MultiView Perceptual Disparity Model for Super MultiView Video

Pablo Carballeira, Jesús Gutiérrez, Francisco Morán, Julián Cabrera, Fernando Jaureguizar, and Narciso García

Abstract—Super MultiView (SMV) video display is the most promising technology for 3-D glasses-free visualization. Although only a few prototypes are currently available, the research on technical and perceptual factors related to this approach is crucial. This paper presents a novel model to capture the subjective perception of SMV, called the MultiView Perceptual Disparity Model (MVPDM), by means of a parametrization of the relation between: 1) capture and scene settings, and 2) perception of speed comfort and smoothness in the viewpoint transition. The MVPDM is based on a novel parameter: the perceptual disparity, that captures appropriately the perceptual cues specific to SMV visualization. The model has been validated using the results of subjective tests on realistic SMV content as benchmark. On the one hand, the subjective results show a high correlation with the MVPDM parametrization, outperforming previous approaches. On the other hand, this test provides useful information about the parameters of the SMV sequences that should be used to guarantee satisfactory visual experience. Thus, the MVPDM constitutes a valuable tool for the design of subjective evaluation and content creation of SMV.

Index Terms—Perceptual model, quality of experience, subjective evaluation, Super MultiView video.

I. INTRODUCTION

Seeking the creation of immersive video displays with depth perception, a great development of 3D display technology has taken place in the last years. First, mainly leading to the arrival of stereoscopic displays to the consumer market. However, stereoscopic systems have not achieved total acceptance of consumers, due to reasons including: the lack of high quality 3D video content, the need of wearing specific glasses, and the absence of motion parallax cues entailing a deficient immersive experience. Although the next generation of multiview auto-stereoscopic displays solves the glasses-free and motion parallax issues up to a certain extent, it can be argued that auto-stereoscopic display technology does not provide a satisfactory Quality of Experience (QoE) for its price [1]. In particular, in addition to the loss of resolution needed to provide multiple views, other critical factors affect the user experience and thus limit the acceptance of auto-stereoscopic monitors. For example: the limited extension of the comfortable viewing cone, the discontinuous transition through viewpoints caused by a low view density, and the accommodation-vergence conflict (the viewer’s eyes focus on the screen while they converge in the plane where the objects are projected, which may cause visual discomfort) [2]. Therefore, an intense research work towards advanced immersive, and glasses-free, display technologies has continued over these years [3]. Holographic, integral imaging, and Super MultiView (SMV) displays are being studied and developed seeking the possibility of offering a high 3D QoE.

SMV displays seem to be the most promising glasses-free 3D visualization technology. They fulfill the so-called SMV condition [4], in which the viewing zone pitch (distance between two adjacent views) is smaller than the pupil diameter, thus alleviating the accommodation-vergence conflict that auto-stereoscopic displays present [5]. As any multiview visualization system, SMV displays generate a set of discrete views and distribute them over the viewing field in front of the screen. Thus, a series of viewing zones are offered where the observers can perceive motion parallax as they move their heads, seeing different perspectives of the scene [2]. To satisfy this condition, SMV displays need to provide dense viewing zones, requiring a large set of discrete views, typically more than 80 [6]. SMV displays can be constructed by means of different technologies [4]: focused light array, multi-projection, flat panel, or time multiplexing. SMV displays present less technical bottlenecks than, for instance, holographic displays [7], [8]. A conceptual difference between SMV and holographic displays is that SMV displays intend to reconstruct light rays while holographic displays intend to reconstruct light wavefronts.

SMV displays are still under research, and not yet widely available. Although only a few prototypes have been presented [8], [9], there is a current need of evaluating the impact on the viewers’ QoE of certain effects related to SMV visualization, such as horizontal-parallax smoothness or visual comfort. Researchers in the video coding community are already interested in coding technology for SMV content [10], [11]. It is thus required to understand the influence of the aforementioned effects, as subjective assessment is essential in the evaluation of new 3D video coding technologies. Some subjective studies have been already presented to evaluate different aspects of SMV technology, such as horizontal-parallax smoothness using a head-up display [12], simulated visual comfort effects [13],
and coding effects [14]. Nevertheless, the research on reliable evaluation methodologies concerning 3D video content, and especially new 3D display technologies, is still an open issue, requiring a revision of traditional methods [3], [15].

To evaluate the viewing experience provided by SMV displays, it is necessary to understand the main factors influencing the 3D QoE, in relation to issues such as horizontal parallax, depth quality, and visual comfort.

This work focuses on QoE factors that collect the viewer perception of view transition, which are not present in fixed-viewpoint stereo systems. This paper shows that the disparity distribution within the SMV view set can provide new insights on how the scene and capture settings affect the perception of view transition. A novel parametrization, the MultiView Perceptual Disparity Model (MVPDM), aims at providing answers to questions such as the relation between: i) camera arrangement, scene depth, view density, and head movement speed, and ii) perceived speed and smoothness in the view transition. While previous parametrizations only considered the camera arrangement geometry, the MVPDM utilizes the disparity perceived by users to compute the parameters that are relevant in that subjective perception, and establish relationships between the two sets. This parametrization is a very valuable tool to guide the subjective evaluation of SMV, and ultimately the SMV content creation, by giving hints to configure scene parameters, such as depth or density of cameras, for an acceptable viewing experience. Previous approaches to this objective have already provided relevant inputs for subjective evaluations of SMV content within MPEG's FTV community [11], [16].

