefit 3D continuity, as is it evidently undisturbed. The gains at 1 source views per degree, however, were greater than in case of 0.66, as such low content angular resolution can prevent the observer from properly perceiving the content itself. We conclude that lowering spatial resolution can indeed support the smoothness of the horizontal motion parallax. This means that during a quality switch that utilizes both resolution parameters, the loss in spatial resolution simultaneously has a potential of lowering QoE by blurring the visualized content but also can compensate for an insufficient angular resolution, reducing the degradations in the parallax effect and thus improving QoE.

It is important to note that dynamic adaptive light field streaming only makes sense if display-independent data is considered (e.g., an array of views), that is to be converted at the client’s side. It is debatable today whether we can consider conversion to be sufficiently fast and efficient for real-time applications. The process fundamentally depends on the input itself; the higher the resolution, the more the time conversion takes. To demonstrate its feasibility, if we take only 18 source views, such as in the live capture system of Adhikarla *et al.* [46], conversion is in the order of 10 ms, which is perfectly suitable for such time-critical use case scenarios. A higher angular resolution, such as 80 source views, can still fit into the order of 100 ms. Using more sophisticated display-independent light field formats can also maintain a low conversion time [55].

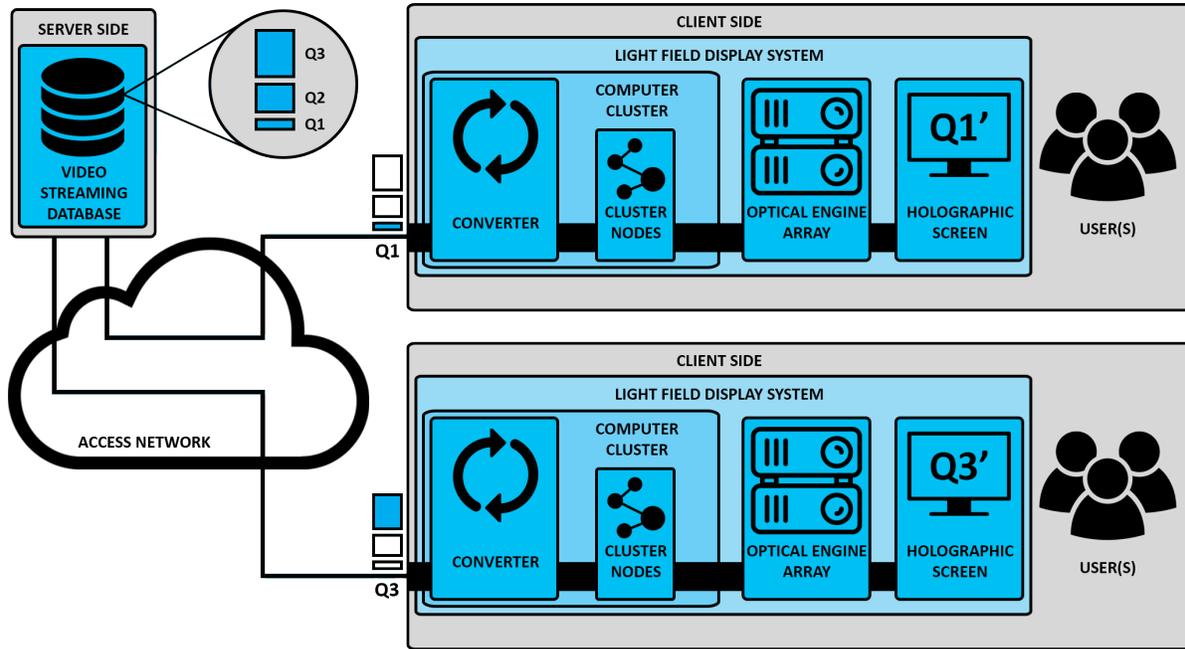


Fig. 3. Dynamic adaptive streaming of $Q1$ and $Q3$ quality representations of a light field video. The illustrated architecture of the light field display system employs a holographic diffuser, analogous to the HoloVizio C80 light field cinema [38].

The feasibility for real-time application can be further exhibited in more details via the light field telepresence solution, presented by Holografika at the T.um technological demonstration exhibition of SKTelecom in late 2017¹ (the paper on the system itself is yet to be published). The system is practically a light field camera system with an integrated light field display. In one room, the light field camera system captures the full body of a standing person with 96 cameras arranged on a 105-degree arc. The camera broadcaster then collects the frames of the cameras and streams them over the network. In the other room, the light field display system, which has a nearly 180-degree FOV, receives the camera frames and on the computer clusters it converts the images to the light field of the display. In this specific implementation, the total system delay is approximately 100 ms, which is measured between the arrival of the given camera frame at the display side of the system and the appearance of the actual visualization on the screen.

Conversion on a given light field display results in fixed parameters of spatial and angular resolution. This means that if a client requests and receives representations of a specific content with quality indicator $Q1$ (low-quality representation) and another client with quality indicator $Q3$ (high-quality representation), the converted and visualized $Q1'$ and $Q3'$ would have the same data volume, but would differ in perceived quality (see Figure 3).

To gain a better understanding of such visualization systems in practice, let us briefly review the functionalities of the primary components and modules present in Figure 3. The views of a given content are captured by a camera array or rendered on computers from virtual scenes. These views are

stored in the video streaming database with different angular and spatial resolutions, resulting in different quality properties and storage/bandwidth requirements. The client continually requests frames of the selected content with a given quality parameter. Once the light field display system receives the frame through the access network, it distributes them to the computers in the cluster that are responsible for rendering the 3D light field on the optical engines, that project light onto the holographic screen. The conversion process in the computer cluster interpolates the continuous horizontal-only-parallax light field from the received discrete camera views, and from this it renders the ideal optical engine images for the specific type of light field display system. The cluster nodes modify these optical engine images by applying the calibration parameters of the particular display system and send these images to the optical engines. The user(s) then is/are able to experience the 3D light field on the display system.

Sending display-specific data can avoid the phase of conversion and thus can make the process of visualization faster. However, in that case, the server side would need to store the corresponding converted data for every single supported display type. Also, if the capabilities of the display system surpass the original parameters of the content, then the converted video is larger as well. As an example in a conventional 2D setting, let us imagine a 720p video. If we want to stream this to an end-user system with a UHD/4K display, that means we would have to transmit a version that is upscaled to 2160p, if we follow the concept of display-specific data transmission. So on one hand, sending videos that are already converted to system specifications eliminates the phase of conversion at the user's end, but it disables the option for dynamic adaptive streaming and can also result in inefficient data transmission rates.

¹<http://tum.sktelecom.com/eng/bbs/bbsDetail.do?bbsSeq=327&bbsTpCd=20>

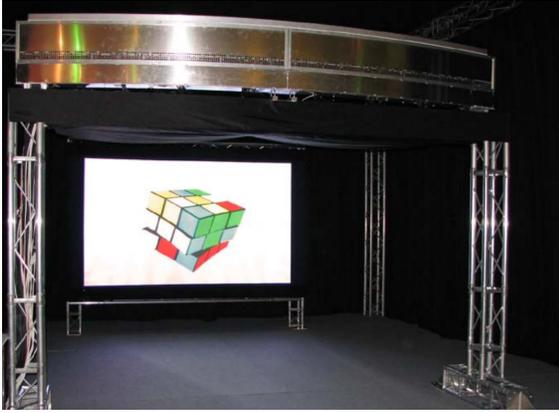


Fig. 4. Interactive Rubik's Cube visualized on the HoloVizio C80 [38].

