Towards the assessment of Quality of Experience for asymmetric encoding in immersive media

Federica Battisti, *Senior Member, IEEE,* Marco Carli, *Senior Member, IEEE,* Patrick Le Callet, *Senior Member, IEEE*, Pradip Paudyal *Student Member, IEEE,*

***Abstract*—The assessment of Quality of Experience for stereoscopic 3D video is a challenging task, especially in 3D video compression and transmission applications. The focus of this contribution is the development of a Quality of Experience assessment framework for understanding the visual effect of asymmetric and symmetric encoding for immersive media.**

Asymmetric stereoscopic video coding exploits the binocular suppression of the human vision system by representing one of the two views with a lower quality. This processing may influence the quality of experience of a person viewing the 3D content. Many studies show that the quality of experience can be thought as the combination of the perceived visual quality, the perceived depth quality, the visual fatigue, and visual discomfort. In this work, we aim at: i) investigating the factors involving visual discomfort in stereoscopic video sequences with a focus on binocular rivalry; ii) exploiting the concept of Preference of Experience for the subjective quality assessment of stereoscopic video; iii) presenting the results of subjective experiments performed, by using the perceptual quality and preference of experience assessment protocols, for evaluating the impact of symmetrical, asymmetrical, and alternate coding schemes.

***Index Terms*—Quality of Experience, Asymmetric video Coding, Immersive media, Standard for subjective testing**

I. Introduction

IMmersive media refers to technologies able to induce the *immersivity* feeling to the user. This status is achieved by techniques, both aural and visual, able to completely engage the user [1]. As stated by Dale Lovell in [2], ”engagement is great, but immersion is the future. Immersion is when you forget the message entirely, forget you are the audience even, and instead fall into a newly manufactured reality”. One of the main challenge is the content creation. It may be generated exploiting the inputs from different sensors (video camera, microphone, depth camera, light field camera, etc.), or it can be computer-generated. In both cases, one severe requirement is that the content should be perceived as high-realistic or ultra-realistic.

In order to create an immersive content the scene should be recorded, by as many points of view as possible. In this case the computational and hardware requirements are too severe to be practical. To cope with this issue a multiview-based approach has been introduced. It is based on the use of multiple videos simultaneously recorded from different camera located in different points of view to provide interactive content. The increase in the data volume due to the number of views, efficient compression and robust transmission are essential. The simplest multi view technique is the 3D in which only two views are considered and rendered to the user. Even if the system complexity is reduced, 3D shares a common element with the multivista framework: the human factor.

Especially when human subjects are involved, the impact of a new technology on the perceived experience is a fundamental issue. If the human-in -the-loop factor is not properly addressed, the novel technology may not be successful. The negative trend of stereo content, especially in home environment [3], is probably due to the fact that the actual 3D content production, delivery, and presentation, are not compliant with 3D Quality of Experience (QoE). The success of the 3D imaging market relies on the ability of 3D systems to provide added value compared to conventional monoscopic imaging (i.e., depth feeling or parallax motion) coupled with high image quality contents. Dealing with these issues can result in the creation of perceivable impairments in the 3D content that may be originated in different points of the 3D chain, from content creation to display techniques. Many artifacts are common to 2D imaging system. However, novel distortions typical of the 3D structure should be considered (i.e., crosstalk or keystone) [4]. All those impairments should be taken into account since their presence highly impacts on the perceived quality (i.e. compression artifacts due to coding). As explained in [5], subjects are prone to prefer 2D contents to 3D ones, as soon as fatigue and discomfort are induced during the content presentation.

The understanding of the quality of the experience is mandatory. However, this task is quite challenging. The quality of 3D content may be addressed in different ways: as high-level concepts such as naturalness and sense of presence, or by means of three perceptual dimensions that may affect the QoE, such as picture quality, depth quality, and visual discomfort. The latter features have been widely used in literature for addressing 3D visual quality, and many efforts are being devoted to quantitative measure such parameters [6]. Therefore, the selection of particular perceptual quality assessment method is crucial for 3D content, since there is no single method that can accommodate all the quality dimension of 3D content.

In literature, many ongoing efforts have been given to the design of encoding methods for stereo video. They can be classified in three groups according to the adopted strategy: frame-compatible methods, Multi-view Video Coding (MVC)-based approaches, and View-plus-Depth (V+D) coding schemes [7]. If we consider MVC and V+D schemes, a promising technique for coding stereo pairs is the asymmetric coding. This

F. Battisti, M. Carli, and P. Paudyal are with the Department of Engineering, Universitá degli Studi Roma TRE, Rome, Italy.