The main advantages of the MVPDM are the following:

1) Its parameters are founded on the disparity between adjacent views, perceived by users, instead of the geometry of camera arrangement (camera angle or distance). Thus, it aggregates the contribution of the different parameters that influence the user perception. Also, the MVPDM covers jointly both linear and arc camera arrangements (or any other).

2) It has a better correlation with the perception of visual comfort than previous approaches. It has been validated subjectively simulating the SMV display functionality in a stereoscopic display by means of view-sweeping through the whole set of views. The correlation results between the MVPDM and the results of subjective tests are shown, outperforming alternative parametrizations [16], [17]. It represents better the perception of visual comfort, taking into account more factors that are relevant to the perception of SMV content than previous parametrizations.

3) It has been assessed using multiview sequences captured by real cameras or computer generated (CG) content that simulates scenes with realistic depth, in contrast to previous approaches.

Furthermore, the MVPDM encompasses other technologies and application scenarios besides SMV displays, and is applicable to any stereoscopic video with viewpoint change. This includes auto-stereoscopic displays and the simulation of SMV displays content using stereoscopic or head-mounted displays. In fact, MVPDM has helped to define the evaluation of the responses to the recent CIE of FTV [10] and, furthermore, could be also applied to other application scenarios, such as Free-Navigation (FN) video content. In particular, a preliminary approach to the MVPDM has been very useful in defining the minimum comfortable camera density in a view path for the FN scenarios in the CIE of FTV [11].

The rest of the paper is structured as follows: firstly, the MVPDM is described in detail in Section II, along with the main contributions that come with it. Secondly, the details of the subjective assessment test carried out with current 3D display technologies are described in Section III. Then, Section IV presents the results of the subjective tests analyzing different factors of SMV perception, validation of the proposed parametrization with the subjective results and discussion of the results. Finally, Section V provides general conclusions.

II. PARAMETRIZATION OF SUBJECTIVE EVALUATION OF SMV: DESCRIPTION OF THE MVPDM

A. On the Relation Between Disparity and Camera Arrangement Geometry

Fig. 1 depicts a typical SMV scenario, where the scene is captured from a set of $N$ viewpoints (typically more than 80) corresponding to real or virtual (obtained by view synthesis [18]) cameras. The camera arrangements currently used for the creation of SMV content are not unique [10], covering different configurations such as linear parallel, linear convergent, or arc
array. Here onwards, an SMV camera arrangement such as those in Fig. 1 is assumed, where camera parameters and baseline or angle are constant. It is also assumed that the depth of the objects with respect to the camera arrangement is that of the central camera. Given these assumptions, the disparity distribution of a given scene is the same for all pairs of consecutive cameras.

Let us now discuss how the disparity among a set of views captures the effect of multiple elements of a given camera arrangement, camera parameters, and scene characteristics namely: (i) field of view (FOV), (ii) camera baseline, (iii) camera angle, (iv) scene depth, and (v) convergence. Such discussion justifies that the parameters of the MVPDM are founded on disparity instead of on camera angle or distance, to better represent the subjective perception of SMV visualization.

Without loss of generality, let us consider the disparity between two cameras, \( C_1 \) and \( C_2 \) that are related by a rotation about axis \( y \) and a translation in the plane defined by axes \( x \) and \( z \), as shown in Fig. 2. The origin of coordinates is located at the optical center of \( C_1 \) (\( O_1 \)), and axis \( z \) is normal to the image plane of \( C_1 \). If the point \( X \) (world coordinates) is considered at depth \( Z \), its projection \( x_2 \) (image coordinates) to the image plane of \( C_2 \) is [19]:

\[
x_2 = K [R] - RT \cdot X \tag{1}
\]

where \( K \) is the intrinsic parameter matrix of \( C_2 \), \( R \) is the rotation matrix between \( C_1 \) and \( C_2 \) for angle \( \theta \), and \( T \) is the translation vector between \( O_1 \) and \( O_2 \). Therefore:

\[
K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \quad T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \tag{2}
\]

where \( f \) is the focal length and \( (c_x, c_y) \) is the principal point. Developing Eq. (1), the expression in Eq. (3) is obtained (shown at the bottom of the page).

Converting from homogeneous to Cartesian coordinates:

\[
x_2 = \begin{bmatrix} Z \sin(\theta) + c_x - t_x \cos(\theta) - t_z \sin(\theta) \\ Z \cos(\theta) + t_x \sin(\theta) - t_z \cos(\theta) \end{bmatrix} + \begin{bmatrix} c_y \\ 1 \end{bmatrix}. \tag{4}
\]

The projection \( x_1 \) of \( X \) to the image plane of \( C_1 \) is:

\[
x_1 = \begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix}. \tag{5}
\]

\[
x_2 = K \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sin(\theta) (Z - t_x \cos(\theta) - t_z \sin(\theta)) \\ 0 \\ \cos(\theta) (Z + t_x \sin(\theta) - t_z \cos(\theta)) \end{bmatrix}
\]

\[
= \begin{bmatrix} f \left[ Z \sin(\theta) - t_x \cos(\theta) - t_z \sin(\theta) \right] + c_x \left[ Z \cos(\theta) + t_x \sin(\theta) - t_z \cos(\theta) \right] \\ c_y \left[ Z \cos(\theta) + t_x \sin(\theta) - t_z \cos(\theta) \right] \end{bmatrix}.
\tag{3}
\]
change are presented, discussing previous parametrizations of those elements and their limitations.

