Quality of Experience in a Stereoscopic Multiview Environment

Felipe M. L. Ribeiro, Student Member, IEEE, José F. L. de Oliveira, Alexandre G. Ciancio, Eduardo A. B. da Silva, Senior Member, IEEE, Cássius R. D. Estrada, Student Member, IEEE, Luiz G. C. Tavares, Student Member, IEEE, Jonathan N. Gois, Student Member, IEEE, Amir Said, Fellow, IEEE, Marcela C. Martelotte

Abstract—In this paper we investigate how visualization factors, such as disparity, mobility, angular resolution and viewpoint interpolation, influence the Quality of Experience (QoE) in a stereoscopic multiview environment. In order to do so, we set up a dedicated testing room and conducted subjective experiments. We also developed a framework that emulates a super-multiview environment. This framework can be used to investigate and assess the effects of angular resolution and viewpoint interpolation on the quality of experience produced by multiview systems, and provide relevant cues as to how the baselines of cameras and interpolation strategies in such systems affect user experience. Aspects such as visual comfort, model fluidity, sense of immersion, and the 3D experience as a whole have been assessed for several test cases. Obtained results suggest that user experience in an motion parallax environment is not as critically influenced by configuration parameters such as disparity as initially thought. In addition, extensive subjective tests have indicated that while users are very sensitive to angular resolution in multiview 3D systems, this sensitivity tends not to be as critical when a user is performing a task that involves a great amount of interaction with the multiview content. These tests have also indicated that interpolating intermediate viewpoints can be effective in reducing the required view density without degrading the perceived QoE.

Index Terms—Multiview, motion parallax, stereoscopic image, quality of experience, visual perception, subjective evaluation.

I. INTRODUCTION

Nowadays one witnesses an increase in production and delivery of 3D content, as the demand for immersive content grows each day. However, the Quality of Experience (QoE) delivery by 3D is limited by the fact that it is based chiefly on stereopsis [1], [2], and it alone cannot provide true immersive experiences [3]. This is so because such systems disregard other important cues for the 3D perception. Among these, a very important one is motion parallax [4], whereby the scene seen depends on the viewer’s position. Motion parallax requires multiple views to produce a natural, glass-free experience [5], increasing the pressure on the acquisition, transmission, storage, coding and representation technologies [6].

Major developments occurred to cope with this pressure, including light field sensors and cameras arrays, new displays technologies, such as Super Multiview Displays, new representations, and coding paradigms [5], [6], [7]. These emerging technologies increased the necessity of correctly identifying and measuring the relevant factors that can enhance the QoE.

Although subjective and objective quality evaluation of 2D video have been widely examined [8], the assessment of 3D QoE is still a challenging problem [9], [10], [11]. The complex characteristics of the human visual system along with the diversity of possible interactions with such environments increases the difficulty in assessing their QoE [12], [13], [14]. Some quality metrics and subjective testing procedures have already been proposed to assess the quality of stereoscopic video [15], [16]. However, most of these metrics were typically oriented towards measuring 3D video quality in face of different compression strategies, instead of being concerned with the assessment of parameters related to the user sense of immersion or the 3D experience as a whole. For example, a reduced reference stereoscopic image quality metric proposed in [17] explores statistics from visual primitives of each view, assessed over different types of distortions. The same database was used in [13] to evaluate a no-reference image quality assessment metric which considers binocular visual perception and local structure distribution. Notable exceptions are the works in [12], [14], [16] and [19]. In [19], the authors conducted subjective experiments to evaluate the QoE in interactive 3D systems, but just focusing on the comparison between stereoscopic and multiview displays. A novel methodology to evaluate 3D subjective QoE, which measures the degree of visual discomfort of the viewers using multi-modal cues, is described in [12]. The authors evaluated factors such as viewing region and position, spatio-temporal complexity and depth of field, with interesting findings. This methodology is latter expanded, adding new modal cues, and further discussed at [14]. Lastly, Wang [10] investigates various aspects of human visual QoE when viewing stereoscopic 3D images/videos and to develop objective quality assessment models that automatically predict visual QoE of 3D images/videos. The author contributed with a new subjective 3D image quality assessment were the subjects evaluated different aspects of their 3D viewing experience, including the perception of 3D image quality, depth perception experience, visual comfort and the overall 3D experience.

In the literature, 3D interactive applications involving motion parallax have been addressed in works such as [20], [21]. However, these works were more concerned about investigating and improving depth perception than considering how configuration and visualization parameters influence the user perception of quality and immersion in a 3D environment.

In this paper we investigate how visualization factors affect...
the QoE in 3D applications with motion parallax. We analyze the visualization factors ‘disparity, mobility, angular resolution and viewpoint interpolation’ by conducting four experiments – one for each factor. We have chosen to deal chiefly with disparity and mobility as they are frequently employed to produce immersive environments [6]. Regarding angular resolution and viewpoint interpolation, the correct selection of these parameters is important for the content producer, particularly when real-life material is being recorded (as opposed to the generation of synthetic material), to cope with the current limitations on the number of views, transmission, storage, and coding. Thus, results such as the ones presented in this paper are relevant to the development of efficient coding solutions. They are aligned with current goals of MPEG and JPEG standardization bodies, that are launching initiatives such as the MPEG Ad Hoc Group on Light Field Compression [22] and the JPEG PLENO [23].

In the first three experiments, we used basic computer graphics models to assess parameters associated to visualization in the user’s perception of the experience. We believe that a good understanding of how the parameters degrade or enhance user experience will help to design more appealing applications, with improved user satisfaction. For the fourth experiment, realistic image scenes were used.

An initial investigation on such topics has been presented previously in works from the same authors [24], [25]. In [24] we investigated how parameters such as disparity, amount of parallax and monitor sizes influence the QoE in stereoscopic multiview environments. Following the conclusions found in [24] and in [25], we investigated the influence of the angular resolution of viewpoints with and without intermediate views interpolation. This paper extends the experiments from [24] and [25], that were carried out using simple computer graphics models, by also using “realistic” synthesized scenes as well as real scenes. In addition, it makes a thorough statistical analysis of the results that is not present neither in [24] nor in [25]. The results in [24] and [25] are reinterpreted using a robust statistical framework, supported by the new findings for both the real scenes and “realistic” synthesized views. The results obtained in the four experiments presented here can contribute to a better understanding of systems based on both multiview [5] or multi-lens (plenoptic) [23] arrays. It is important to point out that interactivity in our case means that, as the viewer moves, the observed scene changes. The way we achieve this is by using an infrared camera and reflective infrared patches or LEDs attached to the 3D glasses.

The paper is organized as follows: Section II describes the investigated scenarios adopted in the experiments. Sections III, IV, V and VI present, respectively, the methodology of each experiment, including the test conditions, equipment, procedures and statistical analysis of results. Finally, Section VII provides the conclusions of this work.

II. INVESTIGATED SCENARIOS

The scenarios adopted in the experiments assessed the effects of ‘disparity, amount of parallax, angular resolution and interpolated views’ on the users QoE. All the experiments were conducted in a 3D multiview system with motion parallax. In this section we briefly overview each of these scenarios. Implementation details are given in Sections III to VI.

A. Disparity

There is a myriad of depth cues that the human visual system uses to produce the perception of a 3-dimensional world. One of the most important is the stereopsis [4], which is currently used to produce image and video content with the illusion of depth. Natural stereopsis is produced by the differences in the images provided by the same scene due to the eyes’ different positions on the head. The depth is produced by the brain computing the disparity or, in other words, the distance between the projections of the same spatial point in the left and right images. Stereoscopic systems simulate this effect by providing different images for the left and the right eyes corresponding to distinct views of the same scene.