P. Le Callet is with Polytech’ Nantes, Nantes, France.

approach is based on the *binocular suppression theory*: the 3D perceived quality of a stereo pair can be close to the view with the highest quality even if one of the views of a stereo pair is altered, provided that the difference between both views does not exceed a threshold [8]–[10]. The perceived impact of these techniques must be studied and standardized systems for evaluating both effectiveness in bit rate consumption and quality impact have to be designed. In other terms, there is the need of understanding how to evaluate the novel coders. In particular, the knowledge of the impact of asymmetric distortions on the quality of stereoscopic video is an important task, and thus, many ongoing efforts are given to develop novel high efficiency asymmetric 3D video coding schemes [11].

In this work, we are dealing with the technological aspect: existing bandwidth limitations force the use of lossy coders, due to the huge amount of data generated by immersive technologies. As detailed in the next section, many efforts are being devoted to design new coders or creating amendments to existing standards. The encoders not only reduce the size of immersive content but also degrades perceptual quality. QoE assessment of 3D content is even more complex because, the stereoscopic 3DTV exploits the characteristics of the human binocular visual system by recreating the conditions that bring about the perception of the relative depth of objects in the visual scene. In more details, in this work the following contributions are presented:

- The factors involving visual discomfort in stereoscopic video sequences are investigated with a focus on binocular rivalry in order to improve objective assessment tools of stereoscopic contents. Among the sources of binocular rivalry conditions, we consider binocular asymmetries brought by asymmetrical coding of stereoscopic sequences. In more details, we study the impact on the visual comfort of the effects of asymmetrically coding the views of 3D sequences exploiting the coding standard HEVC and 3D-HEVC. Our assumption is that, the different structure of the coders may stress different binocular rivalry conditions, and thus, it is worth revising the results for the recent encoders. Particularly, i) visual comfort of symmetrically versus asymmetrically encoded video, ii) influence of (left/right) eye dominance, and iii) visual comfort in terms of preference of experience of the encoded videos, have been investigated.
- From the methodology point of view, the novelty relies in evaluating the visual comfort with the help of overall Quality of Experience (QoE) rather than the perceptual quality only and preference of experience. The normally adopted (discrete/continuous) quality scale is known to be ambiguous in this framework, since subjects usually refer to a measure of picture quality rather than discomfort. Here, we prefer assessing the experience by giving subject the clear instructions and using dedicated innovative methodology and framework, PoE (Preference of Experience) that has been demonstrated to the effective in such context ([12], [13], and IEEE P3333 standard).
- Finally, the study is based on the opinion scores collected on two subjective experiments. In the first, only HEVC encoding and pair comparison protocol are used. The achieved results are confirmed by the second experiment, which is designed by using 3D-HEVC encoding and single stimulus absolute category rating. The details are presented in Sections III and IV.

The rest of the paper is organized as follows. In Section II the state of the art is presented. In Section III a discussion on the quality assessment protocols is presented. In Section IV the performed experiments are detailed while in Section V the results of the experiments are presented; finally, in Section VII the conclusions are drawn.

II. Background analysis

In this section, we briefly revise the state of art on two topics related to the current work: the asymmetric coding and the study of the impact of artifacts introduced by the coding on the human viewer.

A. Video Compression

A review of 3D video coding schemes can be found in [14]–[16]. Stereo video encoding techniques can be classified in three groups: frame-compatible, MVC (Multi View Coding), and Video plus Depth (V+D) coding.

In frame-compatible methods, stereo signals are re-arranged by multiplexing two views into a sequence of frames. Basically, the views are filtered, sub-sampled, and then combined. The fusion can be done in the spatial (side-by-side or top-bottom format) or in the temporal (by interleaving left and right frames) domain [17], [18]. In literature, compression standards (i.e., MPEG2, H.264/AVC, or HEVC) and standardized broadcasting protocols (i.e., MPEG-2 Transport System) for delivering the stereo video [19] have been exploited. The main advantage of these methods is in the full compatibility with existing hardware and communication systems without requiring changes in co/decoders and/or in the network.

