

Subjective Assessment of Super Multiview Video with Coding Artifacts

Rocío Recio, Pablo Carballeira, Jesús Gutiérrez, and Narciso García

Abstract—The subjective assessment of super multiview (SMV) video considers two main perceptual factors: image quality and visual comfort at the viewpoint transition. While previous works only covered raw content with high levels of visual comfort, this work supersedes them by targeting the subjective assessment of SMV content with coding artifacts. The outcome of this analysis yields important conclusions regarding the relationship between these two factors, indicating that 1) the perceived image quality is independent from the view point change speed, and 2) the perceived visual comfort at the view point transition is independent from the image quality. These conclusions facilitate the extension of the scope of existing subjective perception models, designed for raw SMV content, to coded content.

Index Terms—3D-High Efficiency Video Coding (3D-HEVC), multiview perceptual disparity model (MVPDM), quality of experience (QoE), super multiview (SMV) video, subjective assessment, video coding.

I. INTRODUCTION

NOWADAYS, important advances are taking place in three-dimensional (3D) audiovisual systems due to the growing interest in increasing the immersive sensation for users. Different types of 3D displays have been developed in the last few years to enhance the depth sensation of the viewers, evolving from stereoscopic displays to super multiview (SMV) displays [1].

Stereoscopic displays were developed as the first generation of 3D displays, but they present some disadvantages such as the lack of motion parallax (key to immersion engagement) and the need of wearing specific glasses. These displays evolved toward autostereoscopic displays that, despite being a glasses-free technology and solving the lack of motion parallax, present

other drawbacks as low view density or the accommodation-vergence conflict [2], negatively affecting the visualization experience [3].

Currently, important advances have been carried out in order to improve the immersive sensation obtained from the viewpoint change, resulting in the so-called SMV displays [4]. This technology, currently still under development, permits displaying a high view density allowing the user to perceive motion parallax without discontinuities in the view transition and without accommodation-vergence conflict, and seems to be the most promising glasses-free visualization technology.

As in any other audiovisual technology, the analysis of the quality of experience (QoE) of the users is essential in the development of SMV displays and the corresponding video processing/delivery chain. This analysis needs to capture the main factors influencing the 3D QoE, in relation to issues such as horizontal parallax, depth quality, and visual comfort.

Only a few models for the QoE for SMV content have been proposed in the literature [5], [6]. The recently proposed Multiview Perceptual Disparity Model (MVPDM) [7] is an approach to model the user perception of viewpoint change for SMV content that overcomes some of the limitations of previous proposals.

This letter follows the work in [7] addressing the subjective perception of SMV content that was not covered in the validation of the MVPDM, which only comprised uncompressed content, and corresponding to original/physical cameras (not synthesized). Here, new results on the subjective assessment of SMV content with distortion artifacts are presented.

The distortion in SMV content can be related to different blocks of the processing chain such as: coding, transmission, and rendering of virtual views (view synthesis). Very few efforts have been done until now to evaluate the perceptual effects of these processes. As an example, Dricot *et al.* [8] carried out a preliminary subjective evaluation of coding and view synthesis artifacts in SMV video with a light-field display. However, the study in [8] did not address important aspects of SMV video, such as different camera arrangements, and only provided a qualitative preliminary observation regarding perceptual effects of motion parallax.

This letter also considers the coding artifacts in SMV sequences, analyzing, in a systematic way, their effect on the subjective perception of image quality and speed comfort (motion parallax). The subjective assessment has been done using the view-sweep methodology described in [7], considering both monoscopic and stereoscopic viewing, and different view-sweep speeds (VSS), i.e., different viewpoint change speeds. SMV content coded with the 3D-high efficiency video coding (3D-HEVC) [9] scheme has been used in this analysis.

TABLE I
BITRATE AND PSNR VALUES OF THE ANCHOR SMV SEQUENCES

| | Butterfly_Arc (BA) | | | Flowers_Arc (FA) | | | Pantomime (Pa) | | |
|-----------|--------------------|----------------|-----------|------------------|----------------|-----------|----------------|----------------|-----------|
| | QP | Bitrate (Kbps) | PSNR (dB) | QP | Bitrate (Kbps) | PSNR (dB) | QP | Bitrate (Kbps) | PSNR (dB) |
| R1 | 37 | 1665.9 | 37.66 | 35 | 5513.0 | 38.0 | 37 | 7585.8 | 35.34 |
| R2 | 44 | 862.9 | 34.07 | 40 | 3298.5 | 35.1 | 43 | 3188.7 | – |
| R3 | 50 | 563.9 | 30.92 | 45 | 1905.0 | 32.2 | 47 | 1919.6 | – |
| R4 | – | – | – | 50 | 1156.3 | 29.2 | 50 | 1393.3 | 27.98 |

Only three bitrates were provided for Butterfly_Arc in the anchor sequence data of the CfE on FTV [10]. Some PSNR results for Pantomime (bitrates R2 and R3) were not available in that data.