It’s known that the correspondence problem does not have a unique solution, and additional depth cues allow for an easier 3D reconstruction of the visual scene [26]. As a result, conflicts between different depth cues can occur in a stereoscopic system. Some examples are the accommodation–vergence rivalry, puppet theater effect, crosstalk and cardboard effect [13], [27], [28]. These effects can degrade the users’ QoE, producing visual discomfort and other physiological effects [12], [13], [14].

Concerned about the importance of the stereopsis produced by the disparity, we aimed to explore the influence of this parameter on the users’ QoE, in a stereoscopic multiview framework, using subjective quality assessment. While this experiment was previous presented in [24], this paper reintroduces the obtained results under a more rigorous statistical approach. This experiment is presented in Section III.

B. Motion Parallax Configurations

Motion parallax is another of the depth cues in human 3D perception resulting from motion. As we move, objects that are closer to us move farther across the field of view than the objects that are in distance, and sometimes occlusion or overlapping can occur. Motion parallax can be used to produce pseudo-3D effects on a legacy 2D display [21]. Multiview displays provide motion parallax in addition to vergence and binocular disparity. As the technology moves from the stereoscopic displays to the glasses free multiview displays, natural motion parallax becomes increasingly important.

Chen et. al. [29] addressed the effect of disparity in the context of static stereoscopic images, where the 3D QoE was assessed exploring binocular depth using multiple quality indicators. However, the authors did not assess the effect of movement arising from both a dynamic scene or motion parallax. Yano et. al. [30] conducted a study to evaluate the visual fatigue caused by HDTV and stereoscopic HDTV. This study concluded that scenes with a main object located in front of the scene and those containing large amounts of motion cause the most visual discomfort, since the limit of binocular fusion is also reduced with fast moving targets. This result
indicates the effects of model movement or of the motion parallax cannot be neglected in the QoE evaluation.

To investigate the effect of motion parallax in the users’ QoE, we devised a motion parallax system with 3DoF (vertical and horizontal parallax and zoom in/out) and tested different movement configurations. Four configurations were assessed: (i) the standard motion parallax environment, where an object’s point of view on the screen is changed based on its relative position to an observer being tracked; (ii) the denominated hyper parallax, which can be seen as an enhanced motion parallax, where small changes in the user position correspond to large changes in the object point of view (for instance, an object can be rotated 180° when the user sufficiently shifts his position to the side); (iii) hyper zoom, were a user can bring an object closer in 3D space, out of the screen, or take it farther into the screen with slight movements towards or away from the monitor; and (iv) without any motion parallax at all. Implementation details and results are given in Section V. This experiment is also found in [24], but without a thorough statistical analysis.

C. Angular Resolution

When it comes to realistic 3D immersive frameworks, only a limited number of viewpoints is available. As an observer moves from one viewpoint to another, the angular distance between adjacent viewpoints may significantly influence the user experience.

Takaki et al. [31] presents a study on the smoothness of parallax-induced motion provided by multiview displays, where the effect of motion-parallax discontinuity was investigated for multiple distances, number of views and display configurations using a multiview display. The problem of finding the minimum angular resolution required to provide natural motion-parallax on head-tracked multiview displays was approached by [32], [33], [34]. The authors of [20] and [21] addressed the effects of motion parallax in improving depth perception. However, these studies restricted their investigation to the fluidity of the parallax-induced motion or the depth perception, and did not address other aspects of the users’ experience, such as visual comfort or the 3D experience as a whole.

The number of views in our system is only limited by the tracking process resolution. As such, we can change the angular resolution by changing the tracking grid resolution. By reducing the tracking resolution, each viewpoint is associated to a view at the remaining tracking positions. Different tracking resolutions (number of viewpoints) were used to simulate a range of angular densities of the views. Implementation details and results are given in Section V. This experiment is also preliminarily discussed in [25], using only an informal statistical analysis.

D. Viewpoint Interpolation

Multiview applications with a reduced number of available views additionally face the problem of movement discontinuity. To create a realistic motion parallax effect in a 3D interactive scene with a restricted number of views, it is necessary to deal with the transitions and intermediate positions between two viewpoints [9], [10]. To address such effect, we extended our investigation to assess the problem of viewpoint interpolation. Strategies to handle such situation range from simply switching views at an intermediate physical position between available viewpoints, to using sophisticated depth-based interpolation techniques [9], [10]. This scenario is similar to the one described for Free Viewpoint Television [35].

Regarding intermediate view rendering, in order to create an effect that is equivalent to what is seen in commercially available autostereoscopic displays [36], we generated the intermediate views by weight-averaging the images from the two closest available views. This has been done by taking into account their distances to the view being rendered. Implementation details and results are presented in Section V. This experiment is an original contribution of this paper and is a natural continuation of the works in [24] and [25], extending the analysis to both real and “realistic” synthesized images.

III. EXPERIMENT 1: DISPARITY

In this experiment, we aimed to explore how the disparity affects the QoE of 3D images. Computer-graphics basic 3D models were generated and, for each model, 10 levels of disparity were produced. The method and the obtained results are detailed in this section.

A. Method

The assessment of the QoE in a motion parallax environment requires a special room, without any cluttering or obstruction, controlled illumination and a specific visualization system. The session occurs in a very similar manner to the one found in [12] and [14], with the subject relatively free to move around the room, except that the score evaluation occurs at the end of each sequence and the visualization system renders the view corresponding to the observer current viewpoint (physical position). This assertive assumes a tracking system as prerequisite, which finds the current observer position. Here we describe the room and equipment used in the experiment. We also explain the visualization and tracking systems.

1) Physical Setup:

- **Room**: The subjective tests were conducted in a dedicated room for 2D and 3D video quality evaluation, built in the Signals, Multimedia and Telecommunications Laboratory at the Universidade Federal do Rio de Janeiro. This room allowed users to interact with 3D models and watch 3D videos displayed in the monitors. During the tests, windows were covered with a black curtain to allow better illumination control;

- **Equipment**: The equipment used in the experiment included a 23.6” Acer GD235HZ and a 46” JVC 46D310U display, with active and circular polarization technologies, respectively. During the sessions, the distance from the evaluators to the screen was $3H_d$, in the case of the 46” display, and $2H_d$ for the 23.6” display – where $H_d$ is the display height. The 3D models were rendered using an
Visualization System: The framework required the development of a specific software to allow not only the visualization of 3D sequences, but also the introduction of a motion parallax and zoom in/out effect. This was achieved by tracking the user head position by attaching LEDs or IR reflective patches to the 3D glasses for both displays. The 3D views corresponding to the tracked user position was rendered in real time as the observer moved around the display. Also, a number of visualization parameters could be controlled, such as disparity, amount of parallax, tracking delay, interpolation between views, among others. The 3D model manipulation and rendering were implemented using the Open Scene Graph (v3.0.1) library [37].

Tracking system: The tracking strategy consisted in monitoring the images of LEDs or IR patches attached to user glasses captured with video cameras attached to the displays. To perform the tests, we used a Logitech Webcam Pro 9000 to track LEDs with the Acer display. We used also an OptiTrack V:120 Duo IR camera to track the IR patches with the JVC display. The sensor’s resolution for both cameras was 640 × 480 pixels, with a frame rate of 60 frames per second. The cameras’ horizontal and vertical angular resolutions (in pixels per degree (views/deg)) are defined, respectively, by

\[
\rho_H = \frac{W}{\text{fov}_H} \quad \text{and} \quad \rho_V = \frac{H}{\text{fov}_V},
\]

where \( W \) and \( H \) are respectively the horizontal and vertical resolutions of the camera sensor, expressed in pixels, and \( \text{fov}_H \) and \( \text{fov}_V \) are the horizontal and vertical fields of view of the camera, given by

\[
\text{fov}_H = 2 \arctan \left( \frac{w}{2f} \right) \quad \text{and} \quad \text{fov}_V = 2 \arctan \left( \frac{h}{2f} \right),
\]

where \( w, h, \) and \( f \) are the width, height and focal length of the camera, respectively.