In MVC-based methods, one of the two views is compressed with a standard coder. The other view is compressed by exploiting the same compression system with the only difference called Disparity-Compensated Prediction (DCP) for removing the redundancy between the two channels [20]. The category adopting V+D schemes includes methods, exploiting the prediction-based approach, that is the common approach adopted in most of existing 2D video coding standards (i.e., MPEG-X, H.264/AVC, and the new HEVC (High Efficiency Video Coding)). The extension to 3D world results in the MVC H.264/AVC standard extension and in the HEVC standard extensions (both multi-view HEVC (MV-HEVC) and depth-enhanced HEVC (3D-HEVC)). This system may also be used for a generalized multi-view framework: however, the bit rate required for coding multi-view video with MVC, which increases approximately linearly with the number of viewpoints, is too large.
To overcome this problem, Depth Image Based Rendering (DIBR) techniques have been designed. In these systems, additional views are generated at the receiver side by exploiting the transmitted video sequence and depth maps [21]. The estimated depth information is considered as a single

low frequency information that can be compressed with high efficiency [22]. Several works have been devoted to design these systems and understanding their impact on the final user [23]. Furthermore, an international standard has been defined in the MPEG framework [24]. Compression methods based on the suppression theory of binocular vision [25], exploit the assumption that binocular perception can support differences in quality between the two views: therefore, the two views can be represented at unequal resolutions or bit-rates. In asymmetric encoding of stereo pairs the two views are generally compressed with different compression rate while trying to leave the stereo perceived distortion below the Just Noticeable Distortion (JND) level, where the HVS can just distinguish the difference between the quality.

Depending on the modalities adopted for reducing the quality of one of the two views, asymmetric stereoscopic coding methods can be classified into three categories: spatial resolution reduction (spatial filtering) [14], [20], [26]–[29], asymmetric quantization (use of different values for the quantization parameter) [7], [30]–[32] and temporal resolution reduction [33]–[37].

In [20] two techniques for compressing stereo pairs are presented: a disparity-compensated transform-domain predictive coding with the aim of minimizing the mean-square error between the original stereo-pair and the compressed stereo pair, and a mixed-resolution coding, exploits human stereo vision properties for coding stereo pairs. Basically, the two views are represented at different resolution by low-pass filtering one view, while keeping the other view at full quality. The achieved results show that perceived 3D video quality is similar to the one of the original sequences. In [38], the performed subjective experiment resulted in image quality ratings of a symmetric coded pair higher than for an asymmetric coded pair. Those results are valid even with a lower bit rate used for the symmetric pair.

In [27], a compression system based on the hypothesis that a stereo pair consisting of one monochromatic image and one color image is perceived as a 3D color scene, is presented. In [26] an extension of H.264/AVC-based multi-view video coder is presented. The total bit rate is reduced by spatially encoding with low bit rate the down-sampled right frame. The combination disparity-compensated prediction and asymmetric coding allows to achieve high compression rate while preserving the perceived quality. In [28], the authors propose to motion compensate the low-resolution picture from the high-resolution picture based on picture-level adaptive filters. Simulation results show that low complexity and bit rate reduction are achieved for Asymmetric MVC or Asymmetric Stereoscopic Video (ASV) codec. In [29] the problem of understanding the impact of the length of asymmetrically coded video is addressed. The authors perform three experiments comparing two methods of binocular suppression processing in order to verify the rising of fatigue.

In [30] the chrominance information in selected views is discarded and restored at the decoder. This concept is extended in [31], [32] where a bit allocation model and a chrominance reconstruction model for characterizing the view rendering distortion and the binocular suppression respectively, are used. In [7] the authors use different quantization parameters for understanding the impact of the quality, measured by PSNR, of the unequally compressed view on the overall perceived quality and, in more details, for the just-noticeable asymmetry level in terms of PSNR threshold.

Temporal asymmetry is exploited in [33] where, in each view sequence, high-quality frames are interleaved with reduced-quality frames. The performed experiments show that a jerkiness effect can be noticed while viewing the video, unless the frame switch is positioned in presence of scene cuts. In [34] the authors, by performing frame dropping in the right video sequence, show that temporal asymmetry can result in unacceptable perceived quality in presence of high motion video, while it can be used in case of slow motion video. In [35] the authors propose a frame skipping based algorithm for reducing encoding complexity and transmission bit rate. A bilateral interpolation method is adopted at the receiver side for recovering the missing information. In [36] a trade-off between compression performance and perceived quality is obtained by switching left-right side of blurred video at every group of picture occurrence for short videos. However, the onset of fatigue over time is not addressed. More recently, in [37], different temporal-based compression schemes are subjectively evaluated. The results show that the equal spreading of the degradation induced by the coding scheme in both views, results in high perceived 3D quality if the subjects are characterized by asymmetric visual acuity. Furthermore, the authors conclude that the temporal factor, due to motion or frame rate, can strongly affect the perceived quality and that induced visual fatigue is an important issue to be considered.

Summarizing, several efforts are still ongoing in designing efficient asymmetric coding schemes. From a standardization point of view, the main actors in asymmetric coding are H.264, HEVC, and 3D-HEVC. As analyzed in [39], these methods are different concerning complexity, required parameters, and achieved performances.

