

Archivo Digital UPM houses in digital format the academic and scientific documentation (theses, pfc, articles, etc.) generated at the institution and makes it accessible through the Internet, within the framework of the Budapest Open Access Initiative and the Berlin Declaration, of which the Universidad Politécnica de Madrid is a signatory.

El **Archivo Digital UPM** alberga en formato digital la documentación académica y científica (tesis, pfc, artículos, etc..) generada en la institución y la hace accesible a través de Internet, en el marco de la Iniciativa por el Acceso Abierto de Budapest y la Declaración de Berlín, de la que es signataria la Universidad Politécnica de Madrid.

ACCEPTED VERSION

► **To cite this version:**

J. Cubelos, P. Carballeira, J. Gutiérrez and N. García, "QoE Analysis of Dense Multiview Video With Head-Mounted Devices," in IEEE Transactions on Multimedia, vol. 22, no. 1, pp. 69-81, Jan. 2020, doi: 10.1109/TMM.2019.2924575

© 2022 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.

QoE analysis of Dense Multiview Video with Head-Mounted Devices

Javier Cubelos, Pablo Carballeira, Jesús Gutiérrez, and Narciso García

Abstract—This paper presents a system and methodology for the analysis of Quality of Experience (QoE) factors for dense multiview (MV) video using a Head-Mounted Device (HMD). An MV-HMD player has been designed and implemented to immerse the users in a virtual environment, where they are placed in front of a virtual lightfield display that shows a different viewpoint depending on the position of their head. The paper describes a methodology for the analysis of the subjective perception of the transition among views (motion parallax), which is specific to the visualization of multiview content. While previous works simulated the user movement by predefined view paths or used complex devices to track them, this system allows the observer to move freely, varying the perspective of the scene, while easily tracking the observer's position. This work is, up to our knowledge, the first providing a complete framework for the assessment of this subjective factor using an HMD. The subjective results obtained using this framework are used to i) assess the influence of the user movement, display settings, and content characteristics in the perception of smoothness in the view transition, and ii) analyze the performance and limitations of a prediction model for subjective smoothness scores.

Index Terms—Multiview video, lightfield, Head-Mounted Device, Quality of Experience, subjective evaluation, smoothness, motion parallax.

I. INTRODUCTION

AN increase in the production and delivery of three-dimensional and immersive visual content has taken place in the past years, especially notable nowadays due to the broad popularization of virtual reality. This interest led to the arrival of stereoscopic and auto-stereoscopic displays to the consumer market. However, these displays generally entail a deficient immersive experience due to the absent or limited motion parallax cues [1], [2]. To solve these limitations, and to provide the user with satisfactory experiences in windowed-6DoF applications, new 3D display technologies such as holographic, and lightfield displays are currently under development [3].

This work has been partially supported by the Ministerio de Ciencia, Innovación y Universidades (AEI/FEDER) of the Spanish Government under project TEC2016-75981 (IVME) and by the Spanish Administration agency CDTI under project IDI-20180015 (VINEDO). The work of J. Gutiérrez has been supported by the People Programme (Marie Curie Actions) of the EU's FP7/2007-2013 (REA grant agreement N. PCOFUND-GA-2013-609102), through the PRESTIGE Programme (by Campus France).

J. Cubelos and N. García are with the Grupo de Tratamiento de Imágenes, Information Processing and Telecommunications Center and ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid 28040, Spain (e-mail: jco@gti.ssr.upm.es; narciso@gti.ssr.upm.es).

P. Carballeira was with the Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid. He is now with the Video Processing and Understanding Lab, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid 28049, Spain (e-mail: pablo.carballeira@uam.es).

J. Gutiérrez is with the Équipe Image, Perception et Interaction, Laboratoire des Sciences du Numérique de Nantes, Université de Nantes, Nantes 44300, France (e-mail: jesus.gutierrez@univ-nantes.fr).

The broad term lightfield [4] refers to a variety of three-dimensional displays that include auto-stereoscopic or near-eye displays. Lightfield displays include auto-stereoscopic displays that distribute a dense multiview set over the viewing field. Such displays (also called super-multiview displays) satisfy the so-called super-multiview condition [5] and offer wide motion parallax to the observers, that see different views of the scene as they move their heads (view transition). Such displays are currently the most promising solution for 3D glasses-free visualization [6].

Quality of Experience (QoE) analysis of 2D image and video content [7] is traditionally oriented to the study of the image quality [8] and has been widely studied. However, the assessment of QoE for immersive video is more challenging due to the complex characteristics of the interaction of users with the immersive systems [9]. In the QoE analysis of lightfield displays, not only the image quality should be considered [10], but also other factors related to the perception of this view transition are relevant [11]. Understanding how subjects perceive these transitions could improve the design of new lightfield displays, or multiview capture systems, e.g., defining the number of views needed for a natural perception of motion parallax.

Some works [12], [13], [14], have already explored these QoE factors. Nevertheless, these subjective studies were usually conducted in unrealistic environments limited by either: i) the simulation of the user's movement, ii) the use of unwieldy devices to track their behavior, and/or iii) the use of unnatural synthetic content for the subjective tests. Thus, further research is still needed to fully understand the perception of dense multiview content.

Understanding the influence of the subject's movement in the perception of these view transitions is key to correctly model the subjective perception of immersive video. The *MultiView Perceptual Disparity Model* (MVPDM) [14] captures the perceptual cues related to the view transition in multiview video. Nevertheless, due to the low availability of lightfield displays [1], the MVPDM was validated with subjective tests that simulated the head movement by a predefined view-sweeping path in a stereoscopic display.

This work focuses on the design and implementation of a framework for the analysis of the perception of view transition for dense multiview content. A Head-Mounted Device (HMD) has been used to simulate a realistic scenario where the users are placed in front of a lightfield display. Users can move freely, varying their perspective of the scene. This framework allows analyzing the users' perception of these view transitions, relating their subjective votes with their movement.

First, an MV-HMD player has been designed and imple-

mented, which uses an HMD to immerse the users in a virtual environment, placing them in front of a lightfield display, and tracking their movements. Then, we present a methodology for the subjective assessment of dense multiview content, in which the MV-HMD player is used to evaluate the subjective perception of smoothness, i.e., fluidity in the view transition, and how it is affected by the user's movement, display settings, and content. The statistical analysis of the subjective results obtained with this method lead to the conclusion that the smoothness perception is influenced by i) the density of the view array, ii) the user's speed while traveling the views, and iii) the depth of the objects of the scene. Finally, the subjective results have been correlated with the predictions of the MVPDM, designed to predict subjective factors such as the smoothness and the speed comfort in the view transition [14]. This analysis shows that the MVPDM presents certain limitations in the prediction of smoothness scores, indicating that there is room for the development of more precise parameterizations.

The main contributions of this paper are:

- 1) This work is, up to our knowledge, the first providing a complete scenario for QoE assessment of the perception of view transition in dense multiview content using HMDs. This scenario might be used for future assessment, or considered in the elaboration of new lightfield/VR QoE recommendations.
- 2) It provides a proposal for the analysis and characterization of the users' movement while visualizing multiview content, and evaluates its correlation with the subjective smoothness scores.
- 3) The MV-HMD player (publicly available) and methodology proposed in this paper can be used to collect subjective data to develop improved parameterizations that can accurately predict perceptual factors in the view transition.

The rest of the paper is structured as follows: Section II describes previous works on QoE analysis for dense multiview content. Section III introduces the main characteristics of the MV-HMD player. Section IV covers the details of the subjective assessment tests, presenting the results in Section V, where a protocol for the analysis of subjective smoothness results is proposed. Section VI details the correlation of MVPDM's parameterization with the smoothness scores. Finally, Section VII provides the main conclusions of the work.

II. QOE ANALYSIS FOR DENSE MULTIVIEW CONTENT

Emerging immersive technologies entail the revision of QoE factors with respect to previous media technologies (e.g., 2D and stereoscopic 3D content). In this sense, research and standardization efforts are done within international organizations and experts groups (e.g., ITU, MPEG, VQEG, etc.) [15] on the formal definition of QoE factors related to the users' experiences when using immersive technologies (e.g., VR, AR, lightfield, etc.), as well as ways to appropriately evaluate them and their inter-relationships. These efforts shall result on reliable assessment methodologies and a QoE model for immersive media, as it happened in the case of stereo 3D

content [16], [17]. This established QoE model for stereo 3D can be considered a base QoE model for immersive visual media, as it identifies three basic perceptual dimensions for immersive video, i.e., *picture quality*, *depth quality* and *visual discomfort*, and two higher-level perceptual dimensions, i.e., *naturalness* and *sense of presence*. Stereo 3D failed to bring immersive experiences to the consumer public since (among other causes) depth quality was only based on monocular cues (as 2D content) and stereopsis (disparity between the left and right images), while other crucial 3D cues were not provided. For a complete immersive experience, depth perception should be provided by means of stereopsis, motion parallax (i.e., seeing different perspectives of the scene as the user moves the head) and focus changes [18]. Multiview and lightfield (dense multiview) technologies are able to provide an improved immersive experience, adding motion parallax to the user perception, as well as enabling focus adaptation in the case of lightfield systems.

Several technical factors affect the perception of that motion parallax and the transition among views, such as the camera configuration, the density of the arrangement (or angular resolution), and the generation of synthesized views (viewpoint interpolation between two real viewpoints) [11]. These technical factors have a direct influence on perceptual factors such as the perception of the smoothness and the speed comfort when the user changes the viewpoint [14].