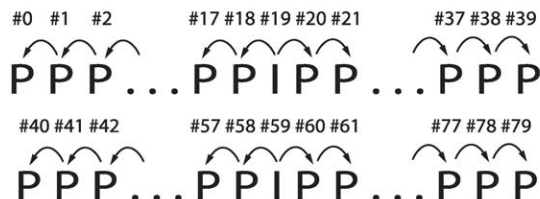


Fig. 1. Interview prediction structure for 3D-HEVC SMV anchors of the two view groups. I represents an anchor view and P represents a view predicted from other reference view.

Relevant conclusions on the subjective perception of the SMV content stem from the subjective results analysis. First, the perception of image quality is independent from speed in the viewpoint change. Second, the perception of speed comfort (perception of visual comfort in the viewpoint change rate) is independent of the level of coding distortion. Finally, there are no substantial differences in regard to perception between the mono and stereo visualization for both image quality and speed comfort. Future work will include the validation of these conclusions in test scenarios in which the user can freely select the view point.

These conclusions are important in the development of QoE models for SMV content, stating that models such as the MVPDM validated for stereo and uncompressed content, are also valid for content with coding artifacts or monoscopic viewing. The remainder of this letter is structured as follows: in Section II, the preparation of the content for the subjective tests is described, Section III presents a detailed description of the test methodology, whose results are described in Section IV. Finally, Section V provides the main conclusion.

II. PREPARATION OF THE CONTENT FOR SUBJECTIVE TESTS

A. Description of Coded SMV Content

The coded content used in the subjective evaluation is the set of SMV anchor sequences used in the MPEG’s Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation [10]. Each SMV anchor sequence comprises 80 views encoded with HTM 13.0 in 3D-HEVC mode (several institutions participated in the selection of the HTM version and preparation of coded content). The 80 views were divided into two groups of 40 views, which were encoded independently using interview prediction. An example of the interview prediction structure of the two view groups is depicted in Fig. 1.

Table I shows the QP, bitrate, and PSNR values for each of the SMV anchors used in the tests. The QP values are those used in [10], and were selected to obtain rate points with sufficient

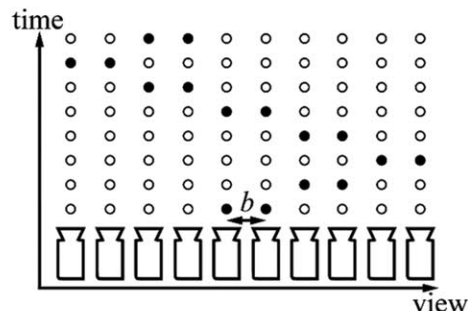


Fig. 2. View-sweeping scheme at VSS \times 2 for stereo baseline $b = 1$. The frames of each view are represented by the dots of the corresponding column. The frames selected to create the view-sweep clip are represented by the black dots.

TABLE II
NUMBER OF SKIPPED VIEWS BETWEEN FRAMES FOR EACH VSS

| VSS id | N° of skipped views | VSS |
|----------------|---------------------|-------------------|
| VSS \times 1 | 0 | \times 1 (1fpv) |
| VSS \times 2 | 1 | \times 2 |
| VSS \times 4 | 3 | \times 4 |
| VSS \times 8 | 7 | \times 8 |

subjective visual quality difference. In addition, uncompressed versions of the videos were also used. Butterfly and Flowers [11] have an arc camera setting while Pantomime [12] have a linear convergent setting.

B. Preparation of Test Clips

The coded SMV content described in Section II-A was shown to the test subjects using the view-sweep methodology used in [13] and [7].