B. *stereoscopic image/video quality assessment*

In literature, many efforts have been given to develop the subjective and objective quality assessment models for 3D content. The knowledge of the perceptual impact of the artifacts is the key to develop objective quality metrics and to optimize processing algorithms. In general, the quality assessment metrics can be classified in three categories. In the first one, some successful 2D image/video quality assessment metrics are directly applied to assess 3D image quality. However, these approaches do not exploit the depth-related information [40]. In the second category, image/video distortion measurement combines depth quality to estimate the overall quality, as in [41]–[43]. In the last category, the binocular rivalry properties of the HVS are modeled into conventional 2D quality assessment approaches [44], [45]. Recently, a new depth perception quality metric is proposed in [46]. A novel HVS model to include the phenomena of binocular suppression and recurrent excitation is developed and an image quality metric based on the novel HVS model is proposed in [47]. Similarly, a binocular rivalry inspired full reference quality prediction model for stereoscopic videos is presented in [11]. The model allows to quantitatively predict the coding gain of different variations of asymmetric video compression, and provides new insight on the development of high efficiency 3D video coding schemes.

C. Perceptual impact of asymmetrically coded video

Besides the performances in compression rate, a new coder has to be evaluated also on the quality of the decoded data. The more perceptually-similar is the rendered video to the original one, the better the designed system is. In [48] an evaluation of the impact of 3D video coding techniques, transmission scenarios, error concealment strategies, and their impact on perceived video quality is presented. Interesting conclusions are drawn concerning the relation between resolution and bit rate vs. perceived quality. When dealing with 3D data, an in depth understanding of the human visual perception system is an important aspect for developing effective 3D quality metrics.

1) 3D QoE: a multidimensional experience: Understanding 3D QoE is not an easy task and the application of conventional subjective quality assessment methods is not enough for an effective validation of 3D quality assessment. Both standardization bodies (i.e., ITU and VQEG) and independent groups are working on this issue. The standardization efforts are producing several documents, starting from ITU-R BT.1438 [49] where some fundamental issues such as assessment factors, methodologies, test viewing conditions, test materials and subject pool selections are defined. Following, in order to define better methods for assessing the 3D field, ITU-R SG6 WP6C and ITU-T SG9 have been defined. At the same time, the Video Quality Expert Group (VQEG) has established a new project called 3DTV devoted to 3DTV video quality assessment. A detailed analysis of existing efforts in reported in [50]. Independent research studies [50], [51] are moving towards a set of QoE indicators, including 2D image quality, depth quantity, visual comfort, depth rendering, naturalness and visual experience, and visual fatigue. The overall conclusions is that a common definition for QoE indicators as well as the methodology for subjective assessment must be studied and standardized. Furthermore, test material is of primary importance and should be tuned according to the aim of the experiment, keeping into account that while in 2D annoyance can be considered a key factor, in 3D content visual discomfort and fatigue caused by 3D are not the same factor and must be properly addressed.

2) Visual discomfort and fatigue: As stated in [52] there is a big difference between visual discomfort and visual fatigue. *Fatigue* is *a decrease in the performance of the human visual system as a consequence of physiological strain or stress resulting from excessive exertion*, while *visual discomfort* can be seen as the *subjective counterpart of visual fatigue, partially reflecting some aspects of the quality of experience (QoE)*. Adaptation mechanisms [53] are used by the human visual system as a response to changes in the environment. These continuous adjustments in sensory processing may improve its performances while causing fatigue. Another difference is that visual discomfort is perceived instantaneously, while fatigue is induced after a discrete duration of effort. Currently, there is no clear understanding of the relation between fatigue and discomfort and how to deal with worsening and improving effects, as well as temporal aspects; furthermore, the correct methodology for measuring visual discomfort is still to be defined.

In this work, we propose an ad hoc protocol for evaluating the visual discomfort. To this aim a video database is needed. Some efforts have been performed in designing 3D image test database and in [54] the importance of the distortion effect is assessed. In particular, the authors designed an image database containing 3D images affected by both symmetric and asymmetric distortions such as noise, blur, or JPEG compression.

TABLE I
EXISTING STEREOSCOPIC VIDEO QUALITY DATASETS.