<table>
<thead>
<tr>
<th>Position</th>
<th>Logitech 9000 Pro</th>
<th>OptiTrack V:120 Duo</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Horizontal</strong></td>
<td>64.1°</td>
<td>55.4°</td>
</tr>
<tr>
<td><strong>Vertical</strong></td>
<td>49.4°</td>
<td>43.0°</td>
</tr>
</tbody>
</table>

Table I shows the horizontal and vertical fields of view (fov) along with the angular resolutions (\( \rho \)), for both cameras. The angular resolution indicated corresponds to the maximum tracking resolution (640 × 480 pixels). At the observing distances used in the experiments, these resolution angles are equivalent to multiview systems with cameras separated approximately by 3 mm. This is in accord with the Super Multiview Video (SMV) condition [38], which states that the distance between views should be smaller than the diameter of a human pupil, which varies from 2 mm to 4 mm.

2) Observers: A team of 15 naive observers, composed by undergraduate students with mixed backgrounds, from both sexes, and with ages ranging from 18 to 24 years, participated in the subjective tests. A screening setup procedure was conducted to test the observers’ visual acuity, composed of stereo and color blindness tests. The stereo blindness test consisted of a set of four anaglyph images with different content presented to the observers, who were inquired about the perceived depth of the scenes. To test for color blindness, observers were screened using the Ishihara Color test plates [39].

3) Models (Stimuli): Four synthetic 3D models were generated for the experiment. They are referred to as computer-graphics ‘basic models’. All these models were designed with an associated task that users should perform while interacting with them. Models and task descriptions are presented in Table II. Models “Earth”, and “Game” are dynamic, with translation and rotation movement. While the remaining models are “static”, their viewpoint dynamically changes considering motion parallax and the observer position. A representative picture of each model can be seen on Figure 1.

4) Experimental Design and Procedure: Ten different disparity values were tested. They were specified by the visualization software and normalized by the mean interocular distance. The normalized values were: \{0.1, 0.3, 0.5, 0.8, 0.9, 1.0, 1.2, 1.3, 2.0, 2.5\}. The disparity value ‘1’ corresponds to the average interocular distance of 6.5 cm. As the disparities values increase, the observer perceives the objects of the scene further away from the display plane.

The aspects of the QoE were assessed by the influence of disparity in task execution and the impacts it causes in the observer overall impression of the 3D experience. Each disparity value was assessed for the two displays (23.6° and 46°), totaling 20 test sessions. The tests were performed for all combinations of the four models, in a total of 80 tests, and each observer assessed the 80 tests.
To evaluate the QoE aspects, after completing the task the observers were asked to answer a task-related question, using a discrete unlabeled scale from 1 (bad) to 5 (excellent). The QoE aspects investigated are represented by items Q1 to Q4:

- **Q1 Visual comfort**: Visual comfort refers to symptoms such as eyes tiredness, headache, nausea and dizziness; the higher the grade, the greater the comfort.
- **Q2 Sense of immersion**: The sense of immersion refers to the sensation of immersion with the environment; the higher the grade, the higher the sense of immersion.
- **Q3 Difficulty to complete the task**: Difficulty to execute the given task, as described in Table II; the higher the grades the smaller the difficulty.
- **Q4 Experience as a whole**: Overall experience of the performed test; the higher the grade, the better the experience.

These QoE aspects are similar to the perceptual dimensions described on the ITU-BT.2021 Recommendation [40], the ones proposed in [41] for 3D QoE assessment, and the ones employed in [16] during its subjective test sessions. However, while in [16] each aspect was evaluated individually during multiple sessions, during this work each aspect was evaluated at the end of the task or sequence.

### B. Analysis Procedure

To investigate the influence of the disparity on QoE, assessed in the subjective tests, we performed an Analysis of Variance (ANOVA) [42] for each QoE aspect listed in Section [11-A4]. They are: visual comfort, sense of immersion, difficulty to complete the task, and experience as a whole. In this study the significance level was set to .05. The significance level, in a statistical test, is the probability of rejecting the null hypothesis despite it being actually true.

In our analysis, the ANOVA F-test (right-tailed) was followed up with a post hoc test – the Duncan’s multiple range test, in order to determine which levels of the factor were statistically different from the others. This procedure consists of a statistical test that compares differences of means between each pair of the factors’ levels, starting by the difference between the smallest and the largest means. It continues until all pairs of means are compared. If any difference is greater than a critical value, defined by the test, the levels of the pair are considered as significantly different [43]. Performing a post hoc test is necessary if the factor has three or more levels. As a result, this test splits the factor’s levels into statistically significant different subsets. In other words, in the subjective testing scenario conducted in this paper, this is equivalent to saying that, on average, users consider the levels of different subsets to be perceptually different. More details on post hoc tests can be found in [43].

In this first experiment, we were interested in the effect of 10 different disparities levels on the aspects of QoE. We also wanted to investigate the effect of the display type, that can be the 23.6” active display or the 46” display with circular polarization. In order to carry this out, we performed four ANOVAs where each QoE aspect was a response variable and with both ‘disparity’ and ‘display’ as factors. Observers were controlled through blocking [43]. Besides this, if both factors showed statistically significant effects, we tested the interaction between them. An interaction between factors means that the factors are not independent. When it occurs, we must examine the levels of one factor together with the levels of the other factor to understand their impact on the response variable. After performing the ANOVA, we applied the Duncan’s multiple range test to the factor ‘disparity’ in order to determine which levels differ from the others.

### C. Results

To present the impact of the different disparities on ‘visual comfort’, ‘sense of immersion’, ‘difficulty to complete the task’, and ‘experience as a whole’, we summarized the post hoc test results on Table III. Each of the subsets represents a range of disparity values (shown between brackets) that were considered statistically equivalent by the Duncan post hoc. Average (avg.) subjective grades for each subset are also given, along with their p-values (p). It is important to notice that these p-values are all greater than .05, indicating that the null hypothesis (all the group means are equal) cannot be rejected. It means that the levels of disparity that belong to the same subset have means that are not statistically different, according to the post hoc test. In other words, these values of disparities affect the response variable equivalently. Complete results for the display tests were made available in [44].

Comments about the four QoE aspects are as follows:

- **Visual comfort**: The F-test results have shown that there was a significant effect of factors disparity (p-value < .001) and display (p-value < .001) on visual comfort. From Table III, we see that the best subjective scores for visual comfort occur for disparity values below or equal to 0.8. The disparity value that resulted in the worst average grade was 2.5. Mean subjective grades obtained for each display indicated that 23” display is better in terms of visual comfort (on average) than the 46” one. Besides this, there is interaction between ‘disparity’ and ‘display’ (p-value = .007), meaning that the display size affects the user’s visual comfort. This interaction is particularly strong for the two largest disparities (2.0 and 6.0).
The obtained results show that despite the fact that users cannot distinguish between closer disparities, they play an important role in the 3D perception, since the number of subsets is reasonably well spread over the range of disparity values. The average grade values also indicate that observers tend to assign inferior grades to larger disparity values, specially those above the average interocular distance. This suggests that systems built around lower disparity values provide the users a better experience, and models with higher disparities, with objects that are very close to the observer, should be avoided.

IV. EXPERIMENT 2: MOBILITY ASSESSMENT

In the second experiment, we investigated the effect of different movement configurations in the perception of the QoE in a parallax-based 3D environment. Four parallax configurations were assessed.