Here, we review works that tackle the perception of view transition, heeding their differences in i) content type (realistic or synthetic), ii) content parameterization (physical or based on disparity), iii) the characteristics of the user movement (free or predefined), and iv) whether the user behavior is considered in the analysis of subjective scores. Then, we discuss the use or simulation of lightfield displays in subjective tests, and how they can be used to solve limitations of these works.

A. View transition in QoE analysis for dense multiview content

The minimum view density required to provide smoothness perception in the view transition was addressed in some works [12], [13]. The work in [12] focused on the preferred camera to head movement ratio, called *gain of motion parallax*. Users visualized a set of emitting diodes (synthetic content) through a window that responded to the user's movement with a mechanical motion parallax tool. The authors show that the preferred viewing geometry is not identical to the shooting one, and a switching distance shorter than the camera baseline is preferred. The authors of [13] analyzed how the smoothness perception varies with the switching distance, using a stereoscopic display and simulating the user's movement with view sweeping. Their study concluded that the perceived smoothness improved with shorter switching distances. However, they did not consider the influence of the view-sweeping speed.

Ribeiro et al. [11] present a multiview player for stereoscopic environments, with motion parallax and zoom in/out capabilities. They analyzed visual comfort, sense of immersion, smoothness of the parallax-induced motion, and overall 3D experience on realistic content. Their results, in accordance to

others' [12], [13], suggest that users are very sensitive to small view-discontinuities, achieving perfect smoothness perception only for angular resolutions over 1.49 views/deg. The results also suggested that the required view density may be reduced interpolating intermediate views. However, user movement was simulated with predefined trajectories in a stereoscopic display, and the view-sweeping speed was not considered in the analysis.

Carballeira et al. [19] analyzed the effects of *View Density* (VD), *View-Sweep Speed* (VSS), and *View Range* (VR) in QoE factors related to multiview video such as smoothness and speed comfort in the perception of view transition, and depth perception, simulating motion parallax with predefined view sweep trajectories in a stereoscopic and an auto-stereoscopic display. These tests were performed for realistic content, but limited in variety, especially in the case of moving sequences, and range of VSS and VD values.

Following this work, a model to capture the subjective perception of the view transition in multiview video was presented in [14]. The MVPDM proposed a parameterization of dense multiview content based on image cues, and not only on the physical setting of the cameras. Specifically, MVPDM is based on the *perceptual disparity* (d_{perc}), that characterizes the disparity histogram between camera pairs, jointly capturing the influence of camera settings (camera density and intrinsic parameters) and the location of the objects in the scene (depth). d_{perc} is computed by a weighted average of the disparity histogram using two factors (α and β) that weight the contribution of disparity values produced by objects in front and behind the convergence plane. MVPDM correlated better with the perception of speed comfort or smoothness in the view transition than parameterizations based only on the physical settings, i.e., camera baseline or angular resolution. However, as in other previous works, the subjective assessment of the MVPDM was performed simulating the user movement through predefined view trajectories in a stereoscopic display.

B. Simulation of lightfield displays in subjective tests

Several of the reviewed works share a limitation: subjective tests have not been performed in real lightfield displays, but user movement has been simulated by predefined view trajectories. This responds to: i) the low availability of commercial lightfield displays, and ii) predefined trajectories allow seamless tracking of the viewpoint observed by the viewer. Nevertheless, in a realistic scenario, the user would freely move in front of the lightfield display.

Only a reduced number of works have used real lightfield display prototypes in the subjective tests. In [20], several lightfield displays were used to test the image quality of compressed dense multiview video. A lightfield cinema system was used to assess the influence of zooming level and number of views in the overall perception of quality in lightfield video [21], [22]. Lightfield displays were also used to validate new objective metrics for dense multiview content, such as the one introduced in [23].

Other works have used head-tracking and HMDs to simulate a lightfield display and/or track the user position. A multiview

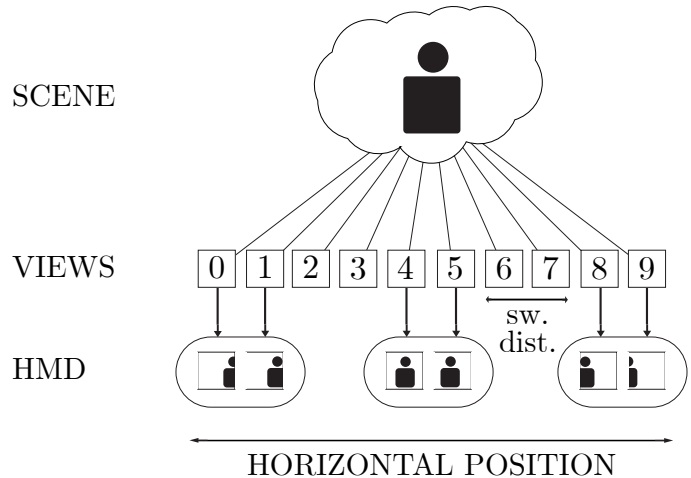


Fig. 1: Basics of the MV-HMD player. Depending on the user's head position, a different pair of views is displayed in the HMD. The switching distance represents the distance between consecutive views.

video streaming system [24] selected transmitted views and bitrate allocation based on the prediction of the head position, which is tested using a single-user stereoscopic display with head tracking [25]. A recent work [10] simulated a lightfield display with a stereoscopic monitor and a LED headlamp tracker. This system was used to collect subjective image quality data on a lightfield dataset with different types of impairments, which is then correlated with objective image quality metrics. A WebGL solution to display narrow baseline lightfield images in a web-browser is provided in [26], focusing on crowd sourcing applications. The work in [27] provides a method to add motion parallax to 3DoF VR videos, and their subjective analysis takes into account the head position which is tracked using an HMD.

We follow a similar approach to address the limitations on the analysis of view transitions for dense multiview videos. An HMD is used to simulate a realistic environment, in which the user is located in front of a virtual lightfield display. Thus, the user can freely vary position and movement speed, which are tracked and used in a posterior analysis of the subjective votes.

III. MV-HMD PLAYER

A. Overview

The MV-HMD player implemented in this work simulates a virtual lightfield display, located in a fixed position within a 3D scene, where the user is immersed. Depending on the user's head position, the system selects a stereo pair of views to display in the virtual display, as shown in Fig. 1. As shown in the figure, the switching distance represents the distance that the user has to move the head to change the viewpoint.

Fig. 2 depicts an example of what the users visualize and how the system reacts to their movement, showing the stereo pair corresponding to the central position with respect to the screen, and the left-most stereo pair for a user reaching the left limit of the array.

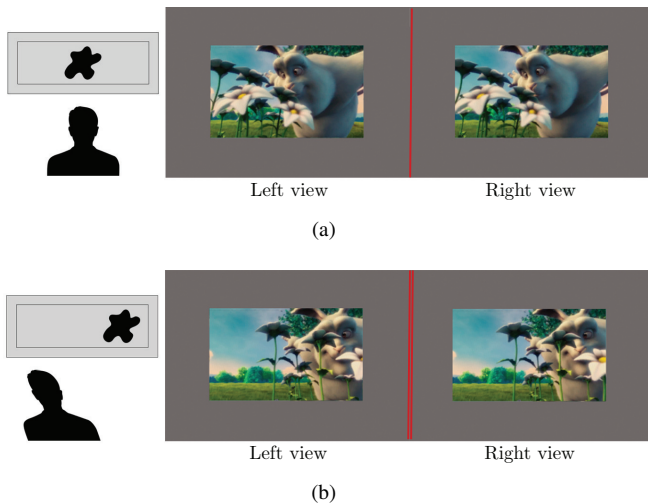


Fig. 2: Example of the stereo viewports displayed in the HMD for different user's positions. (a) Central stereo pair displayed for a user located at the center of the array. (b) Left-most stereo pair displayed for a user located next to the left limit.

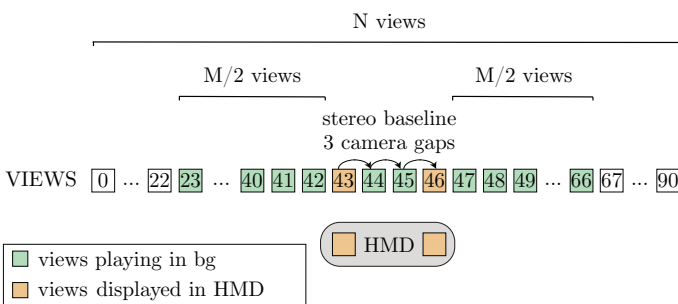


Fig. 3: M views are played in the background around the views displayed to the user. Example for a central user position.

The source code of the MV-HMD player is publicly available in <https://github.com/jcocube/mvplayer>.

B. Implementation

The system is implemented in the Unity 3D game development platform [28]. We use an HMD with a positional tracker to extract the user position and adapt the viewpoint: Oculus DK2 (Development Kit 2) [29]. The seamless integration of Unity 3D platform with multiple HMDs allows a simple migration to other devices.

A key requirement for the system is the fluidity in the view switching, i.e., the synchronization of the user's movement and the corresponding view switching. To fulfill this requirement, the system continuously plays, in a synchronized manner, a subset of M views around the user's position (cf. Fig. 3), while rendering is only enabled for the two views of the stereo pair seen at the user's position. In each Unity 3D rendering refresh, the background views are updated to match the new user's position. Unity 3D is configured with a constant rendering refresh rate of 50 fps, which guarantees a maximum delay in the adaptation of the viewpoint of 20 ms.