Datasets	Encoding
LIVE 3D Databases [55], [56]	H.264 compression
StSD 3D database [57]	H.264 and HEVC
NAMA3DS1-COSPAD1 [58]	H.264
3DVCL@FER video database [59]	H.264/AVC and 3D-HEVC
DML-3D-HFR [60]	3D-HEVC
Tempere 3D video Database [61]	H.264
MMSPG 3D VQA databbase [62]	Camera Distance
Waterloo-IVC 3D video database [11]	3D HEVC

D. Stereoscopic video quality dataset

In literature, many 3D video datasets are proposed and few of them are reported in Table I. With respect to the aim of this article, the available datasets are not usable, since most of them use H.264/AVC encoding, and the datasets created by using HEVC and 3D-HECV encoding are designed for specific purposes, i.e. they do not provide the symmetrically and asymmetrically coded videos and annotated subjective quality scores which are sufficient for the study.

III. SUBJECTIVE QUALITY ASSESSMENT PROTOCOL

ITU Recommendation ITU-R BT.2021 [63], includes many single stimulus and double stimulus methods, for the subjective quality assessment of stereoscopic 3DTV systems.

A. Perceptual quality assessment

In ITU-R BT.2021 [63] the recommended quality assessment protocol for stereoscopic 3DTV systems are summarized, and these protocols are directly taken from ITU-R BT.500: which is primarily devised for 2D image/video. For perceptual quality assessment of 2D/3D visual content, in the-state-of-the-art, all the methods have been exploited.

The usability of the different methods is based on the test material, the purpose of the study, the experiment environment, etc [64], [65]. One of the most accepted method is the Single stimulus–Absolute Category Rating (ACR) [66]. This method is a category judgment, where the test sequences are presented once at a time and are rated independently on a category scale.

It significantly reduces the amount of time needed to conduct subjective studies compared to the double stimulus method, such as PC method [67]. This method is most appropriate method in terms of stability and assessment time to assess the 3D video quality [65]. In fact, it provides the possibility to use more SRCs, in a given experiment time and subjects.

B. Preference of Experience (PoE) assessment

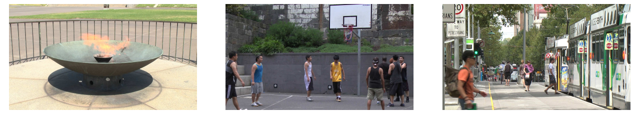
In the context of stereoscopic 3DTV systems, the perceptual quality assessment problem becomes more complex due to the introduction of factors such as depth quantity, occlusion, and visual comfort, together with the one dimensional term ”quality”. The well accepted 2D visual quality assessment methods may not be appropriate to deal with the multi-dimensional problem. In [68] the pair comparison method is considered as a reliable method: as observers just need to answer one question, *which one of the two 3D sequences do you prefer?*. Accordingly, the perceptual quality/QoE is expressed as a *Preference of Experience*.

In PC method the test sequences are presented in pairs, consisting of the same sequence being presented first through one system under test and then through another system, and subjects are asked to select the one based on their preference. Therefore, the outcome of the subjective assessment is referred to as *Preference of Experience (PoE)*. This choice is motivated by the fact that the QoE in 3DTV is considered as a multidimensional subjective impression, resulting from mono dimensional factors, such as image quality, depth quantity and visual comfort [13].

The main advantage of PC is its high discriminatory power, which is of particular value when several test items are nearly equal in quality. Several studies have been performed that show the suitability of the PC approach for image quality evaluation. A detailed discussion is in [68], where a simulation is performed in order to highlight the relationship between the number of subjects, the number of stimuli and the convergence of the quality levels estimated from comparison results. It was found to be the most accurate among other widely used methods: single stimulus, double stimuli, and pairwise similarity judgment method, for visual quality assessment [64]. Moreover, the authors in [69] discuss the use of the PC for evaluating the quality of stereoscopic images.

In this article, our first task has been to obtain visual comfort ratings of the aforementioned asymmetric stereoscopic sequences. As a matter of fact, visual comfort in stereo vision is not always clearly understood by observers. It seems that it is simpler for the observers to answer the question: ”which one of these two 3D sequences do you prefer?” compared to answering ”is the quality of this 3D sequence excellent / good / fair / poor / bad?”. Thus, we consider an assessment methodology involving a simple rating task. Paired comparisons [70] are reliable, accurate and well-adapted to our requirements. Therefore, the PC method is selected as a subjective quality assessment protocol for the PoE assessment.

In another view point, the repeated presentation of the same test sequence limits the use of this technique, if number of test conditions is large.



(a) (b) (c)

Fig. 1. Sample frames extracted from the 3 sequences: (a) Shrine of Remembrance Cauldron and Flame, (b) RMIT University Basketball Players, and (c) Swanston Street Tram Stop.

IV. Experimental Design

In this work, we focus on binocular asymmetries brought by asymmetrical coding of stereoscopic sequences. In particular, two subjective experiment have been performed based on the selected encoding methodology.