A. Method

1) Physical Setup: Room, equipment, visualization system and tracking system were the same as in Experiment 1.

2) Observers: The observers were the same as in Experiment 1.

3) Models (Stimuli): The models were the same as in Experiment 1 (computer-graphics basic 3D models).

4) Experimental Design and Procedure: Four different motion-parallax configurations were evaluated during this experiment:

- hyper parallax, hyper zoom and eye tracking turned on;
- no hyper zoom, just with hyper parallax and eye tracking;
- eye tracking only; and,
- no parallax or eye tracking.

Each configuration was evaluated for both displays (23″ and 46″), resulting in 8 test sessions. These tests were performed for all basic models, comprising 32 tests. The disparity value was fixed to 0.5 during all sessions.

When the hyper-parallax was turned on, the visualization system computed the rotation angle of the models, in radians, as

\[ \theta = \frac{4.36u_y}{\sqrt{1 + 11.6u_y^2}}; \quad \phi = 4.36u_x; \]  

(3)

where \( \theta \) is the rotation angle relative to the horizontal axis of the screen, \( \phi \) is the rotation angle around its vertical axis, and \( u_y \) and \( u_x \) are the coordinates of the unit norm vector corresponding to the direction of the vector connecting the detected LEDs or IR patches to the center of the screen. The hyper-zoom scaling factor was computed by the formula:

\[ S_z = 0.3 + 0.84\frac{d_{ref}}{n_z}, \]  

(4)

where \( d_{ref} \) is the diagonal length of the screen and \( n_z \) is the distance from the point between the LEDs or IR patches to the screen center.
As in the previous experiment, after completing the task, observers were asked to answer a task-related question, as well as use a discrete unlabeled scale from 1 (bad) to 5 (excellent) to assess the QoE aspects, represented by items Q1 to Q4, as explained in Subsection III-A3. These aspects were: (Q1) visual comfort, (Q2) sense of immersion, (Q3) difficult to complete the task, and (Q4) experience as a whole.

B. Analysis Procedure

To assess the impact of the different parallax configurations on the quality of experience (QoE), we performed the ANOVA, where ‘mobility assessment’ was a factor, with four levels. These levels were: hyper parallax and hyper zoom turned on (HPZ); only hyper parallax turned on (HP); standard eye tracking only (P); eye tracking turned off (No Tracking – NP), which corresponded to a static 3D model being displayed on the screen.

As we intended to test the effect of the two different displays, ‘display’ was also considered a factor. Each aspect of QoE (visual comfort, sense of immersion, difficult to complete task, experience as a whole) was a response variable in the ANOVA. Consequently, four ANOVAs were performed. As in the previous experiment, observers were controlled through blocking [43]. The ANOVAs were followed by the post hoc multiple comparison test (Duncan’s multiple range test), as detailed in Subsection III-B.

C. Results

As in Subsection III-B, the ANOVA F-test results for factor ‘mobility assessment’, with all the response variables, are given along the text, and the summarized post hoc test results are presented in Table IV. Complete results for the display can be found in [44]. Comments about each QoE aspect are as follows:

- **Visual comfort:** There is a significant effect of ‘mobility assessment’ on visual comfort (p-value < .001). On Table IV, we can see that the case with static models (NP) had worst result than all the others mobility scenarios. And all mobility scenarios are considered equivalent for the visual comfort aspect, since all three configuration groups are in the same subset on Table IV. In this analysis there was no significant effect of the display size on the visual comfort (p-value = .723).

- **Sense of immersion:** As well as in the ‘visual comfort’ aspect, there is a significant effect of mobility (p-value < .001), and no statistical effect of the display (p-value = .820).

Analyzing Table IV we notice a clear increase in the QoE as we move from a no-tracking situation to a higher degree of mobility (hyper zoom and hyper parallax), with hyper zoom adding less to the QoE (since, for some factors, it can be considered statistically equivalent to the ‘parallax only’ case). This is an interesting result, since designers may opt to introduce parallax, a typically desirable feature, into their systems without compromising visual comfort, with an optional inclusion of hyper zoom.

- **Difficulty to complete the task:** The results are similar to the ones found in ‘sense of immersion’.

- **Experience as a whole:** The results are also similar to the ones for ‘sense of immersion’, indicating that the mobility scenarios hyper parallax (HP) and hyper parallax plus hyper zoom (HPZ) are not significantly different. Regarding the display factor, it showed no significant effect on the experience (p-value = .366), as well as in the other QoE aspects.

From these results we can infer that, while an increase in mobility produces only small gains in QoE, there is a clear increase in QoE when the user experiences motion parallax through head tracking, compared with a no-tracking situation.

V. EXPERIMENT 3: ANGULAR RESOLUTION

In the third experiment, we investigated how a limitation of the number of available viewpoints in a multi-camera, immersive 3D application, influences the user perception of the QoE.

A. Method

1) **Physical Setup:** Room, equipment, visualization system and tracking system were the same as in Experiments 1 and 2.

2) **Observers:** The observers were the same as in Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Assessments</th>
<th>Subsets</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visual Comfort</strong></td>
<td>HP, HPZ, P</td>
<td>NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(avg.=3.91)</td>
<td>(avg.=3.44)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p=.270)</td>
<td>(p=1.000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sense of Immersion</strong></td>
<td>HP, HPZ</td>
<td>P</td>
<td>NP</td>
<td></td>
</tr>
<tr>
<td>(avg.=4.07)</td>
<td>(avg.=3.60)</td>
<td>(avg.=2.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p=.137)</td>
<td>(p=1.000)</td>
<td>(p=1.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Difficulty to Complete Task</strong></td>
<td>HP, HPZ</td>
<td>P</td>
<td>NP</td>
<td></td>
</tr>
<tr>
<td>(avg.=4.09)</td>
<td>(avg.=3.52)</td>
<td>(avg.=2.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p=.133)</td>
<td>(p=1.000)</td>
<td>(p=1.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experience</strong></td>
<td>HP, HPZ</td>
<td>P</td>
<td>NP</td>
<td></td>
</tr>
<tr>
<td>(avg.=4.03)</td>
<td>(avg.=3.69)</td>
<td>(avg.=3.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p=.078)</td>
<td>(p=1.000)</td>
<td>(p=1.000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table IV: Mobility assessment results of the Post hoc test (Mobility configuration groups: No Tracking (NP), Standard Tracking (P), Hyper Parallax Only (HP), Hyper Parallax + Hyper Zoom (HPZ)) (average and p-value in parentheses).
3) Models (Stimuli): The models were the same as in Experiments 1 and 2.

4) Experimental Design and Procedure: In this experiment, no interpolation was used to generate the intermediate viewpoints associated with physical positions between available views (interpolation effects are assessed in Experiment 4). The views from these viewpoints were generated by simply replicating the closest available view.

To reduce the influence of the disparity on the grades, three disparity values were considered: (0.3, 0.5 and 0.8 of the interocular distance). Then, we assessed the effect of a reduction in the angular resolution of the system, which corresponds to reducing the number of cameras in a multi-camera system, or the number of lenses in a multi-lens plenoptic camera.

Observers were asked to assess the QoE for each angular resolution configuration shown in Table V, considering the three disparity values for the two displays (23" and 46"), resulting in 36 sessions. Performing this for each basic model, this experiment produced a total of 144 tests.