Those elements are: i) Viewing Range (VR), ii) View Density (VD), and iii) View-Sweep Speed (VSS). VR is defined as the span that is covered by the whole camera range (N cameras in the arrangement), and can be described in different ways: classically, by the angle covered by the set of cameras or the end-to-end baseline distance. The other parameters that define the perception of SMV content are expressed as follows (T represents the end to end view-sweeping time):

\[
VD = \frac{N}{VR} \quad VSS = \frac{VR}{T}.
\]

Therefore, an SMV sequence with a given VR and VD, is viewed for a certain VSS. These elements are specific to the visualization of multiview content with viewpoint change and do not have to be considered in the evaluation of fixed-viewpoint monoscopic or stereoscopic video.

VR, VD and VSS, or analogous definitions, have been used in previous works which cover the perception of multiview content with viewpoint change [17], [21]. However, previous approaches in defining a parametrization of those elements present several limitations. Mainly: i) the subjective assessment has been done using non-realistic synthetic images [12], [13], and ii) they take into account a limited number of parameters, or parameters that are specific to a certain camera configuration, such as distance or angle between cameras [17], [21]. On the other hand, the work in [16] is the first approach to overcome these limitations, and provides a parametrization which is valid for different camera configurations and realistic scene depth.

At the display side, users do not directly perceive camera arrangement parameters, such as angle or baseline distance. Thus, this work proposes to use an enriched characterization of VR, VD and VSS that takes into account the user perception (speed comfort and smoothness in the view transition). The element that is perceived, and that affects the visual experience, in the view transition is the disparity between adjacent views. As shown in Section II-A the disparity between a camera pair collects the influence of the camera arrangement and scene settings (Eq. (6)), and not only a limited set of camera settings, as in previous parametrizations. Note that this disparity does not refer to the binocular disparity (stereo pair) and can be applied to any pair of cameras. As an example, consider two sets of cameras in the same arrangement. If, for one of them the FOV is wide and for the other is narrow, the subjective perception of speed comfort and smoothness would be radically different. While a parametrization based on angle baseline distance (same for those arrangements) does not reflect this difference, a richer parametrization based on disparity does. Thus, to better reflect the visual experience, we propose a perceptual parametrization of VR, VD and VSS founded on the disparity between views: the MVPDM, that is described in Section II-C.

C. Perceptual Disparity and MVPDM Parametrization

Here, a parametrization of VR, VD and VSS is proposed, based on the disparity \(d\). This parametrization incorporates the effect of all those factors in the subjective perception, and is common to all camera arrangements. Considering the discussion in Section II-A between two adjacent cameras, and recalling the assumption that the camera distance (or angle between adjacent cameras) is constant. VR can be expressed as \(VR = (N-1)d\). As VD and VSS are derived from VR, they also depend on the disparity.

However, as stated in Eq. (6), \(d\) is a function of depth \(Z\) (depicted in Fig. 3). Here, and for the rest of the paper, objects at the convergence plane yield null disparity \((d = 0)\), while objects in front of and behind that plane produce negative and positive disparity, respectively. Therefore, VR, VD, and VSS are also dependent on depth and are defined as functions of \(Z\): VR\(Z\), VD\(Z\), VSS\(Z\). To parametrize a given sequence, it is more useful to have representative values of VR, VD, and VSS, and not functions. A simple approach to obtain those representative values is to use \(d(Z_{min})\) and \(d(Z_{max})\) (disparity values of those space points that are closest to and furthest away from the viewer) [16].

To provide a parametrization better correlated with subjective experience, the MVPDM defines a perceptual disparity value, \(d_{perc}\), to represent the disparity distribution of the scene. It is the fundamental element of the MVPDM, providing advantages with respect to previous approaches based on disparity [16], and is computed as (see Fig. 4):

\[
d_{perc} = \int_{d_{min}}^{d_{max}} \text{hist}(\delta) \cdot \omega(\delta) \, d\delta,
\]
where \( \text{hist}(d) \) is the disparity histogram and \( \omega(d) \) is the following weighting function:

\[
\omega(d) = \begin{cases} 
|d|^\alpha & d \leq 0 \\
|d|^\beta & d > 0 
\end{cases}.
\]

\( \alpha, \beta \) are perceptual factors that modify the relevance of higher disparity values. Instead of using a maximum-disparity representative, \( \delta_{\text{perc}} \) aggregates the depth distribution of the whole scene. Intuitively, it can be understood that the global disparity perceived in a viewpoint change is low if most of the scene content is located at a given depth that produces null disparity, while only a few pixels produce high disparity. The weight \( \omega(d) \) is used to ponderate the perceptual relevance of objects of the scene that present high disparity. Two different factors are used to differentiate the relevance of the disparity of objects closer than the convergence plane (\( d < 0 \)), generally more relevant, and further than the convergence plane (\( d > 0 \)), generally less relevant. An evaluation of the most adequate values of \((\alpha, \beta)\) is presented in Section IV-B by correlating the model with subjective experiments.