As for frame rate, the projector array of such systems can support visualization up to 60 FPS. However, in real-time utilization, although 20-25 FPS is manageable, 60 FPS would be *very* challenging. Also, going below 20-25 FPS in playback can easily threaten the user experience. Thus, in the scope of this research, we did not consider the optimization of the temporal domain, and we limited playback to 25 FPS.

It needs to be noted that the concept of dynamic adaptive streaming for light field content can also be applied to dynamic objects. As a simple example for a dynamic object, let us take the well-known Hungarian invention, the Rubik's Cube (see Figure 4). There are two major cases: either the server renders the cube or the client does. In case of client-side rendering, dynamic adaptive streaming is of course unusable. If the server renders it, the rendered output may either be a series of views, a native light field format (specially rendered for the capabilities of the client's system), or even something in between (e.g., the display-independent light field format proposed by Cserkaszky *et al.* [55]). From these, a logical solution would be to render the cube in the native format, as it is assumed to be controlled by a given user and it is most likely to be viewed by the user(s) on a single display. However, in case of multiple viewers (and also multiple controllers, if the dynamic content is more complex than the Rubik's Cube), rendering views and transmitting them to the clients is a feasible option, as viewers may have different visualization systems. With sufficient computational power at the server side, multiple quality representations can be rendered, and transmitted to the users based on access network capabilities. The example intends to demonstrate that such concept is not necessarily limited to video streaming; however, in the scope of this paper, only streaming multimedia content is investigated.

IV. EXPERIMENTAL SETUP

In this section, we introduce the structure of the tests we carried out, the protocols, the content, and all the important parameters that were selected. To the best knowledge of the authors, this paper is the first work addressing adaptive quality switching for light field video and its relevant subjective quality aspects.

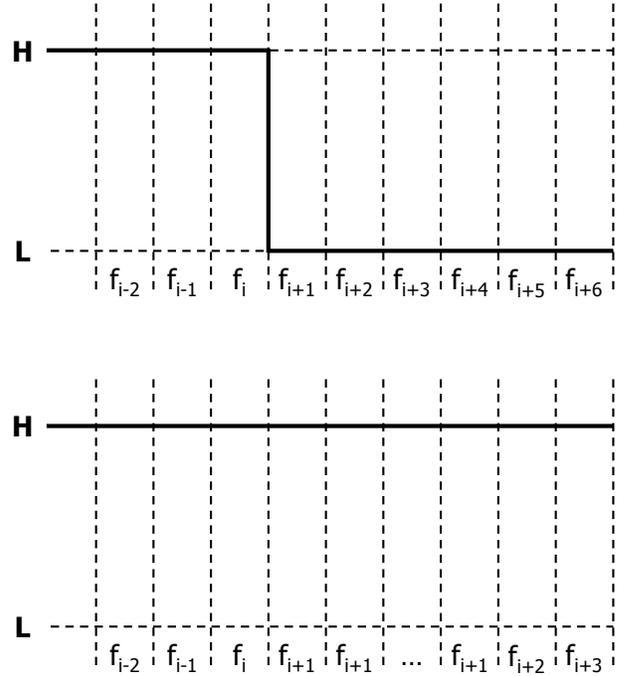


Fig. 5. The implementation of quality switching (top) and stalling (bottom). H and L are the high and low quality representations, respectively, and f represents the frames of the video. For stalling, the length of the event is determined by the number of the repeated f_{i+1} frames.

A. Test Conditions and Video Content

In order to evaluate the concept of dynamic adaptive light field streaming, we created video stimuli with quality switches and stalling events; the variables in the videos were the spatial resolution, the angular resolution and the stalling duration. There were 4 types of test conditions for each source content: (a) quality switching with spatial resolution reduction, (b) quality switching with angular resolution reduction, (c) quality switching with spatial and angular resolution reduction, and (d) stalling event. Each variable in these conditions had two parameters in the test. As the goal of this work was to analyze the effects of stalling and reductions in resolutions on the perceived quality in a controlled manner, instead of considering an actual adaptation strategy based on a real bandwidth model, we chose an ad-hoc approach, and used the following fixed parameters.

The stalling event was either 500 ms or 1500 ms long; these are typical values in the related research [56] [57] [58] [59]. We chose 500 ms as it is above the threshold of being a just noticeable difference (JND), yet it can be easily tolerated. 1500 ms, on the other hand, is more difficult to tolerate and can be considered as a significant stalling. In further experiments, we plan to include stalling durations based on the duration of the source videos or ones that are calculated from the different bandwidth requirements, and also test different buffering schemes. The stalling events in this given study were implemented as frame repetitions (or frame freezing) without graphic indicators.

TABLE I
INVESTIGATED SPATIAL RESOLUTIONS IN QUALITY SWITCHING.

Source ID	High resolution	Low resolution
Red	1024 × 768	640 × 480
Yellow	1024 × 768	800 × 600
Ivy	960 × 540	640 × 360
Tesco	1280 × 720	640 × 360
Gears	1920 × 1080	640 × 360

TABLE II
INVESTIGATED TEST CONDITION PAIRS.

Test ID	Quality switching	Stalling duration
A	Spatial	500 ms
B	Spatial	1500 ms
C	Angular	500 ms
D	Angular	1500 ms
E	Both	500 ms
F	Both	1500 ms

We decided to create video stimuli with stalling events in this manner as this can be considered to be one of the most common approaches for Video on Demand (VoD) services when quality switching is not implemented; the last visualized frame freezes on the screen until playback continues. The only major deviation from the general practice is the lack of a typical spinning, circular graphical indicator. Its implementation would have been possible, either in 2D or 3D, however, we found it visually distracting for the observers, and it would have made stalling duration very explicit, disregarding certain features of the content.

Content angular resolution was either 1 or 2 source views per degree (1 or 0.5 degree). We chose the higher value because according to previous studies [42] [16], it can provide a smooth horizontal motion parallax and thus a good user experience (at least in that regard). The lower value was chosen for multiple reasons. First of all, our most recent work [18] pointed out the potential gain at this content angular resolution, that might compensate in the overall QoE during quality switching with both resolutions. Furthermore, prior researches [36] [14] argue whether this can be considered as a sufficiently high value for light field visualization or not.

Quality switching regarding content spatial resolution varied for each source video (see Table I), and we also used two different aspect ratios. We based our choice of the spatial resolutions on our prior findings in the area [17], and we also aimed at having diversity regarding the differences in perceived quality.

It is vital to point out that this work was focused on “down-switching”, so quality switching was always performed as a sudden change from higher to lower spatial and/or angular resolution. Also, each and every stimulus only contained either one quality switching or a stalling event, and the frame of quality switching and stalling was always at the middle of the content.