In the first experiment the PoE is evaluated by exploiting the PC protocol. To maintain the length of the experiment less than 30 minutes, only 3 SRCs have been used, though the use of large number of SRCs is generally important. It is worth mentioning that, the first experiments have mainly the two shortcomings: i) limited number of SRCs (3 SRCs) are used, and ii) the HEVC encoding is exploited though the 3D-HEVC encoder (HEVC encoder tuned for 3D content) is already available. Therefore it has been considered as a trial. To reduce the impact of these shortcomings, and to confirm the achieved results, another experiment (experiment-II) has been performed. In this experiment 9 SRCs, widely used perceptual quality assessment method (ACR) and 3D-HEVC encoder have been exploited.

A. Experiment-I: HEVC encoding

1) Source sequence: Proper selection of SRC content is important, because the perceptual quality of processed video and level of possible compression are depending on the content [71]. Therefore, the selected SRCs should cover a wide span of key visual quality attributes such as spatial detail of the scene (Spatial perceptual Information - SI) and motion information (Temporal perceptual Information - TI). Considering a large number of SRCs is always good for better understanding the phenomenon under investigation. However, for subjective quality assessment, this number is limited by the processing time, duration of experiment, and available number of subjects for the experiment [72].

We rely on an available dataset of uncompressed HD stereoscopic contents, namely the RMIT3DV [73]. This database has been selected for its availability and its variety of content and conditions. Thus, we selected three different video sequences so that various level of motion and parallax are considered in the test. Only the first five hundred frames are considered. Sample frames extracted from the 3 sequences are shown in Figure 1. The content variations of the selected sequences are explained by Figure 2. Sequences are 1920×1080 pixels (HD resolution), 10-bit 4:2:2 YUV at 25 fps.

The selected sequences are:

- 3D_22 (Shrine of Remembrance Cauldron and Flame): It is a static shot with unpredictable flame and a moderate to strong 3D effect.

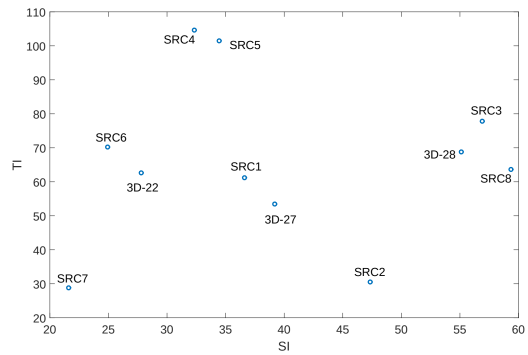


Fig. 2. Spatial-temporal perceptual information plane: SI and TI. Sequences 3D_22, 3D_27, and 3D_28 are used in the first experiment, and the second experiment is designed by exploiting the sequences SRC1–SRC8 and 3D_22.

- 3D_27 (RMIT University Basketball Players): This moving-shot video sequence contains fast motion and a moderate 3D effect.
- 3D_28 (Swanston Street Tram Stop): This static shot video sequence contains complex details and a moderate 3D effect.

2) Hypothetical reference circuits: To create test the sequences, HEVC coding method [74] (Test Model HM 6.1 [HHI Fraunhofer]) has been considered. In general, four to five level of distortions are sufficient to create the test sequences [75]. A preliminary perceptual quality test is performed by the authors to select the quantization parameters that spanned a wide range of visual quality. As a result, four quantization parameters were selected based on the visual quality of the decoded views: QP = 25, 30, 35, 45. The stereoscopic video sequences were built following different HRCs (Hypothetical Reference Circuits) [76]:

- Left view is encoded with QP = 25 and for right view the quantization parameter varies in the range [30, 35, 45]. This leads to three asymmetric stereoscopic sequences. These conditions will be referred, in the following, to as asymmetric right (ASYM R).
- Right view is encoded with QP = 25 and for left view, the quantization parameter varies in the range [30, 35, 45]. This leads to three asymmetric stereoscopic sequences. These conditions will be referred to as asymmetric left (ASYM L).
- Symmetrically coded stereoscopic sequences. Left and right views are encoded with the same quantization parameter in the range [25, 30, 35, 45]. These conditions will be referred to as SYM.
- Stereoscopic views are presented with reduced quality affecting alternately left and right eyes so that the image quality is distributed over both eyes in a balanced manner over time. The alternation is achieved per groups of pictures (GOP) and the size of the group is chosen based on the timing of human visual perception [77]–[79] (we choose a GOP size of 3 for 25fps video sequences). For example, frames 0, 1, and 2 of left view are encoded with QP = 25 and frames 0, 1, and 2 of right view are encoded with QP = 30. Then frames 3, 4, and 5 of left view are encoded with QP = 30 and then frames 3, 4, and 5 of right view are encoded with QP = 25. Then, frames 6, 7, and 8 of left view are encoded with QP = 25 and frames 6, 7, and 8 of right view are encoded with QP = 30; and so on as shown in Figure 3. Considering the three quantization parameters described above, the different combination leads to three asymmetric stereoscopic sequences ([30, 25], [35, 25], [45, 25]). These conditions will be referred to as *Alternate*.