Given \( \delta_{\text{perc}} \), the parameters of the MVPDM, \( VR_{\text{perc}} \), \( VD_{\text{perc}} \), and \( VSS_{\text{perc}} \) are now obtained as:

\[
VR_{\text{perc}} = (N - 1) \delta_{\text{perc}},
\]

\[
VD_{\text{perc}} = \frac{N}{(N - 1) \delta_{\text{perc}}},
\]

\[
VSS_{\text{perc}} = \frac{(N - 1) \delta_{\text{perc}}}{T}.
\]

\( VR_{\text{perc}} \) (pix), \( VD_{\text{perc}} \) (cam/pix) and \( VSS_{\text{perc}} \) (pix/s) are measured in disparity units (pixels) and not angle or distance units.

**D. Parametrization of SMV Sequences and Disparity Histograms**

In this subsection, the parametrization of the sequences in the SMV category [10] is presented, using the MVPDM described in Section II-C. In particular, four different SMV video contents are considered, as well as arc, linear parallel, and linear convergent camera configurations. The main sequence characteristics and the set of values that are relevant to the MVPDM are shown in Table I, and samples of the sequence contents in Fig. 5. More details on the camera arrangements and sequence content can be found in [22] and [23]. BU in Table I refers to Blender Units, used to create the Butterfly and Flowers CG scenes [22]. This test set comprises SMV sequences for different camera arrangement types, includes natural and realistic CG content, and has been used by the MPEG subgroup for the evaluation of the CfE of FTV [10].

As noted in Eq. (8), \( \delta_{\text{perc}} \) depends on the integral of the disparity histogram. Therefore, for the purpose of Table I, to characterize \( d \) and \( VR \), two representative values of those functions have been taken, namely \( d(Z_{\text{near}}) \), \( d(Z_{\text{far}}) \), \( VR_{\text{near}} \), and \( VR_{\text{far}} \). Furthermore, given that 1D-SMV content deals with horizon disparity only, to uncouple the units in the MVPDM from the spatial resolution (that changes from sequence to sequence), disparity \( d \) is measured as a percentage of horizontal resolution \( W \) (width):

\[
d(\%) = 100 \frac{|d|}{W}.
\]

Thus, the MVPDM parameters are measured in the following units: \( VR \) (\%), \( VD \) (cam/\%) and \( VSS \) (\%/s). Table I also includes the values of angular parameters \( (\theta_{\text{near}} \text{ and } \theta_{\text{far}}) \), that are used in alternative parametrizations that are defined and compared to the MVPDM in Section IV-B.

---

**TABLE I**

**CHARACTERISTICS OF SMV SEQUENCES AND MVPDM PARAMETRIZATION VALUES**

<table>
<thead>
<tr>
<th>ID</th>
<th>Camera arrangement</th>
<th>Resolution</th>
<th>frame rate</th>
<th>camis</th>
<th>( Z_{\text{near}} )</th>
<th>( Z_{\text{far}} )</th>
<th>( d(Z_{\text{near}}) )</th>
<th>( d(Z_{\text{far}}) )</th>
<th>VR_{\text{near}}</th>
<th>VR_{\text{far}}</th>
<th>VR_{\text{perc}}</th>
<th>VR_{\text{far}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Champagne</td>
<td>Ch</td>
<td>Linear convergent</td>
<td>1280x960</td>
<td>30 fps</td>
<td>80</td>
<td>2032 mm</td>
<td>7784 mm</td>
<td>-55 pix</td>
<td>4.29 %</td>
<td>0 pix</td>
<td>0 %</td>
<td>339.45 %</td>
</tr>
<tr>
<td>Pantomime</td>
<td>Pa</td>
<td>Linear convergent</td>
<td>1280x960</td>
<td>30 fps</td>
<td>80</td>
<td>3908 mm</td>
<td>8222 mm</td>
<td>-20 pix</td>
<td>1.56 %</td>
<td>0 pix</td>
<td>0 %</td>
<td>123.44 %</td>
</tr>
<tr>
<td>Flowers_Linear</td>
<td>FL</td>
<td>Linear parallel</td>
<td>1280x768</td>
<td>24 fps</td>
<td>91</td>
<td>0.232 BU</td>
<td>595 BU</td>
<td>-32.44 pix</td>
<td>2.53 %</td>
<td>-0.01 pix</td>
<td>0 %</td>
<td>228.09 %</td>
</tr>
<tr>
<td>Butterfly_Linear</td>
<td>BL</td>
<td>Linear parallel</td>
<td>1280x768</td>
<td>24 fps</td>
<td>91</td>
<td>1.270 BU</td>
<td>700 BU</td>
<td>-16.86 pix</td>
<td>1.32 %</td>
<td>-0.03 pix</td>
<td>0 %</td>
<td>118.55 %</td>
</tr>
<tr>
<td>Flowers_Arc</td>
<td>FA</td>
<td>Arc</td>
<td>1280x768</td>
<td>24 fps</td>
<td>91</td>
<td>0.232 BU</td>
<td>595 BU</td>
<td>-25.54 pix</td>
<td>1.99 %</td>
<td>7.23 pix</td>
<td>0.56 %</td>
<td>179.58 %</td>
</tr>
<tr>
<td>Butterfly_Arc</td>
<td>BA</td>
<td>Arc</td>
<td>1280x768</td>
<td>24 fps</td>
<td>91</td>
<td>1.270 BU</td>
<td>700 BU</td>
<td>-3.08 pix</td>
<td>0.24 %</td>
<td>13.93 pix</td>
<td>1.09 %</td>
<td>21.66 %</td>
</tr>
</tbody>
</table>
Fig. 6. Disparity histograms of SMV sequences (Table I) obtained using the central pair of adjacent cameras. (a) Champagne: cams 39–40. (b) Pantomime: cams 39–40. (c) Flowers Linear: cams 45–46. (d) Flowers Arc: cams 45–46. (e) Butterfly Linear: cams 45–46. (f) Butterfly Arc: cams 45–46.