Figure 5 demonstrates the implementation of quality switching and stalling in our experiment. For each source video, f_i represents the frame at the middle of the video (at the half of the duration). As either quality switching or stalling were applied only from f_{i+1} , the stimuli frames between f_1 and f_i were identical, with quality H (high quality) for both angular and spatial resolution. In case of quality switching, if we denote the final frame of a given stimulus as f_n , then from frame f_{i+1} to f_n the stimulus was shown in quality L (low quality), where L either had a reduced spatial resolution, angular resolution or both. As stalling was implemented as frame repetition, f_{i+1} was repeated according to the different stalling durations, and then was followed by f_{i+2} and the rest of the frames until f_n . This means that no frames were skipped, and stalling increased to total duration of the stimulus.

We used 5 short video source contents in our experiment, played back at 25 FPS. *Red* (14.4 sec duration, 130,240 kbps encoding rate) and *Yellow* (13.6 sec, 704,160 kbps) were sets of moving and rotating columns, highly utilizing the available depth budget of the system, and thus also making content angular resolution critical. *Ivy* (10 sec, 38,000 kbps) was a subtle animation of an ivy plant growing around a statue, with fine structural details. *Tesco* (12.5 sec, 76,800 kbps) was a scene with significant motion vectors². *Gears* (7.2 sec, 286,000 kbps) was a collection of rotating gears, basically a fast-paced looped animation. As compression schemes and methods were not addressed by our research, we applied a lossless variation of H.264 to the camera views individually.

In future works, we intend to involve longer sequences in our research, and also target more complex content. Such content could be Big Buck Bunny – which is very well known and used in subjective quality assessments – but in light field format [60].

B. Subjective Quality Assessment Tasks

As there were 5 source videos and 5 test conditions – or Hypothetical Reference Circuits (HRC), according to the terminology of the Video Quality Experts Group (VQEG) – with 3 quality switchings and 2 stalling events, the total number of video stimuli – or Processed Video Sequences (PVS) – was 25. The subjective test was a paired comparison, in which the different quality switching types were compared to the stalling events. There were thus 6 comparisons, which were applied to the 5 sources, so the total number of comparisons was 30. This also means that a test participant viewed 60 video stimuli during the test. Furthermore, this is one of the reasons why the stimulus duration was limited: using longer source videos would have resulted in a prolonged total test duration, as every extra second in average source length would have meant 1 minute more for the total duration. The other reason is that the experiment was centered on the quality switching event itself, and one stimulus only contained a single switching event. Similar video stimulus durations were used in the recent work of Duanmu *et al.* [61], addressing streaming QoE.

²*Ivy* and *Tesco* were created by Post Edison (<http://www.postedison.hu/>) and were provided to Holografika for research purposes.

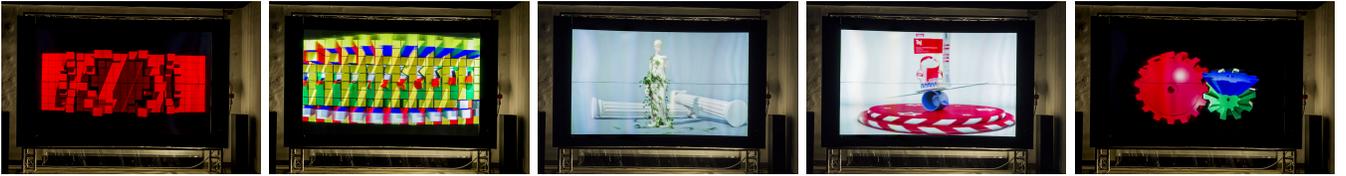


Fig. 6. Source video stimuli of the research (*Red, Yellow, Ivy, Tesco* and *Gears*), visualized on the HoloVizio C80 light field cinema system [38].

The evaluation was performed on a 5-point comparison scale (“*Much worse* (−2)”, “*Worse* (−1)”, “*Same* (0)”, “*Better* (+1)”, “*Much better* (+2)”). The task of the test participants was to compare the second stimulus in the pair to the first one. The comparison task targeted the overall QoE, taking every aspect of perceived quality into consideration. The notion of QoE was of course limited to the visual experience, as the test stimuli contained no audio. The test pairs and also the videos inside the pairs were separated by a 5-second blank screen [23].

We also considered using a 5-point ACR to rate the 25 video stimuli. As the focus of the research questions was on subjective preference, we decided to use a paired comparison in this work of concept evaluation. Different subjective assessment scales and methodologies are to be addressed in future works.

C. Light Field Display and Test Environment

At the time of this paper, there are already several real implementations of light field displays, such as the Nvidia near-eye light field display [21], the nearly 90-inch display of Sang *et al.* [62], the 100-inch display of Lee *et al.* [63], the 200-inch display of Inoue *et al.* [64] or the lenticular-based optical system of Hirsch *et al.* [65] that was inspired by the angle-expanding Keplerian telescopes. However, most of such systems are experimental and are not available commercially. The light field display we used for the subjective tests was the HoloVizio C80 light field cinema [38] [66]. As mentioned earlier, the C80 is a commercially available front-projection system with 80 optical modules. The screen that performs the optical transformations of the light rays is 3 meters wide and has a brightness value of 1500 cd/m². During the research, the display was calibrated for a FOV of 45 degrees. This means that the video stimuli with 0.5 degree content angular resolution consisted of 90 source views, and 1 degree corresponded to 45 views.

The tests were performed in a laboratory environment, isolated from external audiovisual distractions. The lighting condition of the environment was 20 lx. In the selection of the viewing conditions, we had to consider the viewing distance, the viewing angle, and the mobility of the test participants. In related works, Tamboli *et al.* [34] used 1.7H as viewing distance with fixed observer positions in 5 viewing angles, and Dricot *et al.* [42] used 3.3H with 6 fixed positions, and also moving observers with a sideways distance of 2 meters. In our work, we chose a viewing distance of 2.5H, which corresponded to 4.6 meters, and also enabled a sideways movement of approximately 1 meter.

For front-projection light field systems, there is an additional constraint for viewing distance, compared to back-projection

ones. As the optical engine array is located on the same side of the screen as the observer, there is the possibility of light ray occlusion if the observer is viewing the content in a position between the screen and the projectors. Should that happen, the observable visual data would be incomplete, and the shadow of the observer would appear on the screen. Therefore, we chose the viewing distance in our research to be just behind the line of the optical engine array.

We selected the default viewing angle as 0 degree, so initially test participants viewed the content directly from the middle. The sideways movement enabled them to have a better observation of the smoothness of the horizontal motion parallax. However, parameters such as viewing distance, viewing angle and observer motion may affect the perception of light field videos, which we aim to address in further researches.

V. RESULTS

In this section, we introduce the results obtained from the tests on subjective quality comparison. As detailed in Section IV, every comparison pair contained one video stimulus with quality switching and one with stalling. In the analysis of this section, positive values indicate the preference of stalling, and negative values indicate the preference of quality switching (e.g., a score of −2 means that a given stimulus with quality switching was rated “*Much better*” compared to the stimulus it was paired with, containing stalling).

A total of 20 test participants (16 male and 4 female) took part in our experiment. The age interval of the participants was from 20 to 38, and the mean age was 26. Before participating in the tests, each individual was screened for vision using the Snellen charts and Ishihara plates.

Figure 7 shows the mean comparison scores we obtained for the test conditions defined in Table II, Figure 8 provides the distribution of the scores, and Table III contains the corresponding statistical analysis. Both Tukey HSD, Bonferroni and Holm multiple comparisons concluded the same binary significance values (either all deemed a pair to be significantly different or none of them did). These results are also in alignment with the overlaps between the 0.95 confidence intervals of Figure 7.