The head position and orientation are tracked, to enable the analysis of the user's movement data and its relationship with the subjective scores (as explained in Sections V, and VI).

C. System configurable parameters

The MV-HMD player allows the configuration of the following parameters:

- *Content settings*: number of views (N) and spatial resolution of the multiview array.
- *Stereo baseline*: i.e., interocular distance, distance between left and right views of the stereo pair (measured in camera gaps).
- *Number of background views M* : number of views that are played and synchronized in the background.
- *Switching distance*: horizontal distance between two consecutive views, i.e., distance that the user needs to move the head to switch to the adjacent view.
- *Display distance*: distance of the user to the virtual display.

IV. SUBJECTIVE TEST SETUP

Using the MV-HMD player described in Section III, we performed a subjective assessment test to study the user perception of smoothness in the view transition and how it relates to the user's movement and content settings.

In this assessment, the switching distance was explicitly varied, creating sparser/denser multiview arrays at display time (varying the gain of motion parallax [12]). The speed in the viewpoint transition was also implicitly varied among tests, as each user was free to use their own preferred movement, which varied for different contents and switching distances (see Section V-B). We evaluated not only multiview video sequences, but also multiview static frames, to disaggregate the effect of content motion in the smoothness score.

To define and adjust the details of the subjective evaluation methodology in this scenario, which is new up to our knowledge, we first performed pilot tests with a limited number of users. This section describes the final setting of the subjective assessment methodology.

A. Equipment

A high-performance PC with a Nvidia GeForce GTX 970 graphics card, 32 GB RAM, Intel Core i7-4790 CPU was used to execute the MV-HMD player. Additionally, an Oculus Rift DK2 headset and its corresponding tracker were used.

B. Subjective Test Material

For all the subjective tests, we used the four lightfield sequences selected by the MPEG-I subgroup for the dense 1D multiview category [31]. All sequences cover a wide viewing range with high density. Examples of the central view of the four sequences are shown in Fig. 4. As shown in the figure, the test set comprises natural and realistic CG content. More details on the camera arrangements and sequence content can be found in [30], [32], and [33].

TABLE I: Characteristics of multiview sequences and static frames, with MVPDM parameterization values. *Ch2* is a denser version (twice) of the original *Ch* sequence, created adding one synthesized view between each pair of original views [30]. *BA* has been modified removing uniformly two-thirds of the total number of views (from 91 to 31 views), creating a sparser array to avoid saturation of subjective scores in the higher range (conclusion extracted in the pilot studies). The d_{perc} values are computed using the α , and β values proposed in [14]; and correspond to average values for video sequences. The disparity values have been normalized with the image width to compare contents of different resolutions.

<i>Sequence</i>	<i>Id</i>	<i>Camera arrangement</i>	<i>Resolution</i>	<i>Frame rate</i>	<i>Cams</i>	d_{near}	d_{perc}
Pantomime	Pa	Linear convergent	640x480	30 fps	80	1.56%	0.70%
Champagne	Ch	Linear convergent	640x480	30 fps	80	2.20%	1.09%
Champagne 2	Ch2	Linear convergent	640x480	30 fps	80	1.10%	0.54%
Flowers Linear	FL	Linear parallel	640x384	24 fps	91	2.43%	0.58%
Flowers Arc	FA	Arc	640x384	24 fps	91	1.89%	0.23%
Butterfly Linear	BL	Linear parallel	640x384	24 fps	91	1.31%	0.30%
Butterfly Arc	BA	Arc	640x384	24 fps	31	0.71%	0.17%

<i>Static Frame</i>	<i>Id</i>	<i>Camera arrangement</i>	<i>Resolution</i>	<i>Frame number</i>	<i>Cams</i>	d_{near}	d_{perc}
Pantomime hd	PaH	Linear convergent	640x480	#121	80	1.17%	0.92%
Pantomime md	PaM	Linear convergent	640x480	#31	80	0.55%	0.22%
Pantomime ld	PaL	Linear convergent	640x480	#1	80	0.51%	0.08%
Champagne hd	ChH	Linear convergent	640x480	#481	80	3.75%	2.50%
Champagne md	ChM	Linear convergent	640x480	#391	80	2.89%	1.68%
Champagne ld	ChL	Linear convergent	640x480	#1	80	2.20%	0.94%
Flowers Lin. hd	FLH	Linear parallel	640x384	#25	91	2.21%	0.63%
Flowers Lin. ld	FLL	Linear parallel	640x384	#1	91	2.22%	0.56%

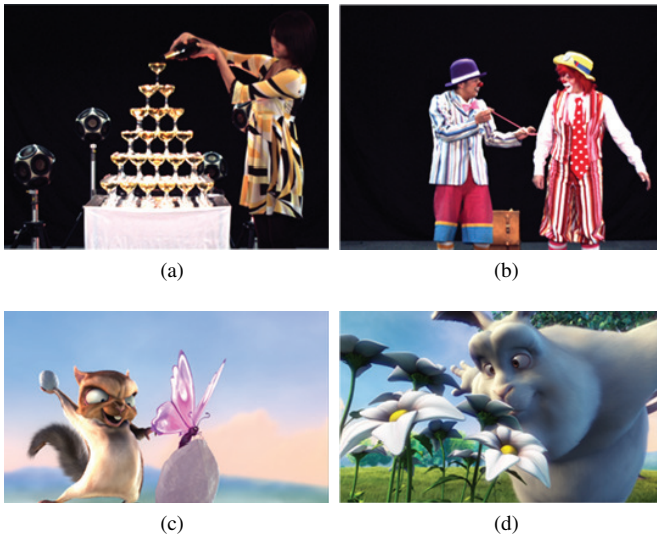


Fig. 4: Samples of the test sequences. (a) Champagne: cam 40. (b) Pantomime: cam 40. (c) Butterfly (linear/arc): cam 45. (d) Flowers (linear/arc): cam 45.

Table I shows the characteristics of the multiview sequences and static frames, with their corresponding IDs. The spatial resolutions have been reduced to one quarter of their original ones. Table I shows, for each content, the disparity of objects at the nearest depth (d_{near}), and d_{perc} , obtained using the MVPDM parameterization [14]. More details of these two parameterizations can be found in Section VI.

The stereo baseline values, measured in camera gaps between left and right views, are the ones used in [14], defined by subjective studies by the MPEG group [34]:

- 1) Linear natural sequences (*Ch*, *Pa*, *Ch2*): 1
- 2) CG sequences: *FL* & *FA* 1, *BL* 2, *BA* 3

C. MV-HMD settings and environment

The users were located in front of a monitor with the positional tracking attached on top of it, located approximately at 1.2-1.4m from the user's head, within the distance range recommended by Oculus for the optimal performance of the tracker. The users were seated on a chair that allowed them to move the head, and eventually their upper body, horizontally to change the viewpoint.

The number of background views in the MV-HMD player was set to 60. For the smallest value of the tested switching distances (1 mm, worst case scenario), the user can move up to 30 mm during each refresh period of the system (20 ms) without leaving the range of background views. Given the 50 fps refresh rate, this corresponds to a speed of 150 cm/s, which is 4 times higher than the maximum speed covered during the tests (see Table IV). Also, the refresh rate is higher than the maximum frame rate of the sequences (30 fps), which guarantees that the frame rate of all clips is correctly perceived.

The background of the 3D scene of the MV-HMD player is set to a mid-gray color as recommended for QoE assessment [35] (as shown in Fig. 2). The virtual lightfield display covers 40° of the horizontal Field Of View (FOV) per eye for all tests. The Oculus HMD provides a horizontal FOV of 60° per eye.

D. Observers

A total of 23 observers (8 female, 15 male) participated in the tests, all of them having normal or corrected vision (glasses or lenses). The ages of the participants were between 19 and 35, with an average age of 27. There were 11 users familiar with HMD technology (47%) and 12 users using an HMD for the first time (53%).

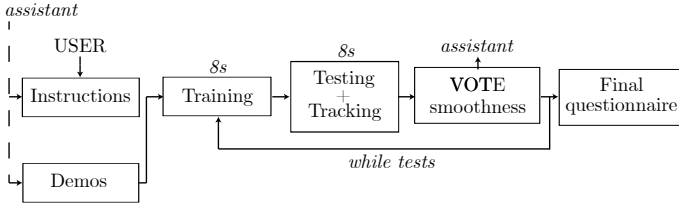


Fig. 5: Subjective assessment methodology.

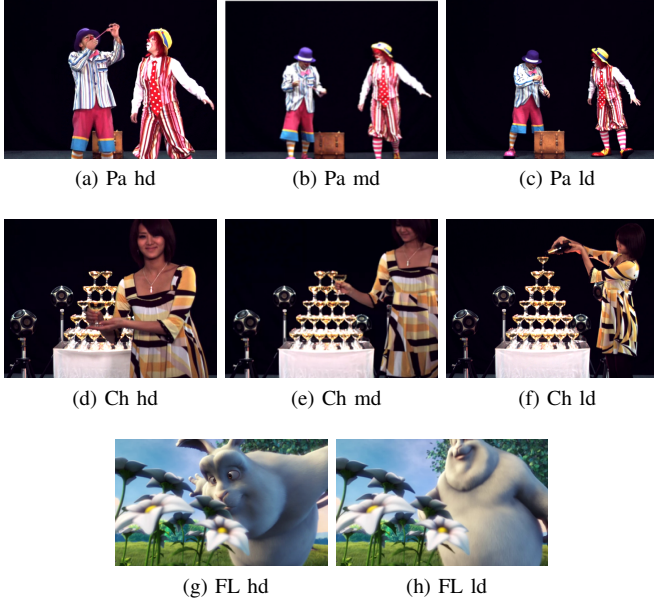


Fig. 6: Experiment 2: Selected multiview static frames.