In this experiment, two new aspects of the QoE were considered: ‘model fluidity level’ and ‘comfort related to the model fluidity level’. The six QoE aspects assessed by the observers, using a scale from 1 (bad) to 5 (excellent), are listed below:

- (Q1) Visual comfort;
- (Q2) Sense of immersion;
- (Q3) Difficulty to complete the task;
- (Q4) Model fluidity level: Model perceived smoothness due to motion parallax;
- (Q5) Comfort related to the model fluidity level: Comfort associated to discontinuities due to viewpoint switching;
- (Q6) Experience as a whole.

B. Analysis Procedure

As in the previous experiments, we performed an ANOVA for each response variable related to the QoE aspects (the variables are listed in Subsection V-A4). The ‘angular resolution’ was a factor, with six levels, as described in Table V. ‘Display’ was also a factor, and we tested the statistical interaction between these factors when both showed a statistically significant effect [43]. We found no statistical interaction between the disparity values and the assessed factors. Mostly likely, this is due the chosen disparity values being below the average interocular distance. As described in Section III, the experiments confirms this fact, since these values are in a statistically equivalent subset. Also, given that the disparity is a binocular cue while angular resolution is related with motion parallax cues, it is reasonable to assume no interaction between them. Observers were controlled through blocking. After this, we applied the Duncan’s multiple range test on the ‘angular resolution’ factor.

C. Results

As before, ANOVA F-test results are given along the text. Table VI summarizes the results obtained in the Duncan’s post hoc test for the statistical effect of a reduction in the angular resolution of the system.

- **Visual comfort:** Both factors, ‘angular resolution’ and ‘display’ have shown statistically significant effect, meaning that the display size affects the user assessment of quality for different angular resolutions. As expected, higher angular resolutions were better evaluated by users. The best ones were in the range [5.88, 11.1] views/deg. The 23" display had better average than the 46" one (4.247 and 4.143, respectively). There was no significant statistical interaction between the factors (p-value = .689), which means that 23" display was always better than the 46".

- **Sense of immersion:** There was no statistical effect of factor ‘display’ (p-value = .918), but factor ‘angular resolution’ has shown a statistically significant effect (p-value < 0.001). The best angular resolutions values were also in the range [5.88, 11.1] views/deg.

- **Difficulty to complete the task:** As in the case of ‘sense of immersion’, there was no statistically significant effect of factor ‘display’ (p-value = .073). However, there is statistical effect of factor ‘angular resolution’ (p-value = .002). In other words, although display size did not influence users on the completion of the task, they were affected by the angular resolution. Here, the three greatest angular resolutions showed no difference, in average. This means that the observers had the same difficulty level to complete the task for these three resolutions.

- **Model fluidity and Model fluidity comfort:** For these two aspects, there was a statistically significant effect of the factor ‘Display’ (p-value = .004 and p-value < .001, respectively), with the 23” display being the one with better average subjective grade. The statistical effects of angular resolution on subjective assessment of ‘Model fluidity’ and ‘Model fluidity comfort’ were also significant (p-values < .001). Results show that the observers could perceive difference between the three finer angular resolutions, which indicates a high sensitivity even to small jumps in continuity.

- **Experience as a whole:** As in the previous aspect result, there was statistical effect of factor ‘Display’ (p-value < .001), where 23” display showed a better average user assessment. Factor ‘angular resolution’ showed significant effect (p-value < 0.001), but in this case the observers couldn’t discern any difference between the two finer angular resolutions [5.88, 11.1] views/deg, meaning that these two resolutions provide the same sense of experience as a whole to the user.

Analyzing the results for all QoE aspects we conclude that...
it is clear, as expected, the perceived discontinuities produced by the limited number of views in coarser angular resolutions resulted in bad to regular assessments, more noticeably regarding visual comfort.

In terms of fluidity, it is interesting to notice that no saturation point could be observed in terms of angular resolution. This means that users can still perceive improvements in the model fluidity due to their movement for angular resolutions as high as 11.1 views/deg. This result is in agreement with the ones obtained in [33] and [34]. There, the required angular resolution for smooth motion parallax in a stereoscopic head-tracked environment were assessed as 12.0 and 14.7 views/deg. respectively. It is important to point out, however, that although improvements can be obtained with resolutions higher than 11.1 views/deg, resolutions of 5.88 views/deg already showed average subjective grades above 4 out of 5.

These results also indicate that the grades of aspects such as visual comfort and the 3D experience as whole were not compromised even in cases where the fluidity of the parallax-induced motion received lower grades.

VI. EXPERIMENT 4: VIEWPOINT INTERPOLATION

In this experiment we investigated the effect of interpolating the intermediate views in an environment with a reduced number of available cameras in a dense multi-camera scenario. The experiments were conducted with realistic scene models.

A. Method

1) Physical Setup: Room, equipment, visualization system and tracking system were the same as in the previous Experiments. However, only the 46” display was used.

2) Observers: A group of 20 naive observers, undergraduate students with mixed backgrounds from both sexes and with ages ranging from 18 to 30 years, participated in the subjective tests.

3) Models (Stimuli): For this experiment, we emulated a more realistic multiview setup, using three realistic static scenes instead of simpler computer graphics basic models. These static scenes have only horizontal parallax, and are shown in Figure 2. The Elephant and Train are light field scenes acquired using an eight-camera array and a linear translating gantry, with 3D horizontal parallax only, and are freely available at [45]. The San Miguel scene was synthetically generated using Physically Based Ray Tracing (PBRT) [46].

![Elephant](a) Elephant ![San Miguel](b) San Miguel ![Train](c) Train

Figure 2. 3D realistic static scenes.

To produce the horizontal parallax, the scenes were composed of multiview arrays that have approximately the same original angular resolution (without interpolation). The angular resolution of the multiview arrays has been computed by manually marking the points with smaller depths in the leftmost and rightmost views, computing the angle between them as measured at the viewer’s position and dividing by the number of views. The procedure conducted to test the observers’ visual acuity was the same as in the previous experiments, described in Subsection III-A2.