Fig. 6 depicts the disparity histograms, hist \( (d) \), for the central camera pair of the SMV content in Table I. Note that depth for Champagne and Pantomime has been obtained using depth estimation algorithms (DERs) [24], whereas the rest of the sequences are generated by computer graphics, where depth can be considered as ground truth. For the Butterfly sequences, the highest bar of the histogram has been truncated for the sake of visibility. Such histograms are used to compute \( d_{\text{perc}} \) using Eq. (8).

III. SUBJECTIVE TEST SETUP

A subjective assessment test was carried out to validate the proposed model and investigate the influence on the perceived QoE of the aforementioned factors related to SMV visualization. Due to the lack of SMV displays, this test has been carried out by simulating the horizontal parallax by means of view-sweeping with a stereoscopic display, which is the method considered by MPEG for current SMV experiments. Alternatively to a user moving his head at a certain speed in front of an SMV display, in the view-sweep sequences the viewpoint is changed with time at a given rate and presented in a stereoscopic display. The main characteristics of the test are detailed in the following subsections.

A. Methodology

The Absolute Category Rating (ACR) [25] methodology was used to evaluate the test sequences. After an initial message indicating the start of the test, the test sequences \( (V_t) \) were shown followed by a message \( (\text{Vote}_i) \) prompting the observers to rate the corresponding video sequence. Also, questionnaires with numbered boxes were used to collect the opinions of the observers, where they were asked to write a mark for the corresponding evaluation. The voting messages had a duration of 12 seconds, to allow the subjects to judge the following factors after watching each test clip:

1) Speed comfort: the participants were asked to evaluate how perceptually comfortable was the view sweeping speed, using the five-grade comfort scale [26].
2) Smoothness: the observers rated the impression of seeing a smooth transition between views (without discontinuities), using the five-grade quality scale [15].
3) 3D quality: the observers were asked to provide a score for their overall 3D QoE, considering depth perception, immersiveness, comfort, etc. They used the five-grade quality scale [15].

The test sessions consisted of a previous visual screening of the subjects, followed by a training process in which some example sequences were shown to them to explain the test methodology and purpose. In particular, the meanings of the three factors to evaluate were carefully explained to the observers. Especially, given the relation between 3D quality and speed comfort, participants were instructed to evaluate both factors identifying whether the source of discomfort (e.g., difficulties to properly perceive 3D) was the sweeping speed or the video content itself. To that end, fixed-viewpoint sequences were used as reference. Also, this training provided a reference to the observers, especially in relation to view sweeping speeds and the 3D capabilities of the display, so they could rate properly the evaluated factors. Furthermore, to reduce contextual effects, each test video was shown twice in each session and different randomizations of the test videos were used (with the condition of not
showing the same source content consecutively). Two observers participated at a time in each session, which lasted 20 minutes approximately.

B. Subjective Test Material

For the subjective tests, the test sequences were prepared by generating view-sweeps with the uncompressed SMV sequences, as depicted in Fig. 7. To generate test sequences with different VSS values for a single content, view-sweeps were shown by changing the view at two different rates: one frame per view (1 fpv) and two frames per view (2 fpv). These two values were selected according to the results of previous subjective experiments [17], trying to have not so long sessions to keep the attention of the participants without tiring them.

The stereo baseline (b) values, measured in camera gaps between left and right views, that was used for each sequence (those were selected to be comfortable in a previous screening), are the following:

1) Linear natural sequences (Ch, Pa): 1
2) Linear CG sequences: (FL, BL): 2
3) Arc CG sequences: (FA, BA): 4

The values of b for the natural sequences are concordant to those used in [27], while the CG sequences present some difference in the b values. All test sequences lasted 10 seconds, and were displayed at a spatial resolution width of 1280 pixels, at its corresponding aspect ratio and frame rate, and covering the whole set of views in the view sweep, that starts from a central stereo pair.

C. Environment

The test area was set according to international recommendations [15], with walls covered by mid-gray curtains, and the ambient lighting conditions were controlled to avoid disturbing reflections. The viewing distance was set at 2.1 m from the display position, which corresponds to a viewing distance of 3H and is within the viewing distance ranges of the displays used.

D. Equipment

A 55” Samsung UE55HU8500L stereoscopic display was used to carry out the subjective tests. It is a curved UHD LED consumer display with active shutter glasses. In addition, the videos were played using a high-performance PC with a Nvidia GeForce GTX 760 graphic card and an updated VLC player [28], which allowed a smooth playback of the sequences.

E. Observers

A total of 20 observers (7 females, 13 males) participated in the tests, all of them having normal or corrected vision. The ages of the participants were between 22 and 48, with an average age of 29. After the tests, a screening of the scores provided by the observers was carried out considering the correlation between their evaluations and the average scores of all the observers [29], which led to discard two observers.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the results for the subjective test described in Section III are presented, and then the validity of the MVPDM is assessed by means of a correlation analysis between its parameters and the subjective test scores.