E. Methodology

The Absolute Category Rating (ACR) [35] methodology was used to evaluate the perception of the smoothness in a five-degree scale. Fig. 5 shows a schematic representation of the different steps comprised in the QoE assessment. Each test clip had an estimated duration of 20 seconds, which is divided in three parts:

- **Training:** First 8 seconds of each test. The user must find the most-comfortable horizontal oscillatory movement for the displayed content, covering a sufficient view range to have satisfactory motion parallax. At the end of the training part, a visual flag appears next to the virtual display, indicating the beginning of the testing part.
- **Testing:** Last 8 seconds of each test. The user must keep the movement of the training phase, maintaining the range of covered views as constant as possible. The user's head position was tracked during these 8 seconds.
- **Vote:** after the training and testing phases, a screen with a *VOTE* message is shown, giving the observer 4 seconds to orally rate the perceived smoothness.

The test sessions started with a training process where some example sequences, different from the ones used in the tests, were shown to the users to explain the test methodology and train the user in the range of view densities shown in the tests. Each test clip was shown once, and different randomizations

TABLE II: Experiment 1: Tested switching distances for multiview sequences.

<i>Sequence</i>	<i>ID</i>	<i>Sw. dist. (mm)</i>			
Pantomime	Pa	2	6	15	30
Champagne	Ch	2	6	12	25
Champagne 2	Ch2	1	3	6	12
Flowers Linear	FL	2	3	6	12
Flowers Arc	FA	2	4	8	12
Butterfly Linear	BL	2	3	6	10
Butterfly Arc	BA	6	10	20	30

TABLE III: Experiment 2: Tested switching distances for multiview static frames.

<i>Static Frame</i>	<i>ID</i>	<i>Sw. dist. (mm)</i>			
Pantomime hd	PaH	2	6	15	30
Pantomime md	PaM	2	6	15	30
Pantomime ld	PaL	2	6	15	30
Champagne hd	ChH	2	6	12	25
Champagne md	ChM	2	6	12	25
Champagne ld	ChL	2	6	12	25
Flowers Linear hd	FLH	2	3	6	12
Flowers Linear ld	FLL	2	3	6	12

were used to reduce contextual effects (with the condition of not showing the same source content consecutively [36], [37]). A set of 7 breaks of 30 seconds were proposed to the user. Finally, a small questionnaire was given to the user after the test session, asking a few simple questions regarding the testing protocol and experienced visual comfort. The average duration of the whole testing session, composed by 60 test clips, was of 28 minutes.

The subjective assessment included two experiments:

1) Experiment 1 - Smoothness of Multiview Video Content:

In this experiment, we evaluated the smoothness perception of the seven dense multiview video sequences (Table I) varying the motion parallax gain, through four different switching distances. Table II shows the tested switching distances for each sequence. There were a total of 28 tests clips, with an estimated total test duration of 11-13 min.

2) Experiment 2 - Smoothness of Multiview Static Content:

In this experiment, we evaluated the smoothness perception of multiview static frames. In video sequences, objects move, and may vary their depth (distance to the camera), and, therefore, the disparity they produce between neighboring views. The MVPDM [14] showed that the smoothness perception is related to this disparity. Thus, this experiment targets to disaggregate the effect of a disparity that varies along time in the smoothness score, testing different individual frames with distinct disparity values.

Individual frames were extracted from sequences with enough disparity variation along time (*Pa*, *Ch* and *FL*), selecting a subset of frames with distinct disparity values (see d_{near} in Table I). Three frames were selected for *Pa* and *Ch* and two for *FL*. Fig. 6 shows the selected frames, marked as high disparity (*hd*), intermediate disparity (*md*) and low disparity (*ld*). Table III shows the switching distances tested for each frame. There were a total of 32 tests clips, with an estimated total test duration of 15-17 min.

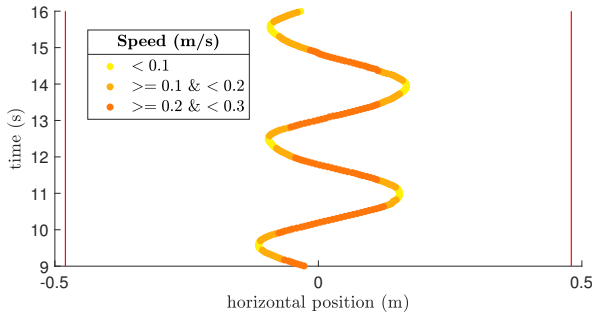


Fig. 7: Example of horizontal positional tracking along time. The speed is represented in a color scale. The vertical red lines represent the location of the left- and right-most views.

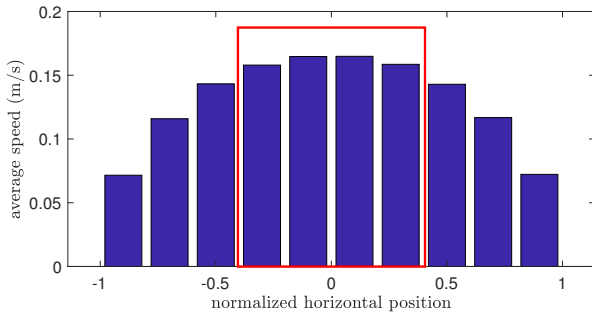


Fig. 8: Average speed vs normalized horizontal position - Mean for all tests and subjects. The red box comprises the central range of speeds averaged for UMS computation.

V. SUBJECTIVE EXPERIMENTAL RESULTS AND DISCUSSION

A. Subject Screening

A screening of the smoothness scores was carried out, considering the correlation between the scores of each subject and the average scores, as recommended in [36], and [37]. After this screening, one observer was rejected for the Experiment 1, and none for the Experiment 2.

B. Characterization of the User's Movement Speed

To analyze how the subject's movement influenced the smoothness score, it is necessary to characterize the subject's speed during each test. However, the user's movement is not completely constant and its speed typically varies with the horizontal position. Fig. 7 depicts an example of a typical user movement along the x-axis, with the speed in a color scale. As it can be seen, the users move at the highest speed at the central part of the oscillation, decreasing when they get closer to their movement limits. Also, each subject covers a different oscillation amplitude for different test clips. The analysis of the movement data, shows that 90% of the users moved with an amplitude lower than 28 cm. These results are in accordance with the ones obtained in [38], where the authors stated that users did not surpass the 30 cm of amplitude, even when the range was not limited.

To characterize the speed of each user (User Movement Speed, UMS), we averaged the speed values covered in the central range of the subject's movement. The user position is normalized by the covered range in the x-axis, and the UMS is defined as the mean speed in the $[-0.4, 0.4]$ interval of normalized positions. Fig. 8 shows the mean speed vs the normalized horizontal position, for all users and tests.

C. Analysis of smoothness MOS using (UMS, sw. dist.) values

Given the definition of UMS, each test clip is defined by a switching distance, a UMS, and a smoothness score. Table IV shows, for all test clips, the 90% upper-bound UMS values, i.e., 90% of the users have moved with slower or equal UMS. For a particular content (sequence or frame), the used UMSs are mainly driven by the switching distance, increasing with larger switching distances (sparser arrays). The influence of the content in the UMS can be analyzed with the results of the static frames. For those tests, the values of the switching distances are the same for all the frames of a sequence (ld, md and hd). In general, UMS increases for frames with further objects (lower disparity). For instance, as shown in Table IV, for *Pa*'s frames, which presents the biggest variation in disparity (see d_{perc} values in Table I), the 90% upper-bound UMS value increases (i.e., viewers move faster) as the disparity decreases. This behavior holds for all the tested switching distances.

To facilitate the analysis of smoothness scores, the UMS values were discretized in a set of 15 bins covering the whole UMS range. To reject outlier scores, smoothness scores with only one subject for a UMS bin in a test clip were discarded.

1) *Experiment 1 - Smoothness of Multiview Video Content:* Fig. 9 shows the subjective smoothness scores provided by the observers. They are represented with boxplots for the discretized UMS values and switching distances of all the test clips in Experiment 1. The boxplots are grouped in blocks corresponding to a single UMS bin value, e.g. in Fig. 9 (f), the four boxplots grouped at $UMS = 13.5$ cm/s correspond to the same UMS value. In addition, the figure shows the smoothness Mean Opinion Scores (MOSs) and trend lines for each group of MOSs that correspond to the same switching distance. Generally, these trends show that: i) for a given UMS, shorter switching distances obtain better smoothness scores than larger ones; ii) for a specific switching distance, faster UMSs lead to better smoothness perception. It is also worth noting that a considerable range of the five-grade scale has been covered for all the sequences. Two sequences present atypical results: i) smoothness scores for *Ch* are concentrated in the lowest range. This indicates the lower view density of this sequence compared to the rest; ii) Smoothness scores for *BL* do not show a clear trend with UMS, which may be related to particular characteristics of the content.