The coded SMV content was shown at different VSS. In addition to view-sweeps at 1 fpv (1 frame per view), view-sweep clips at higher VSS were created by skipping cameras between consecutive frames of the view-sweep clip. Fig. 2 shows an example of view-sweep clips with VSS accelerated by a factor of 2 (VSS \times 2). Table II indicates the number of skipped views in each version of the view-sweep clips.

Note that not all sequences were shown at all VSS rates. Some combinations of content and VSS were found too uncomfortable by means of an expert viewing session, and discarded before the subjective test. Table III shows the combinations of content and VSS that were shown in the test.

For each of the combinations in Table III, monoscopic and stereoscopic clips were prepared. The stereo baseline distances for each content, indicated in Table III, are those used in [10].

TABLE III
VSS AND STEREO BASELINE DISTANCES (MEASURED IN CAMERA GAPS) USED FOR EACH SMV CONTENT

| Content | Id | VSS | Stereo Baseline (b) |
|---------------|----|-------------|---------------------|
| Butterfly_Arc | BA | VSS×1-VSS×8 | 3 |
| Flowers_Arc | FA | VSS×1-VSS×8 | 1 |
| Pantomime | Pa | VSS×1-VSS×4 | 1 |

III. TEST PROCEDURE

A. Methodology

A single stimulus methodology was used to evaluate the test sequences, in particular the Absolute Category Rating method [14]. Thus, after an initial message indicating the start of the test, the test sequences were shown followed by a message indicating the observers to vote for the corresponding video sequence. To collect the opinions of the observers, questionnaires with numbered boxes were used, where they were asked to write a mark for the corresponding evaluation.

The voting messages had a duration of 12 s to allow the subjects to judge the following factors after watching each test clip:

- 1) Image quality using a 5-grade quality scale [15].
- 2) Speed comfort using the 5-grade comfort scale [16], considering how the view-sweeping speed is perceptually comfortable.

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 explaining the tests methodology and purpose. This training provided a reference about the range of qualities used in the tests.

Different randomizations of the videos were used in each session, in order to reduce the contextual effects (with the condition of not showing the same source content twice consecutively). Also, the presentation order of stereo and monosubsets was alternated for each session.

Statistical analysis of previous results in similar tests [7] revealed no statistically significant difference between one or two presentations of the same clip. Therefore, each clip was shown only once to maintain a reasonable session duration.

B. Test Environment and Equipment

The test area was set according to international recommendations [15] and the ambient lightning conditions were controlled to avoid the disturbing reflections. The viewing distance was set at 2.1 m from the display position (3H approx.).

A 55" Samsung UE55HU8500L stereoscopic display was used to carry out the test, connected to a high-performance PC running a VLC player.

C. Observers

A total of 20 observers participated in the test (13 men and 7 women) with ages ranging between 20 and 49 (average 27) and all of them having normal or corrected-to-normal vision. All observers had watched 3D video before the tests: 18 occasionally and 2 only once. Sessions were arranged for two observers simultaneously. After the test, the subjects' scores were screened following standard recommendations [15] [17], which led to discard one observer.