A. Subjective Results for Speed Comfort, Smoothness, and 3D Quality

Figs. 8–10 show the Mean Opinion Scores (MOSs) computed from the evaluations provided by the observers in the questionnaires. In addition, the 95% confidence intervals (CI) are indicated according to the computation recommended by the standard ITU-R BT.500 [15].

Fig. 8 shows the MOSs for speed comfort provided by the observers for each sequence. As shown, the results are highly dependent on the content and, in general, there are no clear statistically significant differences between using fast view transitions (1 fpv) or slow view transitions (2 fpv) in terms of comfort. On the other hand, the results reflect that sequences with an arc configuration of the cameras may tolerate higher values of VSS. However, considering comfortable a speed rated with a MOS above 3 (taking into account also the CI), it should be noticed that Champagne and Butterfly_Line are not comfortable at both values of VSS. In addition, in the case of Flowers_Arc, it is clear that the faster version of the sequence is preferred, which, apart from the content characteristics, could be due to a better smoothness of the video. Therefore, possible interactions between smoothness and speed comfort should be analyzed in future tests.

Fig. 9 shows the MOSs for smoothness provided by the observers for each sequence. As shown, the results are highly dependent on the content and, in general, there are no clear statistically significant differences between using fast view transitions (1 fpv) or slow view transitions (2 fpv) in terms of comfort. On the other hand, the results reflect that sequences with an arc configuration of the cameras may tolerate higher values of VSS. However, considering comfortable a speed rated with a MOS above 3 (taking into account also the CI), it should be noticed that Champagne and Butterfly_Line are not comfortable at both values of VSS. In addition, in the case of Flowers_Arc, it is clear that the faster version of the sequence is preferred, which, apart from the content characteristics, could be due to a better smoothness of the video. Therefore, possible interactions between smoothness and speed comfort should be analyzed in future tests.

Fig. 9 shows the MOSs obtained for smoothness. Considering acceptable a smoothness rated with a MOS above 3 (taking into account also the CI), better results are obtained with fast view transitions, with a satisfactory smoothness for all sequences, as expected. However, only for Butterfly_Arc a slow view transition is considered acceptable. Again, the results are influenced by the content, and arc configurations seem to offer better visual experiences, which makes sense given the convergent orientation of the cameras. Specifically, these configurations imply
lower disparity values than in linear arrangements for typically the most salient region of the scene, when changing between adjacent views.

Besides, to analyze the quality of the overall 3D experience, the MOSs obtained for 3D quality are shown in Fig. 10. Apart from the content properties that may be the main factor influencing the 3D QoE, in general, the results reflect that both smoothness and VSS may affect the 3D perception. Comparing Figs. 8–10, it can be observed, on the one side, that when the smoothness is acceptable, the speed comfort may influence the 3D visual experience, since the results follow a similar trend. On the other side, when the smoothness is bad, the 3D perception seems to be also clearly worsened.

B. Validation of MVPDM With Subjective Results

To validate the MVPDM, a correlation analysis has been performed, between the subjective results (MOS scores) and the parameters of the MVPDM for the SMV content (presented in Section II-C). Specifically, the correlation between i) subjective speed comfort and VSSperc values, and ii) subjective smoothness and the pairs of values (dperc, VSSperc) has been investigated. In this analysis, the MVPDM has also been compared to the following alternative parametrizations:

1) The parametrization used in [16]: the maximum disparity \( d(Z_{\text{near}}) \) is used in Eq. (10) instead of \( d_{\text{perc}} \), obtaining VSS/VD/VRnear.

2) An angular parametrization (similar to that used in [17]): \( \theta_{\text{ang,near}} \) and \( \theta_{\text{ang,far}} \) represent the angular increment between adjacent cameras, with the angle vertex located at the \( Z_{\text{near}} \) and \( Z_{\text{far}} \) planes, respectively. Refer to Fig. 1 (a) as an example of the angle defined by \( Z_{\text{near}} \) and the camera array. In the arc camera settings, the vertex is located at the convergence plane, and thus \( \theta_{\text{ang,near}} = \theta_{\text{ang.far}} \). The angular parametrizations VSS/VD/VRang,near and VSS/VD/VRang,far are obtained using \( \theta_{\text{ang,near}} \) and \( \theta_{\text{ang,far}} \) instead of \( d_{\text{perc}} \) in Eq. (10).

Table I contains the relevant values of SMV sequences for those alternative parametrizations and Table II shows, for each test sequence used in the subjective tests, the VSS values for these alternative parametrizations.

1) Influence of \((\alpha, \beta)\): The dependence of the MVPDM on parameters \((\alpha, \beta)\) of the weighting function \( \omega (d) \), (refer to Eq. (9)), has been analyzed, looking for the optimum values giving the best correlation results.

Although the two correlation analyses are fully described below, let us advance that the highest correlation is always achieved for positive values of \( \alpha \) (\( \alpha > 0 \)) and negative ones of \( \beta \) (\( \beta < 0 \)). Therefore, as expected, closer objects (\( d < 0 \)) are more perceptually relevant, while objects behind the convergence plane (\( d > 0 \)), or even in the background plane, are not perceptually
relevant. It seems to be directly related with the higher saliency of objects located at or closer than the convergence plane.