Given that the obtained sets of subjective scores are not normally distributed, the analysis over the MOSs and their trends is limited to general insights. In order to analyze in depth the statistical significance of the effects of the switching distance and UMS in the smoothness scores, a non-parametric test was performed over the whole set of raw scores. In

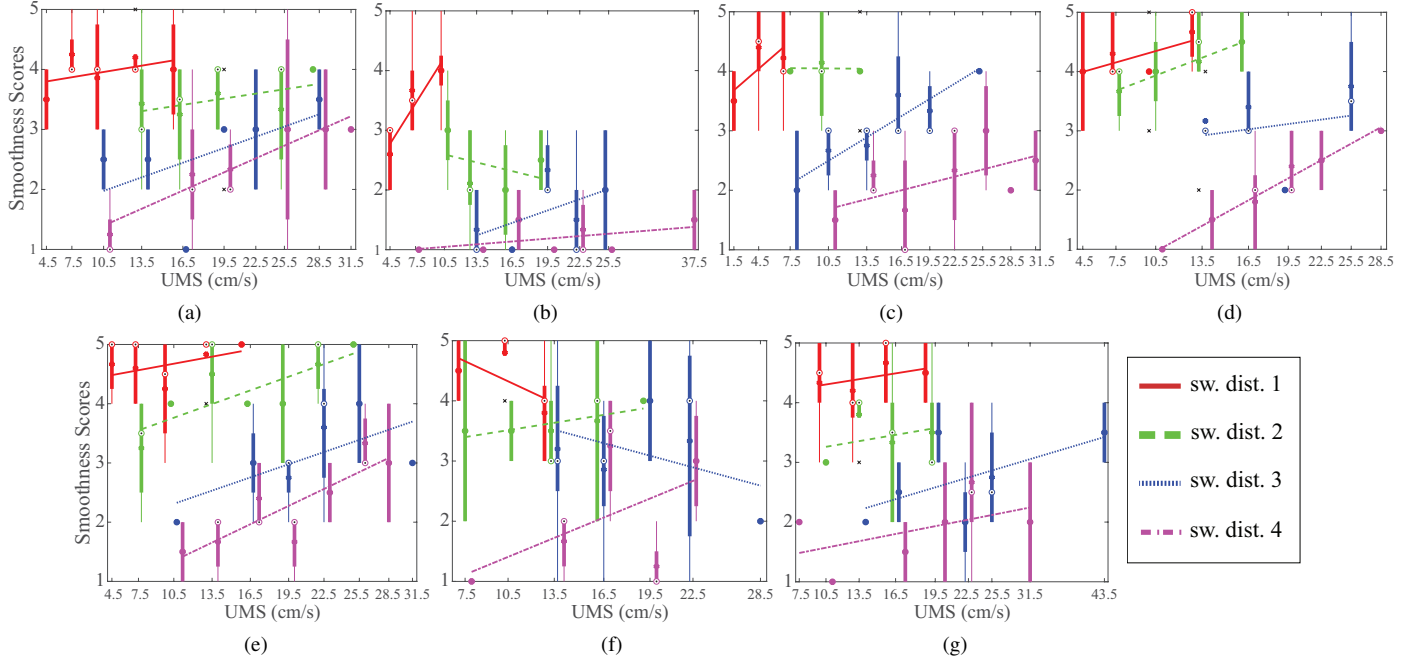


Fig. 9: Boxplots (25-75% percentiles) of smoothness scores for Experiment 1 (varying UMS and switching distance). MOS values are represented by filled dots, medians with black circled dots, and outliers by black crosses. In some cases MOSs occlude the medians. The trend lines group the smoothness MOSs corresponding to the same switching distance. (a) Pa. (b) Ch. (c) Ch2. (d) FL. (e) FA. (f) BL. (g) BA. The switching distances values, labeled from 1 (shortest) to 4 (largest), are shown in Table II.

TABLE IV: 90% upper-bound UMS values (in cm/s) for all test clips, i.e., indicates that 90% of the users have moved with slower or equal UMS than that value for that test clip.

	Sequences							Static Frames							
	<i>Pa</i>	<i>Ch</i>	<i>Ch2</i>	<i>FL</i>	<i>FA</i>	<i>BL</i>	<i>BA</i>	<i>PaH</i>	<i>PaM</i>	<i>PaL</i>	<i>ChH</i>	<i>ChM</i>	<i>ChL</i>	<i>FLH</i>	<i>FLL</i>
<i>sw. dist. 1</i>	15.1	9.6	8.2	13.9	16.9	16.6	21.8	14.3	18.6	23.4	9.9	10.5	11.5	15.9	15.5
<i>sw. dist. 2</i>	26.9	19.2	15.7	26.1	26.3	18.4	26.9	23.6	33.9	32.6	16.4	15.8	18.6	14.8	23.0
<i>sw. dist. 3</i>	33.2	24.6	24.2	26.1	31.8	24.0	44.3	28.0	32.8	40.1	20.0	22.2	24.9	22.2	24.3
<i>sw. dist. 4</i>	40.0	26.6	30.7	28.9	27.7	27.9	31.7	28.8	31.2	40.4	26.8	24.6	27.0	24.2	26.4

particular, a Skillings-Mack test [39] was carried out, which can be considered as a Friedman test (i.e., non-parametric version of a two-way Anova) for unbalanced data (i.e., samples with different size or with missing data) and for samples with reduced size. In the case under analysis, the UMSs were not equally used (i.e., not all observers moved their head with the same speeds), and the smoothness scores were distributed among the different UMSs bins. The results from the Skillings-Mack test showed statistical significance for the effects of both factors: switching distance and UMS (p-values below 0.05 for all source contents). Given the nature of the data (i.e., unbalanced data and reduced samples size), permutation tests were carried out as post-hoc tests to further analyze these influences [40].

First, the influence of the UMS was explored, considering pairs of score sets for different UMSs (same source content and switching distance). In addition to the general trends depicted in Fig. 9 indicating that faster UMSs provide better smoothness, the post-hoc test showed that the effect of UMS is statistically significant for big switching distances. For all source sequences and switching distances 3 and 4, statistically

significant differences exist in most cases when comparing the UMSs that produce the highest and the lowest MOS values. These values typically correspond to the fastest and the slowest UMSs, respectively. Additionally, *FA* shows statistical significance for switching distance 2.

Then, to analyze in detail the effect of the switching distance, a second permutation test compared pairs of score sets for different switching distances (same content and UMS). The results are summarized in Table V, where the ratio of pairs showing statistical significance with respect to the total comparisons are shown. For example, the last cell of *Pa*, shows that there are two UMS values used for both switching distances 1 and 4, and there is statistical significance for one of these two comparisons. Missing data is due to the fact that not always the same UMS was used for two different switching distances (i.e., no overlapping data to compare). The results show: i) there are only a few cases with statistically significant differences between switching distances separated by one step; ii) when the switching distances step is two, more than the half of cases show statistical significance; and iii) almost all cases are significant when comparing switching distance 1 and

TABLE V: Permutation test results for the effect of switching distance in Experiment 1: ratio of pairs showing statistical significance with respect to the total compared pairs between each two switching distances.

SD-Dif	1 step			2 steps		3 steps
	1-2	2-3	3-4	1-3	2-4	1-4
Pa	0/2	0/1	0/1	-	0/3	1/2
Ch	0/1	0/3	1/3	-	1/2	-
Ch2	0/1	1/2	1/2	-	1/1	-
FL	0/3	0/2	1/1	1/1	1/1	-
FA	1/3	0/1	0/2	1/1	1/1	2/2
BL	1/2	0/2	0/3	0/1	1/2	1/1
BA	0/2	1/2	0/2	2/2	1/1	1/1
Avg.	2/14	2/13	3/14	4/5	6/11	5/6
	7/41			10/16		5/6

4. These results prove the effect of the switching distance in the smoothness score, especially when the difference in switching distance is above a certain threshold, which is content dependent (different switching distances were chosen for different sequences, Tables II and Table III).

Experiment 1 uses a subset of sequences in common with the experiments in [14] (*Ch*, *Pa*, *FL*, *FA*, *BL*). Thus, we have compared the smoothness MOSs in [14] (user movement simulated in a stereoscopic display) and the ones obtained here. Smoothness scores were compared for equal values of VSS, measured in views per second (common units for both settings). Smoothness scores in [14] are higher than the ones obtained here (+0.73 points in average). This can be influenced by the fact that, in [14], the user movement was simulated with a uniform speed. Instead, in a realistic user movement, speed is decreased at the limits of the movement (see Fig. 8). As seen in Fig. 9, lower UMS values result in lower smoothness scores, and thus, the perception at the movement limits may influence the subjects to provide a lower smoothness score.

2) *Experiment 2 - Smoothness of Multiview Static Content:* Fig. 10 shows the subjective smoothness scores for Experiment 2 in the same format of Experiment 1. The trends show the same general conclusions for the effect of switching distance than that of Experiment 1. For a particular UMS, shorter switching distances obtain better smoothness scores than larger ones. However, in Experiment 2, a faster UMSs does not always lead to better smoothness perception, as reflected by the trends for *Ch* and *FL* frames for big switching distances (3 and 4). This may indicate that dynamic content is masking some effect in the perception of smoothness. Proceeding similarly to Experiment 1, a Skillings-Mack was performed. The results show statistical significance for the effects of switching distance and UMS in the smoothness scores (p -values below 0.05 for all source frames).

Permutation tests for the UMS factor show that, in all *Pa* frames, significant differences exist when comparing the UMSs that provide the highest and the lowest MOS values for switching distances 3 and 4. Significant differences also exist between the lowest and highest scores for switching distances 1 and 2 in *FLH* and switching distance 2 for *FLL*. For *Ch* frames, the concentration of votes in the lowest range prevents the existence of statistically significant differences among UMS values.

TABLE VI: Permutation test results for the effect of switching distance in Experiment 2: ratio of pairs showing statistical significance with respect to the total compared pairs between each two switching distances.