Table VI

<table>
<thead>
<tr>
<th>Assessments</th>
<th>Subsets</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Comfort</td>
<td>[5.88, 11.1]</td>
<td>[2.86]</td>
<td>[2.86]</td>
<td>[2.86]</td>
<td>[0.58]</td>
<td>[0.58]</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.36)</td>
<td>(avg.=4.23)</td>
<td>(avg.=4.28)</td>
<td>(avg.=4.14)</td>
<td>(avg.=4.15)</td>
<td>(avg.=4.15)</td>
</tr>
<tr>
<td></td>
<td>(p=0.146)</td>
<td>(p=1.000)</td>
<td>(p=1.000)</td>
<td>(p=0.146)</td>
<td>(p=0.146)</td>
<td>(p=0.146)</td>
</tr>
<tr>
<td>Sense of Immersion</td>
<td>[5.88, 11.1]</td>
<td>[2.86, 5.88]</td>
<td>[1.15, 2.86]</td>
<td>[1.15, 2.86]</td>
<td>[1.15, 2.86]</td>
<td>[1.15, 2.86]</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.25)</td>
<td>(avg.=4.19)</td>
<td>(avg.=4.40)</td>
<td>(avg.=4.40)</td>
<td>(avg.=4.40)</td>
<td>(avg.=4.40)</td>
</tr>
<tr>
<td></td>
<td>(p=0.10)</td>
<td>(p=0.10)</td>
<td>(p=0.10)</td>
<td>(p=0.10)</td>
<td>(p=0.10)</td>
<td>(p=0.10)</td>
</tr>
<tr>
<td>Difficulty to Complete Task</td>
<td>[2.86, 11.1]</td>
<td>[1.15, 5.88]</td>
<td>[2.86, 4.28]</td>
<td>[0.58, 1.45]</td>
<td>[0.58, 1.45]</td>
<td>[0.58, 1.45]</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.33)</td>
<td>(avg.=4.28)</td>
<td>(avg.=4.28)</td>
<td>(avg.=4.22)</td>
<td>(avg.=4.22)</td>
<td>(avg.=4.22)</td>
</tr>
<tr>
<td></td>
<td>(p=0.11)</td>
<td>(p=0.11)</td>
<td>(p=0.11)</td>
<td>(p=0.11)</td>
<td>(p=0.11)</td>
<td>(p=0.11)</td>
</tr>
<tr>
<td>Model Fluidity</td>
<td>[11.1]</td>
<td>[5.88]</td>
<td>[2.86]</td>
<td>[1.15, 1.45]</td>
<td>[1.15, 1.45]</td>
<td>[1.15, 1.45]</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.25)</td>
<td>(avg.=3.96)</td>
<td>(avg.=3.74)</td>
<td>(avg.=3.45)</td>
<td>(avg.=3.45)</td>
<td>(avg.=3.45)</td>
</tr>
<tr>
<td></td>
<td>(p=0.00)</td>
<td>(p=1.00)</td>
<td>(p=1.00)</td>
<td>(p=0.00)</td>
<td>(p=0.00)</td>
<td>(p=0.00)</td>
</tr>
<tr>
<td>Model Fluidity Comfort</td>
<td>[11.1]</td>
<td>[5.88]</td>
<td>[2.86]</td>
<td>[1.15, 1.45]</td>
<td>[1.15, 1.45]</td>
<td>[1.15, 1.45]</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.29)</td>
<td>(avg.=4.06)</td>
<td>(avg.=3.85)</td>
<td>(avg.=3.58)</td>
<td>(avg.=3.58)</td>
<td>(avg.=3.58)</td>
</tr>
<tr>
<td></td>
<td>(p=0.58)</td>
<td>(p=1.00)</td>
<td>(p=1.00)</td>
<td>(p=1.50)</td>
<td>(p=1.50)</td>
<td>(p=1.50)</td>
</tr>
<tr>
<td>Experience</td>
<td>[5.88, 11.1]</td>
<td>[2.86]</td>
<td>[1.15, 1.45]</td>
<td>[0.58, 1.45]</td>
<td>[0.58, 1.45]</td>
<td>[0.58, 1.45]</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.24)</td>
<td>(avg.=4.07)</td>
<td>(avg.=3.93)</td>
<td>(avg.=3.87)</td>
<td>(avg.=3.87)</td>
<td>(avg.=3.87)</td>
</tr>
<tr>
<td></td>
<td>(p=0.00)</td>
<td>(p=0.00)</td>
<td>(p=0.00)</td>
<td>(p=0.00)</td>
<td>(p=0.00)</td>
<td>(p=0.00)</td>
</tr>
</tbody>
</table>

Table VI: Angular resolution (views/deg) results of the post hoc test (angular resolution (horizontal × vertical) groups: 0.58 × 0.56, 1.15 × 1.11, 1.45 × 1.39, 2.86 × 2.78, 5.88 × 5.56, 11.1 × 11.1. In the table, only the horizontal resolution is indicated (average and p-value in parentheses).
of views. The models used, along with their number of views and angular resolutions are shown in Table VII. The models with more than 425 views have been truncated to 425 views by discarding the first and last \((N - 425)/2\) views, where \(N\) is the number of views. Also, due to constraints in the visualization system, all images were scaled to 640 \(\times\) 360 pixels. The tracking resolution in each session was equal to the angular resolution.

4) Experimental Design and Procedure: In this experiment we investigated the effect of the number of views and intermediate viewpoint interpolation strategy in a dense multiview array (light-field).

The subject experimental considered three angular resolution/interpolation configurations:

- Full Angular Resolution: The image array contained the total of 425 images;
- Zero-\(q\)-interpolation: Every one out of \(q\) images was kept. The intermediate \(q - 1\) were discarded and replaced by their closest neighbor;
- Alpha-blending \(q\)-interpolation: The intermediate \(q - 1\) images were discarded and each one replaced by their neighbors alpha-blending associated with its position.

For a given intermediate image \(I_n, n \in [1, q - 1]\), the alpha-blending interpolation produces:

\[
I_n = (\alpha_n - 1)I_l + \alpha_n I_r,
\]

where \(I_l\) and \(I_r\) are, respectively, the left and right neighbor, and \(\alpha_n\) is

\[
\alpha_n = \frac{n}{q}.
\]

To increase the number of tests, this experiment was conducted using three disparity values: \(0.4, 0.8, 1.0\). After interacting with the models, the observers were asked to grade the same aspects of QoE as specified in Section V-A4 with the exception of the task-related question. These aspects are: (Q1) Visual comfort; (Q2) Sense of immersion; (Q3) Model fluidity level; (Q4) Comfort related to the model fluidity level; (Q5) Experience as a whole.

The different tracking resolutions, used to simulate a range of angular densities, can be seen in Table VIII.

The impact of these angular resolutions was assessed by the evaluators using the interpolation strategies previously described (zero-\(q\) and alpha-blending) for \(q \in \{2, 4, 8, 12, 16\}\), totaling 11 test cases for each disparity value and model (1 non-interpolated case, 5 zero-\(q\) and 5 alpha-blend interpolation configurations). All the tests in this set were conducted using the 46" (JVC) display and with eye tracking to provide horizontal parallax.

### Table VII

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Number of Views</th>
<th>Angular Resolution [views/deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Miguel</td>
<td>425</td>
<td>5.88</td>
</tr>
<tr>
<td>Train</td>
<td>5000</td>
<td>0.26</td>
</tr>
<tr>
<td>Elephant</td>
<td>460</td>
<td>6.25</td>
</tr>
</tbody>
</table>

### Table VIII

<table>
<thead>
<tr>
<th>Interpolation Factor ((q))</th>
<th>Angular Resolution [views/deg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.88</td>
</tr>
<tr>
<td>2</td>
<td>2.94</td>
</tr>
<tr>
<td>4</td>
<td>1.49</td>
</tr>
<tr>
<td>8</td>
<td>0.74</td>
</tr>
<tr>
<td>12</td>
<td>0.50</td>
</tr>
<tr>
<td>16</td>
<td>0.37</td>
</tr>
</tbody>
</table>

B. Analysis Procedure

As in the other experiments, we used the ANOVA to assess the results. In this experiment, the factor is the ‘viewpoint interpolation’, as described in Table VIII and the observers were controlled through blocking, as before. Again, we found no statistical interaction between the disparity values and the assessed factors. The response variables are each one of the five QoE aspects, listed in Subsection VI-A4. They are: Visual comfort, Sense of immersion, Model fluidity level, Comfort related to the model fluidity level, and Experience as a whole. The ANOVA F-test for each variable was followed by the post hoc test (Duncan’s multiple range test). As mentioned before, all the angular resolution and interpolation configurations inside a subset are, on average, statistically equivalent.

C. Results

The results for the ANOVA F-test show that there is statistically significant effect of factor ‘viewpoint interpolation’ for all response variables, i.e., for all QoE aspects \((p\text{-values } < .001)\). The post hoc test results are presented in Table IX.

It is interesting to observe that the overall results suggest that while users are indeed sensitive to and influenced by the number of views, downsampling up to four times the maximum number of available views (decreasing up to four times the angular resolution) either gives users the same impression or achieves a very close perceptual impression to the one of the original multiview array. In addition, in most of the cases when this is not true, the use of interpolation is effective in providing a perceptual impression equivalent to the one of the original multiview array. Therefore, systems with resolutions higher than 1.49 views/deg can be perceived as being continuous viewpoint systems provided that proper processing is carried out. This result is also in agreement with [33] and [34], and with the recommendation found in [47] for SMV systems.