Observe that the correlation performance is highly dependent on $\alpha$ values. However, the choice of $\beta$ has a negligible influence in the correlation analyses as its optimal values are in the interval $\beta = [-3, -0.2]$, with a very low variance within that range. Therefore, $\omega(d) \approx 0$ for $d > 0$, and, thus, $\omega(d)$ can be approximated by:

$$\tilde{\omega}(d) = \begin{cases} |d|^\alpha & d \leq 0 \\ 0 & d \geq 0 \end{cases}.$$  \hspace{1cm} (11)

In the following, $\tilde{\omega}(d)$ has been used in Eq. (8) and evaluated the correlation results over parameter $\alpha$ only. Future analyses, subject to the availability of a bigger corpus of SMV sequences with positive disparities, would lead to a finer optimization of $\omega(d)$ for $d > 0$.

2) VSS/Speed Comfort: With the premise that, among the parameters in the MVPDM, only VSS influences the perceived speed comfort, the correlation between speed comfort and VSSperc is evaluated. To compare MVPDM with previous approaches, the performance of VSSperc is checked against those of alternative parametrizations: VSSnear, VSSang.near, and VSSang.far.

Fig. 11 shows the Pearson Linear Correlation Coefficient (PLCC) between the speed comfort (MOS) and different parametrizations for all sequences: VSSperc (for a range of $\alpha$), VSSnear, VSSang.near, and VSSang.far (constant values; see Table II; do not depend on $\alpha$). It can be seen that the correlation is negative, i.e., higher values of VSS lead to lower speed comfort scores. The best correlation values are obtained for VSSperc (with a minimum approximately at $\alpha = 2$), indicating that MVPDM is the best of the four parametrizations. High correlation results with low variance are obtained in an interval of $\alpha = [1.5, 2.5]$.

Fig. 12 shows the scatter plots and 12(a) the best-fitting lines for three of the parametrizations: a) VSSperc with $\alpha = 2$, b) VSSnear, and c) VSSang.near, as VSSang.far was discarded given the poor PLCC results (see Fig. 11). Supporting the comparative PLCC results in Fig. 11, Figure shows a better linear correlation between speed comfort scores and VSSperc than the other two parametrizations. To better understand the benefits of the MVPDM compared to other parametrizations, take these two examples:

1) VSSnear: FA_1fpv produces a high speed comfort score. So, while the value of VSSnear is high due to the presence of pixels with high disparity, VSSperc is much lower, and hence more accordant with the subjective score. This is due to the fact that most of the disparity histogram for FA is concentrated around $d = 0$ (see Fig. 5).

2) VSSang.near: the behavior of VSSang.near is the worst, as this parametrization does not capture the influence of scene depth. For instance, FL_1fpv and Ch_1fpv have a similar value of VSSang.near, while they have radically different subjective scores.

In addition, to disaggregate the correlation results, correlation scores have been obtained for the following subsets of sequences: a) linear arrangements (parallel and convergent), b) fast view transition (1 fpv), and c) slow view transition (2 fpv). Arc-arrangement subset was excluded from this analysis, as the low variation of MOS scores and arrangements provides not significant correlation results. Table III shows correlation values (PLCC, Spearman Rank Order Correlation Coefficient (SROCC), and Root Mean Square Error (RMSE)) for

### TABLE III

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>All sequences</th>
<th>Linear</th>
<th>Fast (1 fpv)</th>
<th>Slow (2 fpv)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLCC</td>
<td>SROCC</td>
<td>RMSE</td>
<td>PLCC</td>
</tr>
<tr>
<td>VSSperc (2)</td>
<td>-0.821</td>
<td>-0.692</td>
<td>0.201</td>
<td>-0.806</td>
</tr>
<tr>
<td>VSSnear</td>
<td>-0.705</td>
<td>-0.448</td>
<td>0.218</td>
<td>-0.724</td>
</tr>
<tr>
<td>VSSang.near</td>
<td>-0.488</td>
<td>-0.432</td>
<td>0.303</td>
<td>-0.310</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VSSperc (2.7)</td>
<td>-0.821</td>
<td>-0.692</td>
<td>0.201</td>
<td>-0.806</td>
</tr>
<tr>
<td>VSSnear</td>
<td>-0.705</td>
<td>-0.448</td>
<td>0.218</td>
<td>-0.724</td>
</tr>
<tr>
<td>VSSang.near</td>
<td>-0.488</td>
<td>-0.432</td>
<td>0.303</td>
<td>-0.310</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VSSperc (1.8)</td>
<td>-0.821</td>
<td>-0.692</td>
<td>0.201</td>
<td>-0.806</td>
</tr>
<tr>
<td>VSSnear</td>
<td>-0.705</td>
<td>-0.448</td>
<td>0.218</td>
<td>-0.724</td>
</tr>
<tr>
<td>VSSang.near</td>
<td>-0.488</td>
<td>-0.432</td>
<td>0.303</td>
<td>-0.310</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VSSperc (1.9)</td>
<td>-0.821</td>
<td>-0.692</td>
<td>0.201</td>
<td>-0.806</td>
</tr>
<tr>
<td>VSSnear</td>
<td>-0.705</td>
<td>-0.448</td>
<td>0.218</td>
<td>-0.724</td>
</tr>
<tr>
<td>VSSang.near</td>
<td>-0.488</td>
<td>-0.432</td>
<td>0.303</td>
<td>-0.310</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the MVPDM (particularized for the optimum value of $\alpha$ in each subset) and alternative parametrizations, for all the tested subsets of sequences. The results show that: i) VSS$_{perc}$ is adequate to represent the perceived speed comfort, outperforming alternative parametrizations in all subsets, and ii) the best correlation results are obtained in an interval of $\alpha = [1.8, 2.7]$.