When the quality switches were compared to a high stalling duration of 1500 ms (condition *B*, *D* and *F*), typically the quality switches were preferred; the preference of stalling was 6% or less in all three conditions. Yet it needs to be noted that in this research angular resolution was only reduced from 0.5 to 1 degree, which can be considered a borderline of toleration [14] [15] [16]. We expect that further, larger extents of reductions in angular resolution could easily reverse this

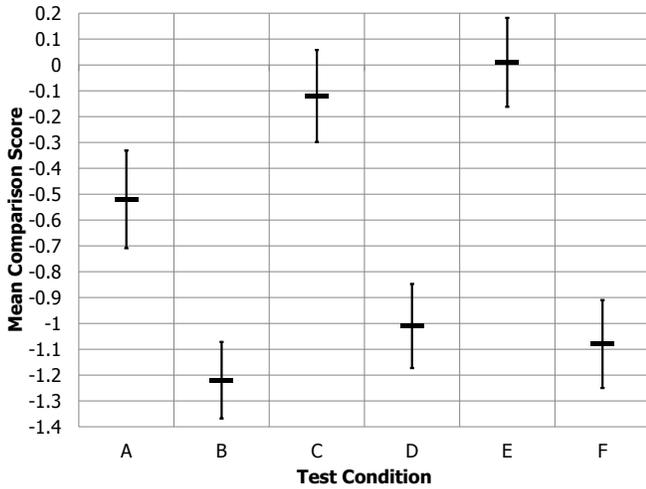


Fig. 7. Mean scores of the investigated test conditions with 0.95 CI.

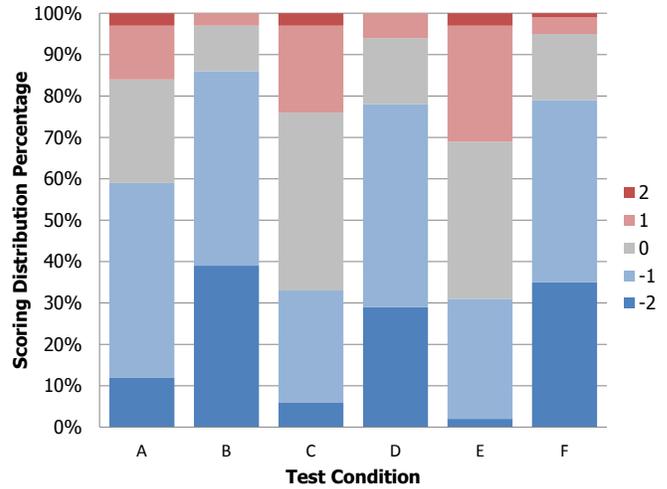


Fig. 8. Distribution of comparison scores per test condition.

TABLE III
STATISTICAL ANALYSIS OF THE INVESTIGATED TEST CONDITIONS. THE P-VALUES OF TUKEY HSD (T), BONFERRONI (B) AND HOLM (H) MULTIPLE COMPARISONS ARE GIVEN, ALONG WITH SIGNIFICANCE (S).

		T	B	H	S
A	B	0.00	0.00	0.00	Yes
A	C	0.01	0.00	0.00	Yes
A	D	0.00	0.00	0.00	Yes
A	E	0.00	0.00	0.00	Yes
A	F	0.00	0.00	0.00	Yes
B	C	0.00	0.00	0.00	Yes
B	D	0.52	1.34	0.35	No
B	E	0.00	0.00	0.00	Yes
B	F	0.85	3.86	0.77	No
C	D	0.00	0.00	0.00	Yes
C	E	0.89	4.40	0.58	No
C	F	0.00	0.00	0.00	Yes
D	E	0.00	0.00	0.00	Yes
D	F	0.89	8.57	0.57	No
E	F	0.00	0.00	0.00	Yes

TABLE IV
DATA RATE REDUCTION THROUGH LOWERED SPATIAL AND/OR ANGULAR RESOLUTION FOR EACH VIDEO CONTENT. 25% IN THE TABLE MEANS THAT ONLY A QUARTER OF THE CORRESPONDING HIGH-RESOLUTION DATA WAS REQUIRED FOR LIGHT FIELD VISUALIZATION.

Source ID	Spatial	Angular	Both
Red	39%	50%	19.5%
Yellow	61%	50%	30.5%
Ivy	44.4%	50%	22.2%
Tesco	25%	50%	12.5%
Gears	11%	50%	5.5%

ratio, as users would rather wait 1500 ms than face severe visual degradations (e.g., see Figure 2).

Generally, it can be stated that the quality switching based purely on spatial resolution (condition A and B) performed better than the other two types, as it is reflected in both mean scores and scoring distribution. However, from a statistical point of view, there is no significant difference between the scores of condition B, D and F.

The same does not apply to the comparisons with a low 500 ms stalling (condition A, C and E). While C and E were rather balanced in preference (see Figure 8), in condition A quality switching was preferred by 59% of the test participants and stalling was chosen by 16%.

The most important finding in these results is the negligible difference between quality switching using angular resolution reduction only (C and E) and using both angular and spatial resolution (D and F). If the difference between them regarding perceived quality is such, then the combined switching can be

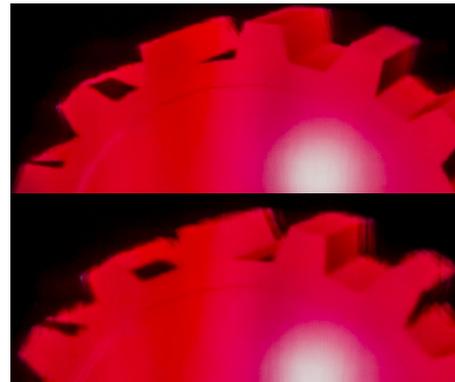


Fig. 9. A part of *Gears* before (top) and after (bottom) a quality switch, reducing both spatial and angular resolution.

used in a practical application of the protocol, which can come with a major reduction in bandwidth requirement compared to angular switching, without compromising user experience.

As an example for bandwidth requirement reduction, let us take the data sizes of *Red*, *Yellow*, *Ivy*, *Tesco* and *Gears* at full spatial and angular resolution. Decreasing spatial and/or angular resolution by selecting lower-quality segments evidently reduces the transmission data rate, as the segment sizes (in bytes) are smaller. Table IV shows how this applies to our source video stimuli, e.g., the size of *Gears* at combined low

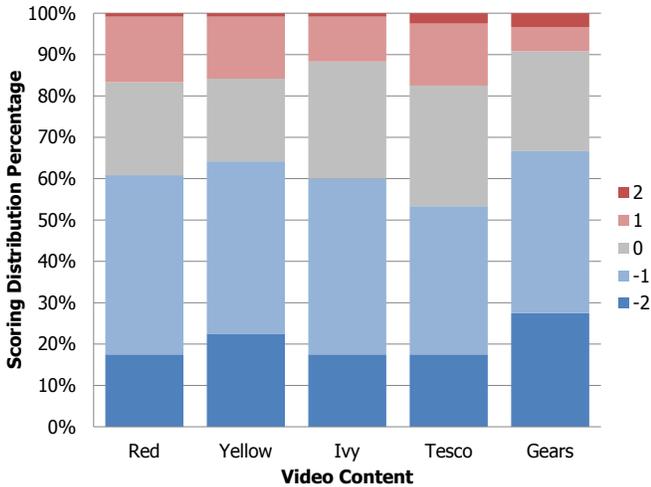


Fig. 10. Distribution of comparison scores per video content.