SD-Dif	1 step			2 steps		3 steps
	1-2	2-3	3-4	1-3	2-4	1-4
PaH	1/1	1/3	2/4	1/1	3/3	1/1
PaM	1/4	1/3	2/2	1/4	3/3	3/3
PaL	0/3	1/2	0/1	0/3	0/2	2/2
ChH	1/2	0/2	0/3	1/2	0/2	1/1
ChM	1/2	0/2	1/4	1/2	1/2	1/1
ChL	1/2	0/4	2/3	1/2	2/3	1/1
FLH	0/1	1/2	1/3	0/1	2/2	1/1
FLL	0/1	2/3	0/2	0/1	2/2	-
Avg.	5/16	6/21	8/22	5/16	13/19	10/10
	19/59			18/35		10/10

Permutation test results for the effect of the switching distance are shown in Table VI (same conditions of Experiment 1). The results support the conclusions extracted for Experiment 1, since there are more significant differences between the compared pairs of sets as the difference between switching distances is bigger (i.e., more statistical differences comparing switching distances 1 and 4 than for smaller steps). However, in contrast to the results in Experiment 1, there are more cases significant differences with smaller steps (e.g., around 30% comparing switching distance 1 and 2 in Experiment 2 against 17% in Experiment 1), which may be explained by the absence of movement in static content. Taking out this factor reduces the dispersion of the vote among participants.

In addition, the results in Experiment 2 allow to analyze the effect of disparity. For a given switching distance and UMS, the smoothness perception is better for frames with lower disparity. For instance, this can be observed in the increasing of smoothness scores from Fig. 10 (a) to (c) corresponding to switching distance 2, which correspond to *Pa hd*, *md* and *ld*, respectively.

D. Conclusion

The analysis of the user behavior reveals that users have moved covering amplitudes below 30 cm, and that the UMS is primarily influenced by the switching distance, increasing for higher distances, and secondarily by scene content, increasing for lower disparities.

From the analysis of the subjective results, it can be concluded that the perception of smoothness for multiview content is influenced by i) the array's density (switching distance), ii) the UMS, and iii) the scene disparity (depth of the objects). The perception of smoothness improves for denser arrays, lower disparities, and faster UMS. Post-hoc analysis of the results have shown statistical significance for the effects of both switching distance and UMS.

Through the comparison of smoothness results in [14] and the ones obtained here, we derive that perception at the movement limits, that typically implies a slower head motion, may influence the smoothness perception, decreasing the score with respect to scores obtained with simulated uniform speed for head motion.

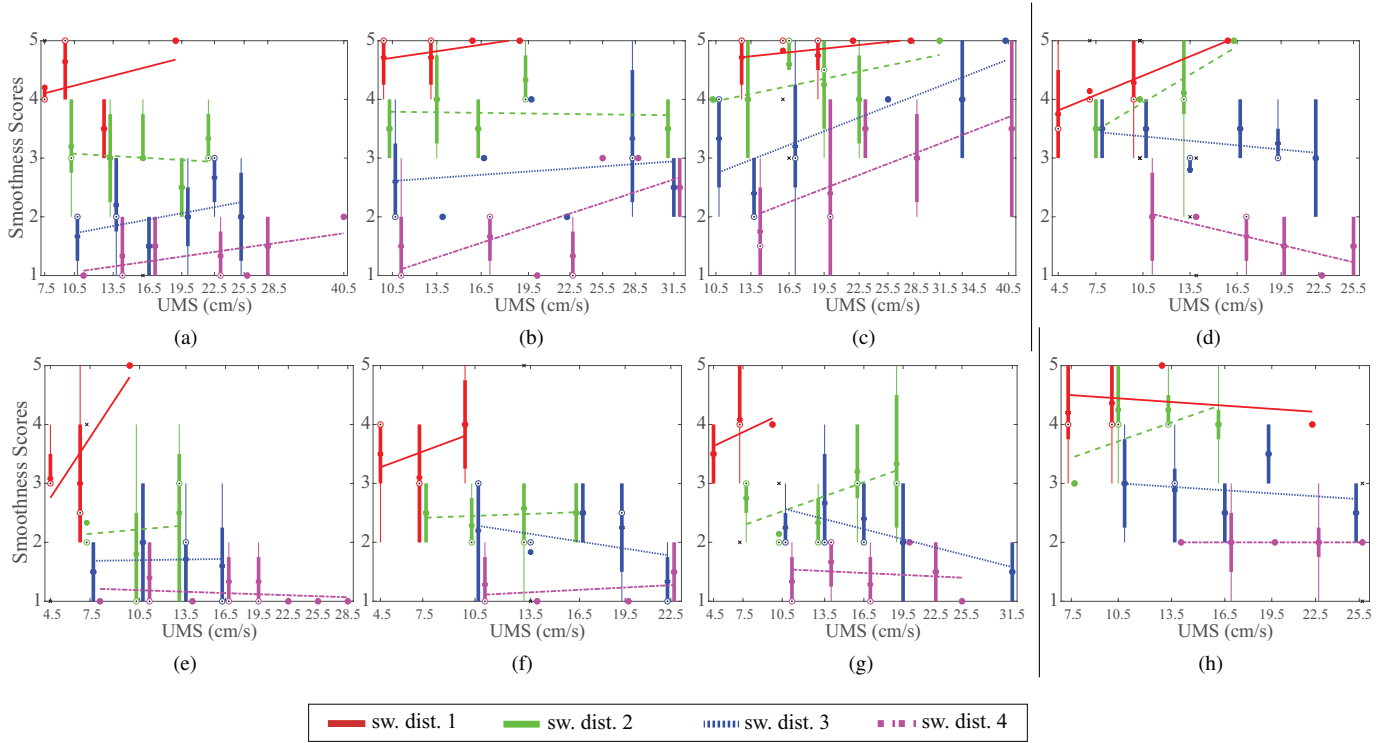


Fig. 10: Boxplots (25-75% percentiles) of smoothness scores for Experiment 2 (varying UMS and switching distance). MOS values are represented by filled dots, medians with black circled dots, and outliers by black crosses. In some cases MOSs occlude the medians. The trend lines group the smoothness MOSs corresponding to the same switching distance. (a) PaH. (b) PaM. (c) PaL. (d) FLH. (e) ChH. (f) ChM. (g) ChL. (h) FLL. The switching distances values, labeled from 1 (shortest) to 4 (largest), are shown in Table III.

VI. CORRELATION OF THE SUBJECTIVE RESULTS WITH THE MVPDM PARAMETERIZATION

We have studied the correlation of the parameterization proposed in MVPDM [14] with the smoothness results (Section V). Additionally, we have compared its performance (based on d_{perc}) with the one based on the disparity of the nearest plane (d_{near}) [14].

The analysis contains two main parts; first, we have verified that the UMSs of the subjects result in a comfortable VSS (using the prediction model for speed comfort [14]). Then, we analyze the correlation of the smoothness scores with the MVPDM parameterization and discuss its limitations.

A. Relation between VSS and UMS values

VSS [14] defines the variation in viewpoint content per time unit, i.e., how fast the viewpoint changes when the user moves the head. VSS depends on UMS (which in [14] was simulated by a predefined view-sweeping trajectory), but also on the switching distance and how the change between two views is parameterized (disparity in the MVPDM). Depending on the basic parameter (d_{perc} or d_{near}) two modes of VSS can be computed, VSS_{near} based on d_{near} , and VSS_{perc} based on d_{perc} :

$$VSS_{near}[\%/s] = \frac{d_{near}[\%] * UMS[cm/s]}{sw_dist[cm]} \quad (1)$$

$$VSS_{perc}[\%/s] = \frac{d_{perc}[\%] * UMS[cm/s]}{sw_dist[cm]} \quad (2)$$

B. Validation of the Speed Comfort

Using the prediction model for speed comfort [14], and the VSS values computed using (2) from subjective test data, we have checked the speed comfort predictions for all test clips. The results show that the predicted values for speed comfort for all tests are within the [4.09, 4.18] range (using a five-degree scale). Since it was requested that users moved at speeds that produced a comfortable perception of viewpoint change, the prediction values provided by the MVPDM accord with the expected behavior of the viewers.

C. Correlation of the Smoothness Results with d_{perc} and VSS_{perc}

The MVPDM [14] proposes that a prediction of the smoothness scores can be obtained as a linear combination of d_{perc} and VSS_{perc} . To assess this proposal, we have trained a linear regression model to predict the smoothness MOS values of Section V from the d_{perc} and VSS_{perc} parameters computed for the multiview sequences and frames.

As explained in Section II, d_{perc} (and therefore, VSS_{perc}) depends on two perceptual factors, α , and β , that modify the relevance of disparity values of objects in front and behind the convergence plane, respectively. In this analysis, the optimum α and β values (α_{opt} , β_{opt}) were derived minimizing the mean prediction error, computed as the absolute difference between the MOS and the prediction values.