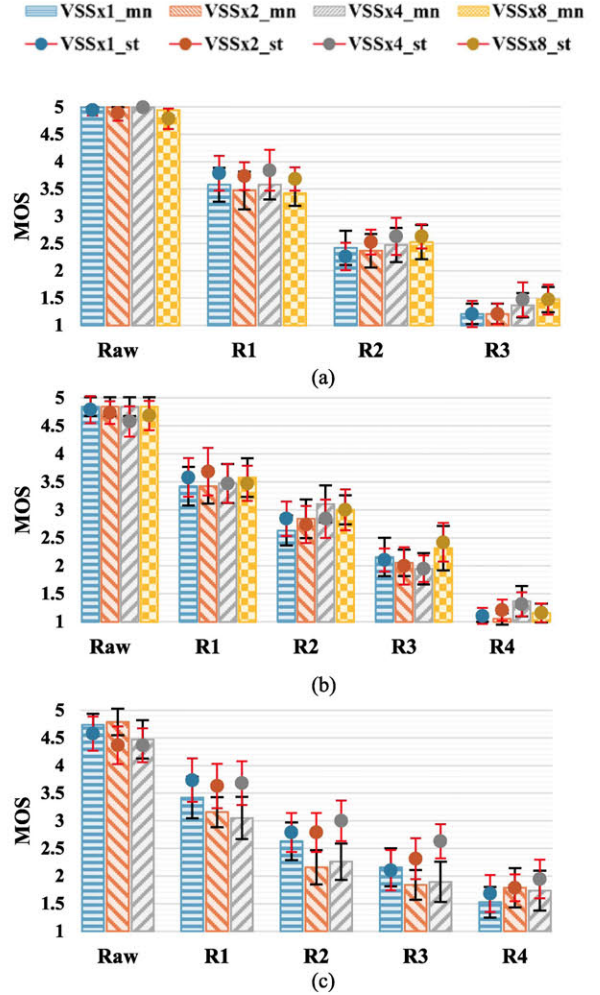


Fig. 3. MOS results of image quality for anchors at multiple VSS. Results for mono (bars) and stereo (dots).

IV. SUBJECTIVE RESULTS

A. Image Quality at Multiple VSS

Fig. 3 shows the image quality mean opinion score (MOS) results for the SMV anchors shown at different VSS. The results for the monoscopic versions are depicted with bars, while the results for the stereoscopic versions are depicted with dots. Each set of bars/dots groups the same bitrate at different VSS for each content. It can be seen that the MOS results within each group are statistically similar in most of the cases. This indicates that the image quality is similarly perceived at different VSS and only varies with the bitrate.

This result yields the following conclusion: The perception of image quality in SMV content with coding artifacts is not dependent on the speed of the viewpoint change rate, and coding artifacts are not masked by high values of VSS, or a rapid movement of the user head in front of a SMV display. These results are in line with the preliminary observations made by Dricot *et al.* [8] stating that compression artifacts do not seem to impact motion parallax.

B. Speed Comfort at Multiple VSS

Fig. 4 shows the speed comfort MOS results for the SMV anchors shown at different VSS. In this case, each set of bars/dots

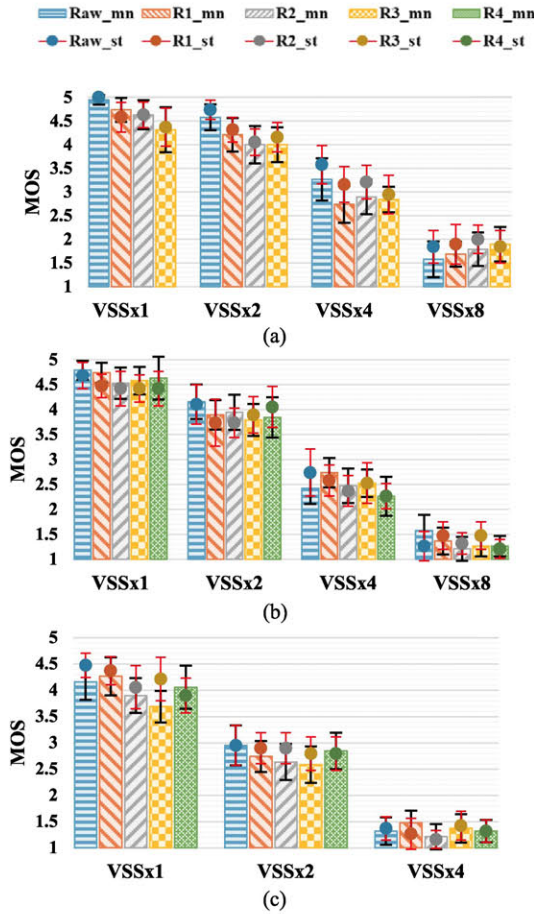


Fig. 4. MOS results of speed comfort for anchors at multiple VSS. Results for mono (bars) and stereo (dots).

groups results for the same VSS at different bitrates. Analogously to the results in Section IV-A, the speed comfort MOS results within each group are statistically similar in most of the cases. This indicates that the perception of speed comfort is independent from the bitrate, and thus, the perceived image quality. These results show that the level of coding artifacts do not influence the perception of speed comfort in the view transition. This conclusion is a key in the development of models of the subjective perception speed comfort for SMV, such as MVPDM, since it allows the extension these models to coded content without modifications.

Moreover, although the speed comfort is influenced by the content of the sequences, the results show that higher speeds than $VSS \times 2$ may considerably degrade the visual experience of the viewers.