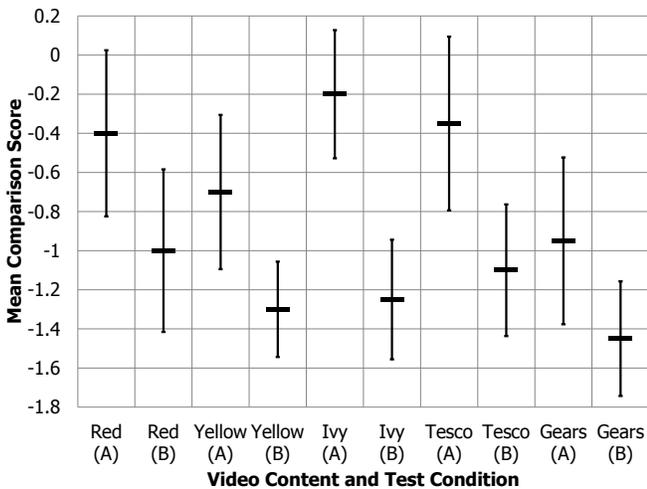


Fig. 11. Mean scores of test conditions A and B per content with 0.95 CI.

resolutions is 5.5% of the size at full spatial and angular resolution (corresponding to a compression ratio of approximately 18:1).

As investigated in our earlier work [18], reducing spatial resolution when angular resolution does not provide a continuous horizontal motion parallax with undisturbed smoothness can lessen the impact of visual phenomena such as the crosstalk effect. Figure 9 compares frame f_i and f_{i+1} (i.e., before and after quality switching) from the sequence *Gears*, where the quality switch included both resolutions. Although certain levels of the crosstalk effect and ghosting were visible, the blur induced by the lowered spatial resolution applied to the entire scene, including the visual degradations that disturbed the parallax effect. This was able to mask the insufficient angular resolution to a given extent. Therefore, the blur reduced the effect of such visual phenomena from a perceptual perspective.

The selection of the source video content used in our experiment did not have a significant impact on the obtained results; no statistically significant difference was found between any two of the contents. There are of course certain extents of

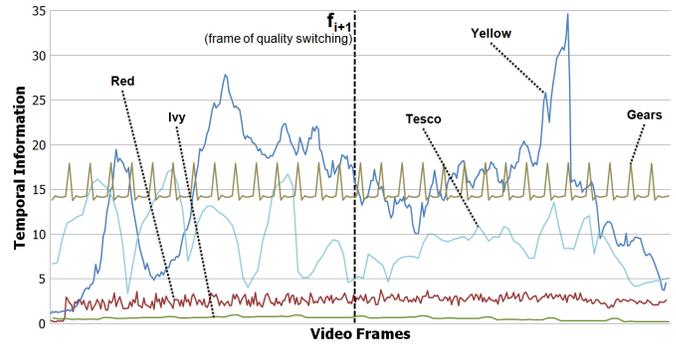


Fig. 12. Temporal Information of the source video stimuli.

differences, visualized in the distribution of the scores per video content (see Figure 10). These primarily originate from condition *A* and *B*, which compared quality switching with spatial resolution reduction to stalling. Condition *C*, *D*, *E* and *F* do not deviate much per source; in fact, some even have the exact same mean values. The only exception is *Tesco*, which obtained less scores favoring quality switching, due to the high mobility of the scene.

As described in Table I, the five contents switched between different spatial resolutions. Based on the selected values, the descending order of switching magnitude (percentage of difference in source pixel count, see Table IV) was *Gears*, *Tesco*, *Red*, *Ivy* and *Yellow*. Theoretically, this would imply that spatial resolution reduction affected *Gears* the most on the level of perception, and contents such as *Ivy* and *Yellow* were less affected. However, the results report the opposite (see Figure 11).

Although there is no statistically significant difference between the source contents in the analysis focusing on *A* and *B*, the relations between the mean comparison scores are quite noteworthy; for both *A* and *B*, *Gears* was the least affected, while *Ivy* received comparably low preference scores for quality switching, particularly for *A*. The results indicate that the content itself had a greater influence than the change in source spatial resolution. Even though *Ivy* was limited to a subtle animation, the additional blur due to quality switching degraded the visual appearance of the statue and the detailed ivy plant growing around it. On the other hand, in case of *Gears*, this quality transition only softened the edges of the rotating gears, even though the source spatial resolution was reduced from 1920×1080 to the same 640×360 as *Ivy* (see Table I). As for *Red* and *Yellow*, their comparison scores were more in alignment with the difference in source spatial resolutions, for both *A* and *B*.

It needs to be noted that when stalling events are investigated, Temporal Information (TI) [54] can support content classification, as it describes the difference between adjacent video frames. Figure 12 shows the application of the conventional 2D TI measurement to the middle source camera view. The subtle animation of *Ivy* barely registered in the measurement, while the multiple rotating columns of *Yellow* resulted in high levels of TI. As *Gears* was a short looping animation, this is well reflected in the repeating TI pattern. However, such

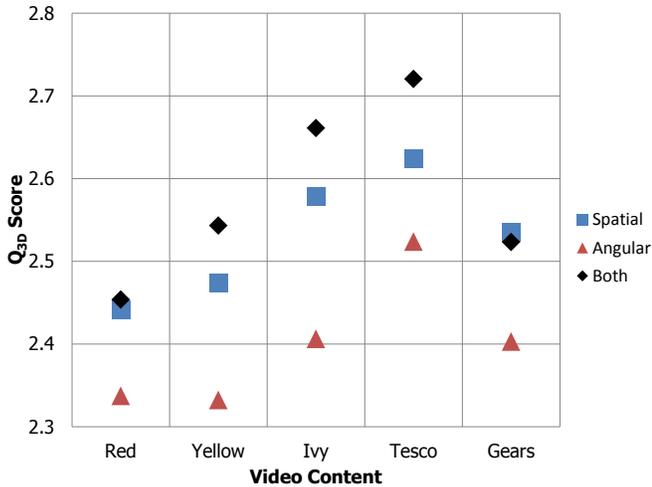


Fig. 13. Q_{3D} scores at frame f_{i+1} per video content and $\alpha = 0.89$. Higher objective scores suggest higher levels of QoE degradation.

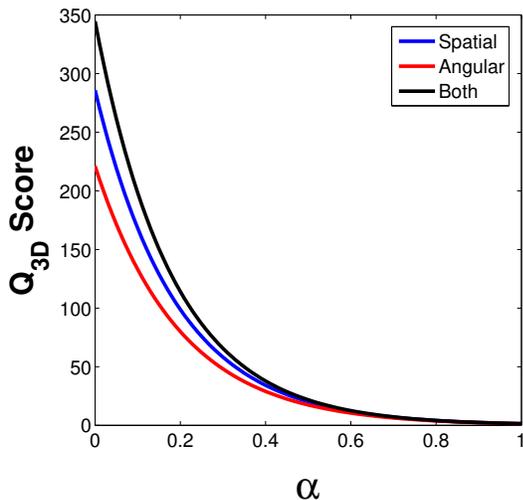


Fig. 14. Q_{3D} scores of *Gears* at frame f_{i+1} . Higher objective scores suggest higher levels of QoE degradation.

application of TI cannot measure motions and alterations along the z axis (depth); TI applied to *Red* mainly measured the shadows cast by the columns that moved closer to the observer, instead of the actual movement. Therefore, in future works, we aim to develop a TI metric for discrete light field content (camera view array), for accurate content classification. Such knowledge regarding the content is particularly important, as the utilization of the depth budget can fundamentally affect perceptual sensitivity towards the parallax effect and thus the requirements for angular resolution.