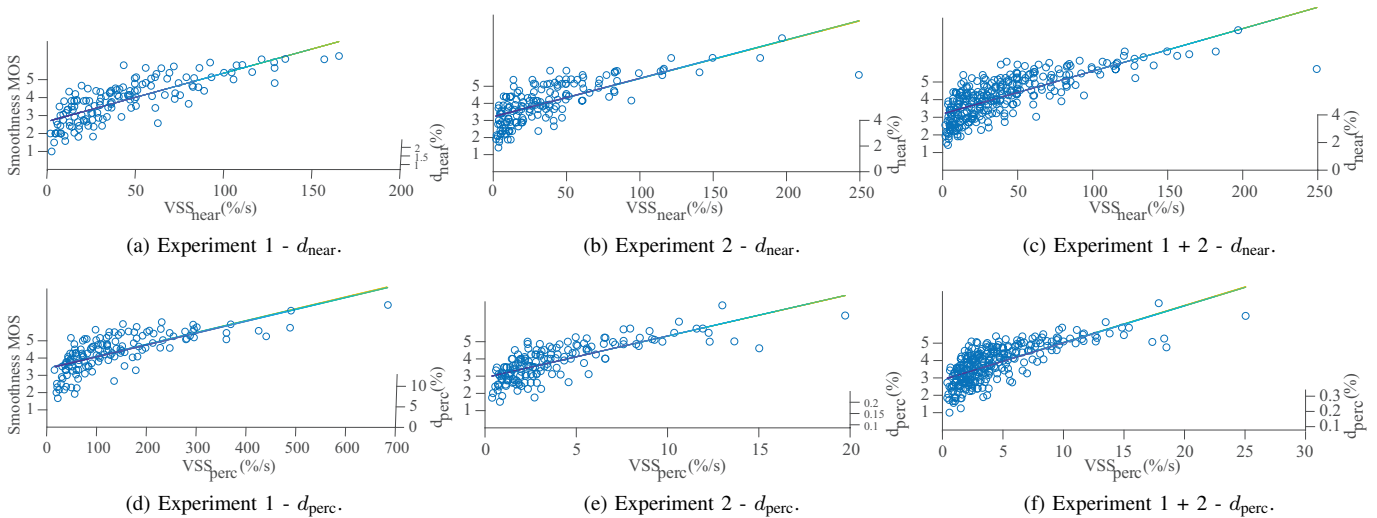


Fig. 11: Best fitting planes for d_{near} and d_{perc} . To plot the 3D graphs, we have chosen a projection that is perpendicular to the fitting plane. This facilitates the interpretation of the prediction errors. Smoothness MOSs are within the [1, 5] range.

TABLE VII: Mean absolute error and standard deviation of the smoothness predictions for Experiments 1, 2, and 1 + 2 using d_{near} and d_{perc} parameterizations. $(\alpha_{\text{opt}}, \beta_{\text{opt}})$ values are shown for each experiment. Plane coefficient values are presented for Experiment 1+2. 10-fold cross-validation results for the d_{perc} model are provided through mean values and 95% confidence intervals. 95% confidence interval results are provided by the value of half-interval size.

		Experiment 1 (sequences)				Experiment 2 (static frames)				Experiments 1 + 2						
		α_{opt}	β_{opt}	\bar{Err}	$\sigma(Err)$	α_{opt}	β_{opt}	\bar{Err}	$\sigma(Err)$	α_{opt}	β_{opt}	\bar{Err}	$\sigma(Err)$	A	B	C
Trained using all samples	d_{near}	-	-	0.60	0.41	-	-	0.67	0.49	-	-	0.64	0.47	0.02	-0.79	3.20
	d_{perc}	1.9	1.3	0.58	0.46	0.4	0	0.55	0.40	0.5	0	0.60	0.43	0.21	-8.91	3.61
10-fold cross-val (d_{perc})	mean value	2.02	1.44	0.61	-	0.4	0	0.56	-	0.49	0	0.57	-	0.22	-9.19	3.62
	95% CI	0.08	0.10	0.09	-	0.00	0.00	0.06	-	0.02	0.00	0.05	-	0.01	0.58	0.04

Fig. 11 depicts the plane that fits best to the smoothness MOS values, for the sequences, the static frames, and the combination of both, and the two previously mentioned parameterizations (d_{perc} and d_{near}). This plane represents the smoothness prediction values from d and VSS values. The corresponding mean absolute errors, standard deviations and $(\alpha_{\text{opt}}, \beta_{\text{opt}})$ pairs are presented in Table VII. In [14], an angular resolution parameterization was also analyzed, but it has been omitted here as it provided the worst results. In addition, a 10-fold cross-validation test has been performed for the d_{perc} model. Mean absolute error results and 95% confidence intervals for the cross validation test are reported in Table VII. The cross validation results show stable results with low confidence intervals for α β and mean error, demonstrating that the MVPDM model does not overfit to the data.

1) *Experiment 1: Smoothness Prediction of Video Content:* In the case of the video content, both planes (d_{near} and d_{perc}) have similar mean absolute errors (see Table VII). Moreover, we observe a significant variability in the smoothness scores for similar (d_{perc} , VSS_{perc}) values, especially for the lowest values of (d_{perc} , VSS_{perc}), as shown in Fig. 11 (d). This indicates certain limitations in MVPDM as a model to predict precise smoothness scores for video sequences. This may be influenced by two factors:

- A linear model may be over-simplistic to reflect the

user perception of smoothness. The use of more complex models needs to be studied. However, the training of such models requires a higher amount of data to avoid overfitting problems. The MV-HMD system and the methodology proposed in this work can enable the extraction of that data.

- As stated in Section V, the user perception of smoothness varies for different disparity distributions of a scene. Thus, for sequences with disparity variation along time, it is not clear which are the scoring criteria of the subjects, e.g., is the user scoring the worst instant, the best instant, or averaging the perception along time? While further subjective research could clarify this second factor, Experiment 2 aims to disaggregate the effect of a variable disparity along time.

2) *Experiment 2: Smoothness Prediction of Static Content:* The prediction accuracy of smoothness scores improves for the static content (see Fig. 11 (e)), with a slightly lower mean prediction error (Table VII), and outperforming the d_{near} parameterization. This improvement is related to the use of static content that maintains the scene disparity constant in time. Fixing the disparity eliminates a possible variability in the user's scoring criteria for the presence of a disparity that varies within a test clip.

Nevertheless, a dispersion of smoothness scores still exists

for low (d_{perc} , VSS_{perc}) values, and thus, future work should focus on the development of more accurate models for the prediction of smoothness perception in view transitions.

3) *Experiment 1 + 2: Combining the Results of Video and Static Content:* Finally, Figs. 11 (e) and (f) are obtained combining the data of Experiments 1 and 2. The prediction error (Table VII) sits between the values obtained for Experiments 1 and 2 separately.

While (α_{opt} , β_{opt}) values differ between Experiments 1 and 2 (see Table VII), the (α_{opt} , β_{opt}) values for Experiment 1+2 are very similar to the ones obtained for Experiment 2, i.e., ($\alpha_{\text{opt}} = 0.4/0.5$, $\beta_{\text{opt}} = 0$). This indicates that, within the limitations of the model, these values are valid for both static frames and sequences, i.e., prediction errors for Experiment 1 samples do not increase significantly the mean prediction error. For this experiment, Table VII also shows the value of the plane coefficients (A , B , C) for the d_{perc} and d_{near} parameterizations:

$$\text{Smoothness} = A * VSS_{\text{perc/near}} + B * d_{\text{perc/near}} + C. \quad (3)$$

D. Conclusion

The MVPDM has been used to predict the speed comfort scores for the set of tests. The high values of the speed comfort predictions accord with the perception of viewers, who were requested to move with a speed that produced a comfortable viewing experience. Also, the analysis of the correlation of smoothness scores with the MVPDM parameterization indicates that MVPDM is useful to establish an approximate prediction of smoothness scores. Results for Experiments 1+2 indicate that a single model can be derived to predict the smoothness scores for static and moving content. However, improved models must be developed to increase the accuracy of these predictions, especially for low disparity values. For that, more subjective data will be necessary, and the MV-HMD player and methodology presented in this work will be useful to collect that data.

VII. CONCLUSIONS

This paper presents a system and methodology for the assessment of QoE factors specific to dense multiview content using HMDs. Given the limited availability of commercial lightfield displays, the use of an HMD provides a more realistic environment than previous approaches, allowing the user to move freely, varying its perspective of the scene. Hence, this work might be used for future assessment, or even considered in the elaboration of new lightfield/VR QoE recommendations.

Using the MV-HMD player and the presented methodology, we have analyzed how the variation of the switching distance, content, and the user's movement affect the perception of smoothness in the view transition. We can state that the array's density (switching distance) and the user movement speed (UMS) influence the smoothness perception, increasing for smaller switching distances and higher UMS. The statistical analysis of the smoothness scores show the statistical significance of UMS and switching distance in the perception of smoothness.

Finally, we have used the subjective data to assess the performance of the prediction metric MVPDM in different

aspects. First, we have verified that MVPDM yields high speed comfort prediction values for all test clips and subjects, which were requested to move with a speed that produced a comfortable viewing experience. Second, we have studied the correlation of the smoothness results with the MVPDM parameterization. The results indicate that MVPDM is useful to establish an approximate prediction of smoothness scores, but improved models must be developed to increase its accuracy.

Future work will include the use of the proposed methodology, with a higher number of subjects and content variety, to collect more data that can enable the development of accurate models for the prediction of subjective factors specific to multiview content (Section VI). Also, further research could focus on the analysis of the typical scoring criteria of a user who watches multiview content with a head-speed that necessarily varies along time (Section V). This analysis could lead to an improvement in the characterization of the UMS.