3) VD, VSS/Smoothness: To evaluate the correlation of the MVPDM with the perceived smoothness, two parameters have to be taken into account: VD and VSS. It is clear that VD influences the perceived smoothness, considering that the density of views is a key factor for the perceived smoothness in view transition, but also VSS has to be taken into account. For example, for two sequences, with the same VD, seen at two different VSSs, the perceived smoothness tends to be different, with higher scores for higher VSS values. This is evident in the results of Fig. 9, where smoothness scores are rather different for the same content (same VD) at the two view-change rates (1 fps and 2 fps). Thus, the correlation of the perceived smoothness with VSS and VD has been evaluated by a multiple linear regression analysis using: i) smoothness (MOS) as the dependent variable, and ii) VSS and $d$ as the independent variables. For convenience in the analysis, $d$ is used instead of VD, (refer to Eq. (10) for the relation between $d$ and VD). This regression analysis can be seen as finding a fitting plane for a set of 3D points.

Fig. 13 shows the Sum of Square Distances (SSD) of the points to the plane using different parametrizations for all sequences (note that VSS depends on $d$; i) ($d_{perc}$, VSS$_{perc}$) for a range of $\alpha$, ii) ($d_{near}$, VSS$_{near}$), and iii) ($\theta_{ang,near}$, VSS$_{ang,near}$). To provide comparable results among the different parametrizations, the independent variables (VSS and $d$, $\theta$) were normalized (divided by the maximum value in each case). It can be seen from the results that, again, MVPDM has a better correlation with the smoothness scores than alternative parametrizations (lower SSD values), with an optimum correlation score around $\alpha = 1.2$. High correlation results with low variance are obtained in an interval of $\alpha = [0.8, 1.5]$. Fig. 14 shows the scatter plot of the 3D points and the best-fitting plane for the ($d_{perc}$, VSS$_{perc}$) parametrization ($\alpha = 1.2$). Table IV shows correlation values (SSD) for the MVPDM and alternative parametrizations, both for all sequences and linear sequences.

4) Remarks From the Validation: The MVPDM presents a high correlation with the results of subjective experiments (both speed comfort and smoothness), outperforming previous approaches based on the angle between cameras [17] or maximum disparity [16]. The MVPDM presents a high correlation with the results of subjective experiments (both speed comfort and smoothness), outperforming previous approaches based on the angle between cameras [17] or maximum disparity [16].

Although, the optimum value of $\alpha$ differs in each of the analyses, both show overlapping ranges of values of $\alpha$ with a minimum deviation from the optimum correlation value. As observed in Fig. 11, high correlation results with low variance between VSS$_{perc}$ and speed comfort MOS scores are obtained in an interval of $\alpha = [1.5, 2.5]$. Besides, Fig. 13 shows the MVPDM parameters ($d_{perc}$, VSS$_{perc}$) present high correlation results with the smoothness score in an interval of $\alpha = [0.8, 1.5]$. Thus, $\alpha = 1.5$ presents a good compromise of high performance in the two correlation analyses: speed comfort and smoothness. This compromise value is valid to parametrize the whole set of test sequences. More extensive analyses subject to the availability of a larger set of SMV test sequences might change slightly this compromise value.

C. Guiding Rules for Subjective Evaluation and Creation of SMV Content

The good correlation results of the MVPDM with the subjective perception of speed comfort and smoothness makes this model a useful tool in guiding the subjective evaluation of SMV, and ultimately the SMV content creation. Specifically, it provides a model to predict the influence of: scene depth distribution, camera arrangement and head-movement speed, in the subjective perception of speed comfort and smoothness.

For example, given the depth distribution of a scene and the camera arrangement, the MVPDM provides solid predictions of speed comfort and smoothness scores at a given user head-movement speed. Another example regarding the SMV content
creation: given the depth distribution of a scene and the distance to a camera arrangement with fixed intrinsic parameters, the MVPDM defines the minimum view density (original or virtual cameras) that is needed for a smooth perception of viewpoint change at a given user head-movement speed.

V. CONCLUSION

This paper presents a model, the MVPDM, that characterizes the main factors that are specific to the subjective perception of SMV. The MVPDM is novelly founded on the perceptual disparity, instead of angle or camera distance. This provides several advantages with respect to previous approaches: i) it covers jointly both linear and arc camera arrangements, and ii) it aggregates the contribution of the set of capture and scene settings that influence the perception of SMV content. In fact, the core of the MVPDM is applicable to visualization of stereoscopic video with viewpoint change, regardless of specific display technologies. For the validation of the model, a correlation analysis between the MVPDM and subjective assessment results has been carried out, using subjective testing procedures currently active for the assessment of SMV processing and compression technologies [11]. As conclusions from this analysis, it can be derived that the MVPDM presents a high correlation with subjective scores, outperforming previous approaches, in both tested subjective factors: speed comfort and smoothness. Overall, the MVPDM is a useful tool in the design of future subjective tests to investigate the QoE for SMV, and much more importantly, for the creation of comfortable SMV sequences. Future work will include the extension of the presented study by considering more sequences and the influence of other factors, such as coding impairments and the optical propagation characteristics of existing SMV displays.

REFERENCES