The obtained subjective test results were also compared to the full reference (FR) objective quality metric proposed by Tamboli *et al.* [34], which was selected due to its consideration for the angular quality component. The Q_{3D} values of the metric are calculated as:

$$Q_{3D} = Q_{2D} \cdot \left(\frac{Q_{\theta}}{Q_{2D}}\right)^{\alpha} \quad (1)$$

where Q_{2D} is the spatial component, which is based on the transformation of images into a parameter space and their comparison in that space, using a steerable pyramid decomposition; Q_{θ} is the angular component, calculating MS-SSIM for optical flow vectors; and α is a parameter whose value is between 0 ($Q_{3D} = Q_{2D}$) and 1 ($Q_{3D} = Q_{\theta}$).

The metric was applied to frame f_{i+1} (see Figure 5) of each stimulus with a quality switching event, and the reference was the corresponding frame with high angular and spatial resolution. We used for this evaluation the α value of 0.89 set by Tamboli *et al.*, which was determined via a 1000-fold cross-validation, based on their subjective quality assessment scores. The objective Q_{3D} scores (see Figure 13) fit into an interval of 0.4 (the difference between the highest and lowest value was 0.39). This indicates small differences between the degradations of the contents, as this metric in practice can provide Q_{3D} scores between 5 and 10 for distorted content at an α of 0.89. However, in our experiment, the contents were not degraded visually (e.g., via added noise), but only by reducing spatial and/or angular resolution. The Q_{θ} value, which runs between 0 (most extreme extent of angular distortion) and $\sqrt{2}$ (no measurable angular distortion) deviated the most in case of *Gears* (1.37), due to the sharp edges in the scene. Its Q_{3D} scores based on the possible α values are shown in Figure 14.

According to these objective results, shifting down from the angular resolution of 2 views per degree to 1 had a lesser impact on the estimated scores than the changes in spatial resolution, for all video content (see Figure 13 and 14). However, this is in contradiction with the obtained subjective results (see Figure 7 and Table III). Yet it needs to be noted that the authors of the metric strictly used reference visual data with an angular resolution of 1 view per degree during every process of creation, including calibration and evaluation.

Furthermore, as the objective metric was designed for still content, it is difficult to efficiently apply it to light field video frames, as such frames might not be as clear and sharp as a static scene, due to the motion and changes in the content. In such case, motion blur and other visual degradations affect the reference of the FR metric, also modifying the quality of the image (video frame) to which spatial and/or angular resolution reduction is applied. For accurate QoE estimation, light field video metrics would be necessary. However, at the time of this paper, no objective quality metric for light field video exist, as video in general for this visualization technology is currently under-investigated.

VI. CONCLUSION

In this paper, we introduced the concept of dynamic adaptive streaming for light field video and its evaluation through QoE studies. Our results indicate that quality switching that applies to spatial and/or angular resolution is clearly preferred instead of long stalling events and can also be preferable compared to short ones. The primary finding of our work is that the overall QoE of quality switching using both spatial and angular resolution is statistically indistinguishable from reducing angular resolution only. We conclude that in case of

adaptive streaming, it is beneficial to choose a video representation with both spatial and angular resolution reduced, if the smoothness of the continuous horizontal motion parallax cannot be guaranteed.

In future works, we aim to further investigate the QoE of light field video streaming, with longer sequences and with different stalling distributions, including various event frequencies and durations. We shall also investigate an extended range of angular resolution variation, as severe reductions (below 1 degree) may even result in the clear preference of longer stalling events. In the area of measurement methodologies, our targeted research questions cover subjective quality assessment scales and viewing conditions, with particular focus on viewing distance and observer movement for video rating tasks. Our long-term goal is to develop and test streaming solutions and schemes for cost-efficient real-time transmission of light field videos.

ACKNOWLEDGMENT

The research in this paper was done as a part of and was funded from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 643072, QoE-Net, and No. 676401, ETN-FPI. The authors would also like to thank Roopak R. Tamboli and the Indian Institute of Technology Hyderabad for the support in objective light field quality estimation.