REFERENCES

- [1] J. Gutiérrez, F. Jaureguizar, and N. García, "Subjective comparison of consumer television technologies for 3D visualization," *Journal of Display Technology*, vol. 11, no. 11, pp. 967–974, 2015.
- [2] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," in *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV*, vol. 7690. Int. Society for Optics and Photonics, 2010, p. 76900B.
- [3] M.-C. Park and S. Mun, "Overview of measurement methods for factors affecting the human visual system in 3D displays," *Journal of Display Technology*, vol. 11, no. 11, pp. 877–888, 2015.
- [4] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, 2017.
- [5] Y. Takaki, "Development of super multi-view displays," *ITE Trans. on Media Technology and Applications*, vol. 2, no. 1, pp. 8–14, 2014.
- [6] J.-Y. Son, H. Lee, B.-R. Lee, and K.-H. Lee, "Holographic and light-field imaging as future 3-D displays," *Proceedings of the IEEE*, vol. 105, no. 5, pp. 789–804, 2017.
- [7] L. Janowski and M. Pinson, "The accuracy of subjects in a quality experiment: A theoretical subject model," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2210–2224, Dec 2015.
- [8] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.
- [9] T. Kim, J. Kang, S. Lee, and A. C. Bovik, "Multimodal interactive continuous scoring of subjective 3d video quality of experience," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 387–402, Feb 2014.
- [10] V. Kiran Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [11] F. M. Ribeiro, J. F. de Oliveira, A. G. Ciancio, E. A. da Silva, C. R. Estrada, L. G. Tavares, J. N. Gois, A. Said, and M. C. Martelotte, "Quality of Experience in a Stereoscopic Multiview Environment," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 1–14, 2018.
- [12] D. Runde, "How to realize a natural image reproduction using stereoscopic displays with motion parallax," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 376–386, 2000.
- [13] F. Speranza and W. Tam, "Perceived smoothness of viewpoint transition in multi-viewpoint stereoscopic displays," in *Stereosc. Displays and Virtual Reality Systems XII*, vol. 5664, Mar. 2005, pp. 72–82.
- [14] P. Carballeira, J. Gutiérrez, F. Morán, J. Cabrera, F. Jaureguizar, and N. García, "Multiview perceptual disparity model for super multiview video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 113–124, 2017.
- [15] ISO/IEC JTC1/SC29WG11, "Overview of standards activities related to immersive media," Output doc. N18130, 125th MPEG meeting, Marrakech, MA, Jan. 2019.
- [16] Rec. ITU-R BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems," Aug. 2012.

- [17] P. Lebreton, M. Barkowsky, A. Raake, and P. Le Callet, "3D Video," in *Quality of Experience: Advanced Concepts, Applications and Methods*, S. Möller and A. Raake, Eds. Springer, 2014, pp. 299–313.
- [18] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The Road to Immersive Communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 974–990, Apr. 2012.
- [19] P. Carballeira, J. Gutiérrez, F. Morán, J. Cabrera, and N. García, "Subjective evaluation of super multiview video in consumer 3D displays," in *International Workshop on Quality of Multimedia Experience*, Costa Navarino, GR, May 2015, pp. 1–6.
- [20] 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.
- [21] 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)*, Dec 2016, pp. 1–6.
- [22] P. A. Kara, M. G. Martini, P. T. Kovacs, 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 *2016 International Conference on 3D Imaging (IC3D)*, Dec 2016, pp. 1–7.
- [23] R. 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.
- [24] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, "Client-driven selective streaming of multiview video for interactive 3dtv," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1558–1565, Nov 2007.
- [25] K. Hopf, P. Chojeci, F. Neumann, and D. Przewozny, "Novel autostereoscopic single-user displays with user interaction," in *Proc. SPIE, Three-Dimensional TV, Video, and Display V*, vol. 6392, Oct. 2006.
- [26] E. Upenik, I. Viola, and T. Ebrahimi, "A rendering solution to display light field in virtual reality," in *26th European Signal Processing Conference (EUSIPCO)*, 2018.
- [27] A. Serrano, I. Kim, Z. Chen, S. DiVerdi, D. Gutierrez, A. Hertzmann, and B. Masia, "Motion parallax for 360° RGBD video," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1817–1827, May 2019.
- [28] Unity3D, [Online]: <https://www.unity3d.com>. Accessed in 2018.
- [29] Oculus, [Online]: <https://www.oculus.com>. Accessed in 2018.
- [30] T. Senoh, K. Yamamoto, N. Tetsutani, and H. Yasuda, "Proposal of new Champagne Tower 2 sequence," Contrib. M38063, 116th MPEG meeting, Ghengdu, CH, Oct. 2016.
- [31] ISO/IEC JTC1/SC29/WG11, "Overview of MPEG-I Visual Test Materials," Output doc. N17718, 123th MPEG meeting, Ljubljana, SI, Jul. 2018.
- [32] P. T. Kovacs, A. Fekete, K. Lackner, V. K. Adhikarla, A. Zare, and T. Balogh, "Big Buck Bunny light-field test sequences," Contrib. M36500, 112th MPEG meeting, Warsaw, PL, Jun. 2015.
- [33] M. Tanimoto, T. Fujii, and N. Fukushima, "1D Parallel Test Sequences for MPEG-FTV," Contrib. M15378, 84th MPEG meeting, Archamps, FR, Apr. 2008.
- [34] ISO/IEC JTC1/SC29/WG11, "Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation," Output doc. N15348, 112th MPEG meeting, Warsaw, PL, Jun. 2015.
- [35] Rec. ITU-R BT. 500-13, "Methodology for the subjective assessment of the quality of television pictures," Jan. 2012.
- [36] Rec. ITU-T P.913, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," Mar. 2016.
- [37] Rec. ITU-T P.915, "Subjective assessment methods for 3D video quality," Mar. 2016.
- [38] J. Thatte and B. Girod, "Towards perceptual evaluation of six degrees of freedom virtual reality rendering from stacked omnistereo representation," *Electronic Imaging*, vol. 2018, no. 5, pp. 1–6, 2018.
- [39] M. Chatfield and A. Mander, "The Skillings-Mack test (Friedman test when there are missing data)," *Stata Journal*, vol. 9, no. 2, pp. 299–305, Apr. 2009.
- [40] M. D. Ernst, "Permutation Methods: A Basis for Exact Inference," *Statistical Science*, vol. 19, no. 4, pp. 676–685, Nov. 2004.



Javier Cubelos received the BSc of Engineering in Telecommunication Technologies and Services in 2016, and the MSc in Telecommunication Engineering in 2018, both from the Universidad Politécnica de Madrid (UPM), Spain. He was a member of UPM's Grupo de Tratamiento de Imágenes (Image Processing Group) from 2015 to 2018. Since then, he works at BDEO Technologies SL as a deep learning engineer. His research interests include immersive media, computer vision and deep learning.



Pablo Carballeira received the degrees of Ingeniero de Telecomunicación (five-year engineering program) in 2007, and the Doctor Ingeniero de Telecomunicación (Ph.D. in Communications) in 2014, both from the Universidad Politécnica de Madrid (UPM), Spain. From 2008 to 2017 he was a member of UPM's Grupo de Tratamiento de Imágenes (Image Processing Group). Since 2017, he is an Assistant Professor at Universidad Autónoma de Madrid, and a member of its Video Processing and Understanding Lab. His research interests include video coding, computer vision and quality of experience evaluation for immersive visual media, such as free-navigation and lightfield video. He has been actively involved in European and Spanish research projects on these topics, and in the international standardization activities of MPEG (Moving Picture Experts Group, formally ISO/IEC JTC 1/SC 29/WG 11) related to free-viewpoint and immersive 3D video.

include video coding, computer vision and quality of experience evaluation for immersive visual media, such as free-navigation and lightfield video. He has been actively involved in European and Spanish research projects on these topics, and in the international standardization activities of MPEG (Moving Picture Experts Group, formally ISO/IEC JTC 1/SC 29/WG 11) related to free-viewpoint and immersive 3D video.



Jesús Gutiérrez received the Telecommunication Engineering degree (five-year engineering program) in 2008 from the Universidad Politécnica de Valencia (UPV), Spain. He received the Communications Technologies and Systems Master degree (two-year MS program) in 2011, and the Ph.D. degree in Telecommunication in 2016, both from the Universidad Politécnica de Madrid (UPM), Spain. He was a member of UPM's Grupo de Tratamiento de Imágenes (Image Processing Group) from 2010 to 2016. Since then, he is a post-doctoral researcher

(currently Marie Curie PRESTIGE fellow, previously Marie Curie fellow within the PROVISION ITN) at the IPI team of the LS2N laboratory of the Université de Nantes, France. His research interests are in the area of image and video processing, immersive media, evaluation of multimedia quality of experience, human behavior and visual perception.



Narciso García received the Ingeniero de Telecomunicación degree (five years engineering program) in 1976 (Spanish National Graduation Award) and the Doctor Ingeniero de Telecomunicación degree (Ph.D. in Communications) in 1983 (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Spain. Since 1977, he has been a member of the faculty of the UPM, where he is currently a Professor of Signal Theory and Communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group),

UPM. He has been actively involved in Spanish and European research projects, also serving as an evaluator, a reviewer, an auditor, and an observer of several research and development programs of the European Union. He was a Co-Writer of the EBU proposal, base of the ITU standard for digital transmission of TV at 34–45 Mb/s (ITU-T J.81). He was an Area Coordinator of the Spanish Evaluation Agency (ANEP) from 1990 to 1992 and he was the General Coordinator of the Spanish Commission for the Evaluation of the Research Activity (CNEAI) from 2011 to 2014. He has been the Vice-Rector for International Relations of the Universidad Politécnica de Madrid from 2014 to 2016. He was a recipient of the Junior and Senior Research Awards of the Universidad Politécnica de Madrid in 1987 and 1994, respectively. His current research interests include digital video compression, computer vision, and quality of experience.