Also, an important conclusion can additionally be drawn from the subsets obtained in the post hoc test: as a multi-camera system moves towards further reducing the number of available cameras (coarser angular resolution), typical interpolation strategies (such as the alpha-blending) have a negative effect in the user perception. This may be attributed to two reasons: first, since such strategies show “average” images on the intermediate viewpoints, the resulting image seen by a user has a blurring effect, reducing its quality; and second, in a non-interpolation scenario for large camera gaps users are subject to a much larger skew-distortion effect, which gives them the impression of 3D movement as they move.
Table IX

Viewpoint interpolation results of the post hoc test (Groups: Original sequence (O) (maximum angular resolution of 5.88 views/deg) along with Frame Repetition (R) and Frame Interpolation (I) for resolutions 2.94 views/deg (q=2), 1.49 views/deg (q=4), 0.74 views/deg (q=8), 0.50 views/deg (q=12) and 0.37 views/deg (q=16) (average and p-value in parentheses).

<table>
<thead>
<tr>
<th>Assessments</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visual Comfort</strong></td>
<td>R:[5.88],</td>
<td>R:[0.74, 2.94],</td>
<td>R:[0.37, 1.49],</td>
<td>R:[0.50],</td>
<td>I:[0.37, 0.50]</td>
<td>I:[0.37, 0.50]</td>
<td>I:[0.37, 0.50]</td>
</tr>
<tr>
<td></td>
<td>[0.37],</td>
<td>[0.37],</td>
<td>[0.74],</td>
<td>[0.74],</td>
<td>(avg.=3.29)</td>
<td>(avg.=3.11)</td>
<td>(avg.=3.11)</td>
</tr>
<tr>
<td></td>
<td>I:[2.94]</td>
<td>I:[1.49]</td>
<td>I:[1.49]</td>
<td>I:[0.74]</td>
<td>(p=1.13)</td>
<td>(p=0.065)</td>
<td>(p=0.065)</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.15)</td>
<td>(avg.=3.99)</td>
<td>(avg.=3.98)</td>
<td>(avg.=3.78)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p=0.171)</td>
<td>(p=.072)</td>
<td>(p=.055)</td>
<td>(p=.084)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sense of</strong></td>
<td>O:[5.88],</td>
<td>R:[0.74, 2.94],</td>
<td>R:[0.37, 0.74],</td>
<td>R:[0.37, 0.74],</td>
<td>I:[0.37, 0.50]</td>
<td>I:[0.37, 0.50]</td>
<td>I:[0.37, 0.50]</td>
</tr>
<tr>
<td><strong>Immersion</strong></td>
<td>R:[1.49],</td>
<td>I:[1.49]</td>
<td>I:[0.74]</td>
<td>I:[0.74]</td>
<td>(avg.=3.75)</td>
<td>(avg.=3.75)</td>
<td>(avg.=3.75)</td>
</tr>
<tr>
<td></td>
<td>2.94],</td>
<td>I:[1.49]</td>
<td>I:[1.49]</td>
<td>I:[0.74]</td>
<td>(p=1.18)</td>
<td>(p=0.065)</td>
<td>(p=0.065)</td>
</tr>
<tr>
<td></td>
<td>[1.49],</td>
<td>[0.37]</td>
<td>[0.74]</td>
<td>[0.74]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.94]</td>
<td>I:[1.49]</td>
<td>I:[1.49]</td>
<td>I:[0.74]</td>
<td>(p=0.32)</td>
<td>(p=1.11)</td>
<td>(p=1.11)</td>
</tr>
<tr>
<td></td>
<td>(avg.=4.10)</td>
<td>(avg.=3.98)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(p=0.080)</td>
<td>(p=.072)</td>
<td>(p=.032)</td>
<td>(p=.032)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fluidity</strong></td>
<td>O:[5.88],</td>
<td>R:[2.94],</td>
<td>R:[0.74],</td>
<td>R:[0.37],</td>
<td>R:[0.37],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
</tr>
<tr>
<td></td>
<td>2.94],</td>
<td>R:[0.74],</td>
<td>I:[1.49],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
</tr>
<tr>
<td></td>
<td>[1.49],</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
</tr>
<tr>
<td></td>
<td>2.94]</td>
<td>[0.37]</td>
<td>[0.74]</td>
<td>[0.74]</td>
<td>[0.74]</td>
<td>[0.74]</td>
<td>[0.74]</td>
</tr>
<tr>
<td></td>
<td>(avg.=3.97)</td>
<td>(avg.=3.86)</td>
<td>(avg.=3.86)</td>
<td>(avg.=3.86)</td>
<td>(avg.=3.86)</td>
<td>(avg.=3.86)</td>
<td>(avg.=3.86)</td>
</tr>
<tr>
<td></td>
<td>(p=0.466)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
</tr>
<tr>
<td></td>
<td>(p=0.466)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
<td>(p=.141)</td>
</tr>
<tr>
<td><strong>Fluidity</strong></td>
<td>O:[5.88],</td>
<td>R:[1.49],</td>
<td>R:[0.74],</td>
<td>R:[0.37],</td>
<td>R:[0.37],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
</tr>
<tr>
<td><strong>Comfort</strong></td>
<td>2.94],</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
</tr>
<tr>
<td></td>
<td>[1.49],</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
</tr>
<tr>
<td></td>
<td>2.94]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
</tr>
<tr>
<td></td>
<td>(avg.=3.97)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
</tr>
<tr>
<td></td>
<td>(p=0.24)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
</tr>
<tr>
<td></td>
<td>(p=0.24)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
<td>(p=.096)</td>
</tr>
<tr>
<td><strong>Experience</strong></td>
<td>O:[5.88],</td>
<td>R:[1.49],</td>
<td>R:[0.37],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
<td>I:[0.37],</td>
</tr>
<tr>
<td></td>
<td>2.94],</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
<td>2.94]</td>
</tr>
<tr>
<td></td>
<td>[1.49],</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
</tr>
<tr>
<td></td>
<td>2.94]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
<td>[1.49]</td>
</tr>
<tr>
<td></td>
<td>(avg.=3.97)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
<td>(avg.=3.84)</td>
</tr>
<tr>
<td></td>
<td>(p=0.57)</td>
<td>(p=.092)</td>
<td>(p=.092)</td>
<td>(p=.092)</td>
<td>(p=.092)</td>
<td>(p=.092)</td>
<td>(p=.092)</td>
</tr>
</tbody>
</table>
sideways even though the image they are seeing is the same. For example, users may perceive, in some cases, the non-interpolated (zero-hold) 0.37 views/deg configuration as being better than other finer resolutions. Overall, for large steps, the impression caused by the skew-distortion effect seems to be more pleasant to users than observing an interpolated, blurred image. These results allow us to conclude that interpolation techniques only improve the user experience above a certain angular resolution (around 1.49 views/deg). However, it is possible that more sophisticated techniques of interpolation, such as Depth Image Based Rendering [9], [10], [48] could circumvent this limitation.

Specifically regarding visual comfort, the large overlap of subsets, particularly for finer resolutions, make it clear that as long as resolutions are not too coarse, users are equally comfortable for large ranges of angular resolutions. No-interpolation strategies lead to good results for angular resolutions higher than 0.37 views/deg. A similar conclusion holds when one is interested in the sense of immersion or the experience as a whole, provided by such systems.

The larger number of subsets and the small number of groups inside each subset for the model fluidity factor also show that although users tend to group a larger number of angular resolutions in terms of visual comfort and sense of immersion, they are also capable of differentiating between them in terms of model fluidity. This allows us to conclude that the common idea that model fluidity should be of main concern when it comes to designing a multiview system may be exchanged by finding a trade-off between the desired overall experience and the number of views (or angular resolution).