REFERENCES

- [1] C. T. Hewage, M. G. Martini, and N. Khan, "3D medical video transmission over 4G networks," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*. ACM, 2011, p. 180.
- [2] Y. Narita, S. Tsukagoshi, M. Suzuki, Y. Miyakita, M. Ohno, H. Arita, Y. Saito, Y. Kokojima, N. Watanabe, N. Moriyama, and S. Shibui, "Usefulness of a glass-free medical three-dimensional autostereoscopic display in neurosurgery," *International journal of computer assisted radiology and surgery*, vol. 9, no. 5, pp. 905–911, 2014.
- [3] D. S. Magalhães, R. L. Serra, A. L. Vannucci, A. B. Moreno, and L. M. Li, "Glasses-free 3D viewing systems for medical imaging," *Optics & Laser Technology*, vol. 44, no. 3, pp. 650–655, 2012.
- [4] T. Balogh and P. T. Kovács, "Holovizio: The next generation of 3D oil & gas visualization," in *70th EAGE Conference and Exhibition-Workshops and Fieldtrips*, 2008.
- [5] Holovizio 721RC light field display, www.holografika.com/Documents/Holovizio_721RC-emailsize-nomovie.pdf (retrieved June 2017).
- [6] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.
- [7] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the Internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, 2011.
- [8] T. Stockhammer, "Dynamic adaptive streaming over HTTP—standards and design principles," in *Proceedings of the second annual ACM conference on Multimedia systems*. ACM, 2011, pp. 133–144.
- [9] O. Ognenoski, M. M. Nasralla, M. Razaak, M. G. Martini, and P. Amon, "DASH-based video transmission over LTE networks," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 1783–1787.
- [10] O. Oyman and S. Singh, "Quality of experience for HTTP adaptive streaming services," *IEEE Communications Magazine*, vol. 50, no. 4, 2012.
- [11] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: Between the devil and the deep blue sea," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*. IEEE, 2012, pp. 1–6.
- [12] P. A. Kara, M. G. Martini, and S. Rossi, "One Spoonful or Multiple Drops: Investigation of Stalling Distribution and Temporal Information for Quality of Experience over Time," in *International Conference on Telecommunications and Multimedia (TEMU)*, 2016.
- [13] P. A. Kara, A. Cserkaszy, A. Barsi, M. G. Martini, and T. Balogh, "Towards Adaptive Light Field Video Streaming," *IEEE COMSOC MMTC Communications – Frontiers*, 2017.
- [14] P. A. Kara, M. G. Martini, P. T. Kovács, S. Imre, A. Barsi, K. Lackner, and T. Balogh, "Perceived quality of angular resolution for light field displays and the validity of subjective assessment," in *International Conference on 3D Imaging (IC3D)*, Liege, 2016.
- [15] P. A. Kara, A. Cserkaszy, S. Darukumalli, A. Barsi, and M. 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.
- [16] P. A. Kara, P. T. Kovács, S. Vagharshakyan, M. G. Martini, A. Barsi, T. Balogh, A. Chuchvara, and A. Chehaibi, "The Effect of Light Field Reconstruction and Angular Resolution Reduction on the Quality of Experience," in *12th International Conference on Signal Image Technology & Internet Based Systems (SITIS) 3rd International Workshop on Quality of Multimedia Services (QUAMUS)*, Naples, 2016.
- [17] P. A. Kara, P. T. Kovács, M. G. Martini, A. Barsi, K. Lackner, and T. Balogh, "Viva la Resolution: The Perceivable Differences between Image Resolutions for Light Field Displays," in *5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS 2016)*, Berlin, 2016.
- [18] P. A. Kara, A. Cserkaszy, A. Barsi, T. Papp, M. G. Martini, and L. Bokor, "The Interdependence of Spatial and Angular Resolution in the Quality of Experience of Light Field Visualization," in *2017 International Conference on 3D Immersion (IC3D)*. IEEE, 2017.
- [19] G. Lippmann, "Epreuves reversibles. photographies intégrales," *Comptes-Rendus Academie des Sciences*, vol. 146, pp. 446–451, 1908.
- [20] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma, "Lytro camera technology: theory, algorithms, performance analysis," in *Multimedia Content and Mobile Devices*, vol. 8667. International Society for Optics and Photonics, 2013, p. 86671J.
- [21] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 220, 2013.
- [22] L. D'Acunto, J. van den Berg, E. Thomas, and O. Niamut, "Using MPEG DASH SRD for zoomable and navigable video," in *Proceedings of the 7th International Conference on Multimedia Systems*. ACM, 2016, p. 34.
- [23] ITU-T Rec., "BT.2021: Subjective methods for the assessment of stereoscopic 3DTV systems," 2015.
- [24] —, "P.915: Subjective assessment methods for 3D video quality," 2016.
- [25] I. Viola, M. Rerabek, and T. Ebrahimi, "A new approach to subjectively assess quality of plenoptic content," in *Applications of Digital Image Processing XXXIX*, vol. 9971, no. EPFL-CONF-221562. SPIE, 2016, pp. 99710X–1.
- [26] S. Darukumalli, P. A. Kara, A. Barsi, M. G. Martini, T. Balogh, and A. Chehaibi, "Performance comparison of subjective assessment methodologies for light field displays," in *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2016, pp. 28–33.
- [27] Society for Information Display, "The International Display Measurement Standard v1.03," 2012.
- [28] P. T. Kovács and T. Balogh, "3D visual experience," *High-Quality Visual Experience*, pp. 391–410, 2010.
- [29] —, "3D display technologies and effects on the human vision system," in *2011 2nd International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2011, pp. 1–33.
- [30] —, "3D light-field display technologies," *Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering*, pp. 336–345, 2013.
- [31] P. T. Kovács, R. Bregović, A. Boev, A. Barsi, and A. Gotchev, "Quantifying spatial and angular resolution of light-field 3-D displays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1213–1222, 2017.
- [32] P. T. Kovács, A. Boev, R. Bregović, and A. Gotchev, "Quality measurements of 3D light-field displays," in *Proc. Eighth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2014.
- [33] P. T. Kovács, K. Lackner, A. Barsi, Á. Balázs, A. Boev, R. Bregović, and A. Gotchev, "Measurement of perceived spatial resolution in 3D light-field displays," in *International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 768–772.
- [34] R. Tamboli, B. Appina, S. Channappayya, and S. Jana, "Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display," *Signal Processing: Image Communication*, vol. 47, pp. 42–55, 2016.

- [35] R. Tamboli, K. K. Vupparaboina, J. Ready, S. Jana, and S. Channappayya, "A subjective evaluation of true 3D images," in *2014 International Conference on 3D Imaging (IC3D)*. IEEE, 2014, pp. 1–8.
- [36] R. Tamboli, A. B., S. Channappayya, and S. Jana, "Achieving high angular resolution via view synthesis: quality assessment of 3D content on super multiview lightfield display," in *2017 International Conference on 3D Immersion (IC3D)*. IEEE, 2017.
- [37] Holografika, "Holovizio 80WLT light field display," http://www.holografika.com/Documents/HoloVizio_80WLT-emailsize.pdf (retrieved June 2017).
- [38] —, "HoloVizio C80 light field cinema system," www.holografika.com/Documents/HoloVizio_C80.pdf (retrieved June 2017).
- [39] A. Cserkaszky, P. A. Kara, A. Barsi, and M. G. Martini, "To Interpolate or not to Interpolate: Subjective Assessment of Interpolation Performance on a Light Field Display," in *IEEE International Conference on Multimedia and Expo (ICME) 8th Workshop on Hot Topics in 3D Multimedia (Hot3D)*, Hong Kong, 2017.
- [40] S. Smirnov, M. Georgiev, and A. Gotchev, "Comparison of cost aggregation techniques for free-viewpoint image interpolation based on plane sweeping," *Ninth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2015.
- [41] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 4, pp. 333–340, 1980.
- [42] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, F. Dufaux, P. T. Kovács, and V. K. Adhikarla, "Subjective evaluation of super multi-view compressed contents on high-end light-field 3D displays," *Signal Processing: Image Communication*, vol. 39, pp. 369–385, 2015.
- [43] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *Picture Coding Symposium (PCS), 2016*. IEEE, 2016, pp. 1–5.
- [44] I. Viola and T. Ebrahimi, "A new framework for interactive quality assessment with application to light field coding," in *Proceedings of SPIE*, no. EPFL-CONF-230109, 2017.
- [45] P. Paudyal, F. Battisti, M. Sjöström, R. Olsson, and M. Carli, "Towards the perceptual quality evaluation of compressed light field images," *IEEE Transactions on Broadcasting*, vol. 63, no. 3, pp. 507–522, 2017.
- [46] V. K. Adhikarla, F. Marton, T. Balogh, and E. Gobbetti, "Real-time adaptive content retargeting for live multi-view capture and light field display," *The Visual Computer*, vol. 31, no. 6–8, pp. 1023–1032, 2015.
- [47] P. Paudyal, J. Gutierrez, P. Le Callet, M. Carli, and F. Battisti, "Characterization and selection of light field content for perceptual assessment," in *9th International Conference on Quality of Multimedia Experience (QoMEX 2017)*, 2017.
- [48] V. K. Adhikarla, J. Sodnik, P. Szolgay, and G. Jakus, "Exploring direct 3D interaction for full horizontal parallax light field displays using leap motion controller," *Sensors*, vol. 15, no. 4, pp. 8642–8663, 2015.
- [49] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and Usability Engineering Group*. Springer, 2008, pp. 63–76.
- [50] I. Viola, M. Rerabek, 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.
- [51] P. A. Kara, Z. Nagy, M. G. Martini, and A. Barsi, "Cinema as large as life: Large-scale light field cinema system," in *2017 International Conference on 3D Immersion (IC3D)*. IEEE, 2017.
- [52] P. A. Kara, A. Cserkaszky, A. Barsi, and M. G. Martini, "The couch, the sofa, and everything in between: Discussion on the use case scenarios for light field video streaming services," in *International Young Researcher Summit on Quality of Experience in Emerging Multimedia Services (QEEMS)*, 2017.
- [53] P. A. Kara, P. T. Kovács, M. G. Martini, A. Barsi, K. Lackner, and T. Balogh, "From a Different Point of View: How the Field of View of Light Field Displays affects the Willingness to Pay and to Use," in *2016 Eighth International Workshop on Quality of Multimedia Experience (QoMEX)*, Lisbon, 2016.
- [54] ITU-T Rec., "P.910: Subjective video quality assessment methods for multimedia applications," 2008.
- [55] A. Cserkaszky, P. A. Kara, A. Barsi, and M. G. Martini, "Towards display-independent light-field formats," in *2017 International Conference on 3D Immersion (IC3D)*. IEEE, 2017.
- [56] S. Van Kester, T. Xiao, R. E. Kooij, O. Ahmed, and K. Brunnstrom, "Estimating the impact of single and multiple freezes on video quality," *Proceedings of SPIE*, 2011, vol. 7865, 2011.
- [57] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching," *IEEE Transactions on Broadcasting*, vol. 59, no. 1, pp. 47–61, 2013.
- [58] R. K. Mok, E. W. Chan, and R. K. Chang, "Measuring the quality of experience of HTTP video streaming," in *2011 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2011, pp. 485–492.
- [59] P. A. Kara, W. Robitza, M. G. Martini, C. T. Hewage, and F. M. Felisberti, "Getting used to or growing annoyed: How perception thresholds and acceptance of frame freezing vary over time in 3D video streaming," in *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2016, pp. 1–6.
- [60] P. T. Kovács, A. Fekete, K. Lackner, V. Adhikarla, A. Zare, and T. Balogh, "Big buck bunny light-field test sequences," in *Proc. 112th Moving Picture Experts Group Meeting*, 2015.
- [61] Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang, "A quality-of-experience index for streaming video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 154–166, 2017.
- [62] X. Sang, F. C. Fan, C. Jiang, S. Choi, W. Dou, C. Yu, and D. Xu, "Demonstration of a large-size real-time full-color three-dimensional display," *Optics letters*, vol. 34, no. 24, pp. 3803–3805, 2009.
- [63] J.-H. Lee, J. Park, D. Nam, S. Y. Choi, D.-S. Park, and C. Y. Kim, "Optimal projector configuration design for 300-Mpixel multi-projection 3D display," *Optics express*, vol. 21, no. 22, pp. 26 820–26 835, 2013.
- [64] N. Inoue, S. Iwasawa, and M. Okui, "Public Viewing of 200-Inch Glasses-Free 3D Display System," *New Breeze*, vol. 4, p. 10, 2014.
- [65] M. Hirsch, G. Wetzstein, and R. Raskar, "A compressive light field projection system," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, p. 58, 2014.
- [66] T. Balogh, Z. Nagy, P. T. Kovács, and V. K. Adhikarla, "Natural 3D content on glasses-free light-field 3D cinema," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013, pp. 86 480F–86 480F.