C. Comparison of Subjective Results for Mono and Stereo

Figs. 3 and 4 show the MOS results of image quality and speed comfort for the mono (bars) and stereo (dots) versions. On one side, the data reveals that image quality is perceived similarly for mono and stereo viewing. Only a conclusive preference (nonoverlapping confidence intervals) is found for Pantomime at R2, R3, and $VSS \times 4$. On the other side, speed comfort is perceived similarly in mono and stereo for all of the cases, since none of the tested clips present statistically significant differences. This result permits to extend the application of

models developed for stereoscopic viewing, such as MVPDM, to monoscopic viewing.

V. CONCLUSION

A subjective analysis of image quality and speed comfort has been conducted for SMV content with coding artifacts and different viewpoint change rates using the view-sweep methodology. The subjective results yield highly relevant conclusions. First, the perception of image quality in SMV content with coding artifacts is not dependent on the speed of the viewpoint change rate. Second, the perception of speed comfort in the viewpoint change rate for SMV content is also independent of the bitrate or the perceived image quality. Finally, monoscopic and stereoscopic versions yield statistically similar results for image quality and speed comfort.

These conclusions allow extending the scope of perceptual models for SMV, validated for uncompressed content and stereoscopic viewing, to coded content and monoscopic viewing. Future work will include similar subjective tests that include synthesis artifacts, and validation of the results in a realistic scenario in which the user can freely select the viewpoint.

REFERENCES

- [1] J. Geng, "Three-dimensional display technologies," *Adv. Opt. Photon.*, vol. 5, no. 4, pp. 456–535, Dec. 2013.
- [2] S. Reichelt, R. Haussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," *Proc. SPIE*, vol. 7690, Apr. 2010, Art. no. 7690.
- [3] J. Gutiérrez, F. Jaureguizar, and N. García, "Subjective comparison of consumer television technologies for 3D visualization," *J. Display Technol.*, vol. 11, no. 11, pp. 967–974, Nov. 2015.
- [4] Y. Takaki, "Development of super multi-view displays," *ITE Trans. Media Technol. Appl.*, vol. 2, no. 1, pp. 8–14, Jan. 2014.
- [5] D. Kim and S. Kim, "Visual comfort range in the super-multi-view display," in *Proc. Int. Display Res. Conf.*, Sao Paulo, Brazil, Nov. 2010, pp. 1–4.
- [6] F. Speranza and W. Tam, "Perceived smoothness of viewpoint transition in multi-viewpoint stereoscopic displays," *Proc. SPIE*, vol. 5664, pp. 72–82, Mar. 2005.
- [7] P. Carballera, J. Gutiérrez, F. Morán, J. Cabrera, F. Jaureguizar, and N. García, "Multiview perceptual disparity model for super multiview video," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 113–124, Feb. 2017.
- [8] A. Dricot *et al.*, "Subjective evaluation of super multi-view compressed contents on high-end light-field 3D displays," *Signal Process., Image Commun.*, vol. 39, pp. 369–385, Nov. 2015.
- [9] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro, and Y. K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, quality of experience (QoE) pp. 35–49, Jan. 2016.
- [10] ISO/IEC JTC1/SC29/WG11, "Call for evidence on free-viewpoint television: Super-multiview and free navigation—update," Output doc. N15733, 113rd MPEG Meeting, Geneva, CH, Oct. 2015.
- [11] P. T. Kovacs, A. Fekete, K. Lackner, V. K. Adhikarla, A. Zare, and T. Balogh, "Big buck bunny light-field test sequences," Contrib. M36500, 112th MPEG Meeting, Warsaw, Poland, Jun. 2015.
- [12] M. Tanimoto, T. Fujii, and N. Fukushima, "1D parallel test sequences for MPEG-FTV" Contrib. M15378, 84th MPEG meeting, Archamps, FR, Apr. 2008.
- [13] ISO/IEC JTC1/SC29/WG11, "Revised summary of call for evidence on free-viewpoint television: Super-multiview and free navigation," Output doc. N16523, 116th MPEG meeting, Chengdu, CN, Oct. 2016.
- [14] ITU-T, "Subjective video quality assessment methods for multimedia applications," Rec. ITU-T P.910, Apr. 2008.
- [15] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Rec. ITU-R BT.500-13, Jan. 2012.
- [16] ITU-R, "Subjective methods for the assessment of stereoscopic 3DTV systems," Recommendation ITU-R BT.2021, Aug. 2012.
- [17] Video Quality Experts Group, "Test plan for evaluation of video quality models for use with high definition TV content!" Jun. 2009.