We believe that the indications provided by the results found in this study may help designers to better find the best trade-off points for their applications.

VII. CONCLUSION

This work investigates how factors such as disparity, amount of parallax and angular resolution (with or without intermediate views interpolation) can influence the QoE in stereoscopic multiview environments by conducting experiments on each factor. To achieve this we produced a specific visualization and manipulation software for stereoscopic multiview environment with motion parallax and zoom in/out. We also devised a dedicated testing room and conducted extensive subjective experiments with a team of subjects, assessing aspects such as visual comfort, sense of immersion, fluidity of the parallax-induced motion, and 3D experience as a whole. We evaluated these different aspects using multiple sequences, including dynamic and static basic computer graphic models, computer-generated realistic scenes, as well as real scenes. The obtained results were interpreted using a statistical framework, producing relevant findings for the optimal values of disparity and the effect of motion parallax. In addition, we could also confirm previous finds in the literature on the effect of angular resolution and interpolation of intermediate views.

The following conclusions were obtained:

- To have motion parallax is better than not having any motion parallax at all;
- As long as motion parallax is used, user experience in an immersive environment is not as critically influenced by configuration parameters such as disparity and amount of parallax as initially thought;
- Users are very sensitive to angular resolution in 3D multiview systems, only perceiving continuity with head movements for resolutions at or above 1.49 views/deg. However, this sensitivity tends not to be as critical when a user is performing a task that involves a great amount of interaction with the multiview content;
- The results suggest that interpolating the intermediate views may reduce the required view density without degrading the perceived 3D QoE.

We believe that our conclusions may be useful for future implementations of plenoptic systems, such as those based on large camera arrays (super multiview systems) or light field cameras. As such, they may be relevant to the studies being carried out by standardization bodies such as MPEG [5], [22] and JPEG [24]. They also indicate that a better understanding of how a satisfactory 3D experience can be obtained still requires the design and conduction of far more comprehensive tests.

REFERENCES

Felipe M. L. Ribeiro was born in Rio de Janeiro, Brazil, in 1987. He received his B.Sc. degree in Electronics Engineering from Universidade Federal do Rio de Janeiro (UFRJ) in 2012; he received his M.Sc. from COPPE/UFRJ in 2014 and is now pursuing his D.Sc. degree from COPPE/UFRJ, both in Electrical Engineering. His research interests include image processing, video/image quality evaluation, computer vision, machine learning, and pattern recognition-tracking.

José F. L. de Oliveira has graduated in Electrical Engineering (1994) from the Universidade Federal do Rio de Janeiro (UFRJ) and received M.Sc. (1997) and D.Sc. (2003) in Electrical Engineering from the Universidade Federal do Rio de Janeiro (UFRJ). His research interests include signal processing, image compression, and pattern recognition-tracking.

Alexandre G. Ciancio received the B.Sc. and M.Sc. degrees in Electrical Engineering from Universidade Federal do Rio de Janeiro (COPPE/UFRJ), Brazil, in 1999 and 2001, respectively, and the Ph.D. degree in Electrical Engineering from the University of Southern California, Los Angeles, CA, in 2006. He was a Post Doc researcher at COPPE/UFRJ from 2006 to 2007 and took part in research projects at COPPE/UFRJ from 2007 to 2015, including a collaboration with HP on multimedia quality assessment. His main areas of interest are distributed compression algorithms and quality assessment of digital video. He is currently at the Brazilian Patent Office (INPI), where he was a patent examiner at the Telecommunications Division from 2009 to 2015. He is now Technical Assistant of the Director of Patents of INPI.
Eduardo A. B. da Silva (M’95, SM’05) was born in Rio de Janeiro, Brazil. He received the Electronics Engineering degree from Instituto Militar de Engenharia (IME), Brazil, in 1984, the M.Sc. degree in Electrical Engineering from Universidade Federal do Rio de Janeiro (COPPE/UFRJ) in 1990, and the Ph.D. degree in Electronics from the University of Essex, England, in 1995.

He was with the Department of Electrical Engineering at Instituto Militar de Engenharia, Rio de Janeiro, Brazil in 1987 and 1988, with the Department of Electronics Engineering, UFRJ since 1989 and with the Department of Electrical Engineering, COPPE/UFRJ since 1996. He is co-author of the book “Digital Signal Processing – System Analysis and Design”, published by Cambridge University Press, in 2002, that has also been translated to the Portuguese and Chinese languages, whose second edition has been published in 2010.

He has served as associate editor of the IEEE Transactions on Circuits and Systems – I and II, and of Multidimensional, Systems and Signal Processing. He is Deputy Editor-in-Chief of IEEE Transactions on Circuits and Systems I. He has been a Distinguished Lecturer of the IEEE Circuits and Systems Society. He was Technical Program Co-Chair of ISCAS2011. His research interests lie in the fields of signal and image processing, signal compression, digital TV, and pattern recognition, together with its applications to telecommunications and the oil and gas industry.

Cássius R. D. Estrada was born in Rio de Janeiro, Brazil. He received the Electronic and Computer Engineering degree from Universidade Federal do Rio de Janeiro (UFRJ), Brazil, in 2008, and the M.Sc. degree in Electrical Engineering from Universidade Federal do Rio de Janeiro (COPPE/UFRJ) in 2011.

He was a TV Systems Researcher at Rede Globo between 2006 and 2015. He is currently Executive Supervisor of Exploratory Research at Rede Globo.

He has experience in Electronic Engineering and Computer Science, with emphasis on Signal Processing, working mainly on the following topics: image processing, video coding, digital TV, and quality evaluation.

Luiz G. C. Tavares was born in Rio de Janeiro, Brazil, in 1989. He received the Electronics and Computing Engineering degree from Universidade Federal do Rio de Janeiro (UFRJ) in 2013 and the M.Sc. degree on Electrical Engineering from COPPE/UFRJ in 2016. He is currently working at the Brazilian Army Technological Center (CTEx), and has interest in radar and image signal processing.

Jonathan N. Gois was born in Rio de Janeiro, Brazil, in 1990. He received the Electronics engineering degree from Universidade Federal do Rio de Janeiro in 2013 and the M.Sc. degree in Electrical Engineering, in 2016, from the same university. Since 2016, he is a Department of Electrical Engineering’s Professor at Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, in Rio de Janeiro, Brazil. His research interests include image processing, video processing, video fusion, machine learning, and subsea communications.


Amir Said (S’90–M’95–SM’06–F’14) received the B.S. and M.S. degrees in electrical engineering from University of Campinas, Brazil, and the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY. After working at IBM, University of Campinas, HP Labs, and LG Electronics, in 2015 he joined Qualcomm Technologies, where he is now principal engineer. His current research interests are in the areas of multimedia signal processing, compression, and 3D visualization, and their efficient implementation in new processing architectures. He has more than 100 technical publications among book chapters, conference and journal papers, more than 30 US patents and applications. Dr. Said received several awards including Best Paper Award from the IEEE Circuits and Systems Society, the IEEE Signal Processing Society Best Paper Award for his work on multi-dimensional signal processing, and the Most Innovative Paper Award at the 2006 IEEE International Conference on Image Processing. Among his technical activities, he was Associate Editor for the SPIE/IS&T Journal of Electronic Imaging, and IEEE TRANSACTIONS ON IMAGE PROCESSING; a member of the IEEE SPS Multimedia Signal Processing, and the Image, Video, and Multidimensional Signal Processing Technical Committees, was technical co-chair of the 2009 IEEE Workshop on Multimedia Signal Processing, the 2013 Picture Coding Symposium, and has co-chaired conferences at the SPIE/IS&T Electronic Imaging since 2006.