Peter A. Kara received his M.Sc. degree in Computer Engineering from the Budapest University of Technology and Economics in Hungary in 2013. He participated in the CONCERTO project of the European Union's 7th Framework Programme, and he is currently a research associate at the Wireless Multimedia Networking Research Group of Kingston University and a fellow of the H2020 QoE-Net project of the EU. His research interests include multimedia streaming, quality assessment of services, light fields and cognitive bias.



Aron Cserkaszky received his M.Sc. degree in physics from the Budapest University of Technology and Economics in Hungary in 2011. He is currently a researcher at Holografika Ltd. in the area of full parallax imaging. He previously worked on GPU photon simulations and image reconstructions of PET and SPECT systems.



Maria G. Martini is (full) Professor in the Faculty of Science, Engineering and Computing in Kingston University, London, where she also leads the Wireless Multimedia Networking Research Group and she is Course Director for the MSc in "Networking and Data Communications". She received the Laurea in electronic engineering (summa cum laude) from the University of Perugia (Italy) in 1998 and the Ph.D. in Electronics and Computer Science from the University of Bologna (Italy) in 2002. She is a Fellow of The Higher Education Academy (HEA).

She has led the KU team in a number of national and international research projects, funded by the European Commission (e.g., OPTIMIX, CONCERTO, QoE-NET, Qualinet), UK research councils (EPSRC, British Council, Royal Society), UK Technology Strategy Board / InnovateUK, and international industries. Her research interests include wireless multimedia networks, cross-layer design, joint source and channel coding, 2D/3D error resilient video, 2D/3D video quality assessment, and medical applications. She is the author of approximately 150 international scientific articles and book chapters, and the inventor of several patents on wireless video.



Attila Barsi received the M.Sc. degree in Informatics Engineering from Budapest University of Technology and Economics in 2004. From 2004 to 2007, he was attending the Ph.D. program of Budapest University of Technology and Economics, researching real-time graphics. Since 2006, he is in employment at the light field display manufacturer company, Holografika Ltd., where he is researching real-time light field capture and rendering. Since 2009, he is employed as the lead software developer for the company. He was working on several EU

research projects and is currently participating as a supervisor in the QoE-NET and ETN-FPI European training networks. His research interest includes real-time rendering, light fields, multi-camera systems and video streaming on high speed networks.



László Bokor graduated in 2004 with M.Sc. degree in computer engineering from the Budapest University of Technology and Economics (BME) at the Department of Telecommunications. In 2006 he got an M.Sc.+ degree in bank informatics from the same university's Faculty of Economic and Social Sciences. After years of research work in multiple EU funded and national R&D projects he received his Ph.D. degree from the BME Doctoral School of Informatics. Currently he is with the Department of Networked Systems and Services as assistant professor and leads the Commsignia - BME HIT Automotive Communications Research Group. He is a member of the IEEE and HTE, member of the Hungarian Standards Institution's Technical Committee for Intelligent Transport Systems (MSZT/MB 911), and the BME's Multimedia Networks and Services Laboratory (MEDIANETS) where he participates in researches of wireless communications and works on projects related to advanced mobile technologies. His research interests include IPv6 mobility, SDN/NFV-based mobile networks, mobile broadband architectures, network performance analyzing, network simulation, heterogeneous future mobile networks, mobile healthcare infrastructures and V2X communication in cooperative intelligent transportation systems. He received the UNKP-16-4-I. Postdoctoral Fellowship (2016) from the New National Excellence Program of the Ministry of Human Capacities of Hungary.



Tibor Balogh CEO and Founder of Holografika, has extensive experience in the field of holography and optical engineering. He graduated as an electric engineer from the Technical University of Budapest and worked for SZKI (one of the major software houses in Hungary). Later, he was an assistant professor at the Eotvos Lorand University. Today he is responsible for the overall management of his company, supervises the work of the research and development team and forms the overall business strategy. He was awarded the Joseph Petzval medal, the Kalmar prize and the Dennis Gabor Award for his work, was World Technology Award finalist in 2006. He has several patents and publications and actively follows the developments of 3D display technologies around the world.