For the three different video contents, the following sequences have been generated:

- 3 sequences with condition ASYM R;
- 3 sequences with condition ASYM L;
- 4 sequences with condition SYM
- 3 sequences with condition *Alternate*.

Summarizing, 39 HD video sequences have been created (3 video contents * (3+3+4+3)). Since the uncompressed versions of each sequence have also been included in the database, 42 sequences have been rated in total.

3) Assessment methodology: As mentioned before, in PC method, the complexity of the test increases with the number of video sequences to assess, since the number of comparisons grows according to $\frac{N(N-1)}{2}$ where N is the number of stimuli. Thus, the test becomes difficult to carry out. To overcome this problem, we considered the methodology proposed by Li et al. [80], [81]. It is a square-design based sub-set paired comparison method that reduces the number of comparisons while producing comparably precise results under some assumptions.

As explained above, we considered 48 stimuli: 42 (as mentioned in Section IV-A2) and 6 (2 stimuli (ASYM L QP40 and SYM QP40) per SRCs), since there was a possibility to add more 6 stimuli during the experiment design. Based on [80], [81], this leads to 144 paired comparisons per observer: there are 3 different video contents and 16 HRC per content. To avoid visual fatigue caused by long time watching (which could affect the experimental results), the subjective assessment session was split into two sub-sessions, with 10 min breaks between sub-sessions. A typical test session should last 60 min including breaks.

Thirty observers participated in the experiment. They were tested for visual acuity, stereo vision and eye dominance. Responses from 30 observers are two discrete grades representing the frequency that a video sequence i is preferred over stimulus j. To analyze the subjective quality, a preliminary step is to convert the preference frequencies in a continuous quality rating scale. In this article, the conversion of the preference frequencies to continuous-scale quality scores is performed using the popular Bradley-Terry (BT) model [82]. A detailed description of the BT score and Confidence Interval (CI) estimation procedure is available in [83]. This leads to a perceptual continuum. Where, Higher BT score indicates higher preference rate. Figures 7 to 9 illustrate these results. The analysis of the results considers three groups of observers: the whole set of 30 observers, the 8 left-eye-dominant observers from the whole set, the 22 right-eye-dominant observers from the whole set. The bars are ranked from the most preferred

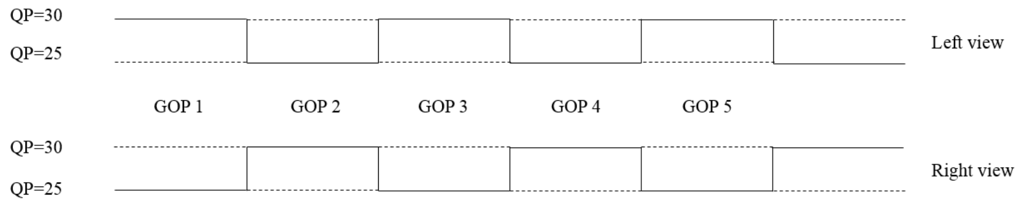


Fig. 3. Structure for the creation of asymmetrically coded sequences.

Fig. 4. Sample frames extracted from the left view of 9 SRCs (Type1: (a) city–SRC1, (b) fire/$3D\_22$, and (c) lake–SRC2; Type 2: (d) elephants dream 1–SRC3, (e) elephants dream 2–SRC4, and (f) elephants dream 3–SRC5; Type 3: (g) Hansel 1–SRC6, (h) Hansel 2–SRC7, and (k) Hansel 3–SRC8).

video sequence to the less preferred. In other words, the higher the quality score, the more preferred. It can be observed that the original video sequence is always preferred over the other sequences, which is expected.

B. Experiment-II: 3D-HEVC encoding

This section briefly describes the second experiment: SRCs, HRCS, and setup.

1) Source video contents: As mentioned in [84], nine stereoscopic sequences (shown in Figure 4) of the duration of 2 minutes are selected. The selected SRCs are categorized as follows: Type 1: objects and landscapes with low to moderate motion within the scene; Type 2: animation short-film with different shots and moderate to high motion within the scenes; and Type 3: commercial Hollywood 3D movie containing different shots, moderate to high motion, and action, suspense, and soft-horror scenes. Content variation of the selected SRCs is proven by a wide range of key quality attributes, as shown in Figure 2.

2) Encoding method: For this experiment 3D-HEVC (Test Model 16.2), an extension of HEVC/H265 [85], has been used. To create test sequences, the same configuration as specified in Section IV-A2 has been used. In *Alternate* group the number of frames used are 2 and 4, and these are referred as GOP4 and GOP8. The different number of frames are included to study the impact of the number of frames in *Alternate* group on the perceptual quality. Therefore, total number of Processed Video Sequences (PVSs) were 16, 3 ASYM R, 3 ASYM L, 4 SYM, and 6 *Alternate* (3 GOP4 and 3 GOP8), for each SRC.

3) Experiment setup: Experiment setup is configured by considering the following issues:

- To reduce the fatigue that may be experienced, when evaluating stereoscopic images, the overall duration of a test session should be less than 30 m [71];
- To find the JND threshold, more then 3 SRCs are needed [86];
- For the experiment, more than 15 expert/non-expert subjects are necessary [71].

Collecting opinion scores for all the test sequences (9×16 = 144) within 30 minutes was not possible. Therefore, 12 test videos were selected form the pool of 144 test videos in pseudorandom order, and evaluated by each subject. The pseudorandom order is maintained in such a way that each stimulus is shown for more than 15 subjects. To collect opinion scores for 16 HRCs from more than 15 subjects, in total, 40 subjects (18 left-eye dominant, 16 right-eye dominant, and 6 subjects are included without eye dominance test) were used.

It is important to notice that all HRCs (16 test conditions) are not evaluated by 40 subjects to each SRC. In this situation, the impact of the content may influence the results, since the perceptual quality of video varies to the different content (SRCs) even at a same level of compression [87]. The usability of the collected opinion scores and test videos may be limited, but it is sufficient to confirm the results achieved from experiment-I.

V. Experimental Results

A. Perceptual quality of symmetric vs asymmetrically encoded video

1) HEVC encoded video: When considering Figure 5, that is the ratings for the three different video contents from the whole set of observers, we observe that the video sequences having the lowest quality scores are the symmetrically coded and asymmetrically coded pairs, and with alternation with $QP = 45$.

For the analysis, with the help of K-means [88] centroid ([25.28, 24.17]; [37.86, 35.71]), the HRCs are grouped as:

- Group A: QP25 S, QP30 L, QP30 R, and QP30 A
- Group B: QP30 S, QP35 L, QP35 R, QP35 A, QP40 L QP40 S, QP45 L, QP45 R, QP45 A, and QP45 S

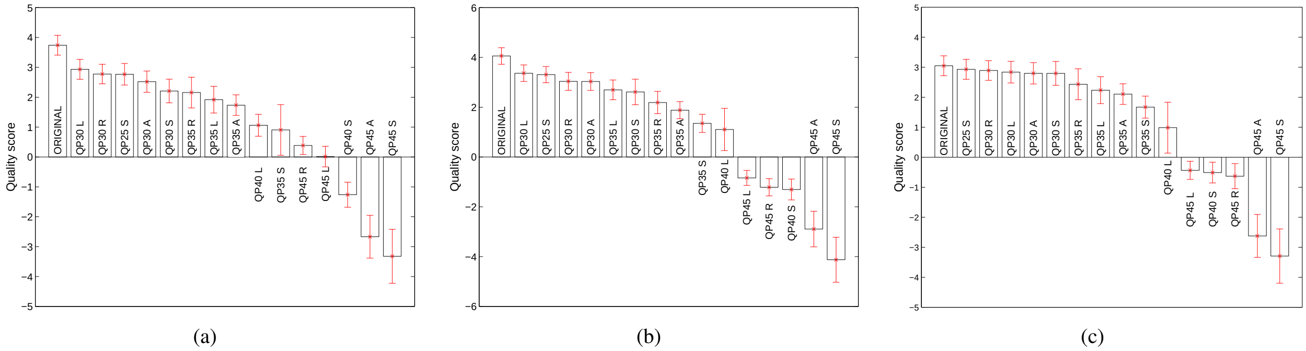


Fig. 5. Subjective test results for the sequence 3D_22. In the plots S = Symmetric encoding, A = Asymmetric encoding, R = Only right frames coded, L= Only left frames coded, QP = Quality parameter. (a) shows the results for sequence 3D_22, (b) shows the results for sequence 3D_27, and (c) shows the results for sequence 3D_28.

Group A: Figure 5 shows that the perceptual quality is slightly changed for low level of QPs (Group A). For understanding the quality difference is significant, the ANOVA test is performed. It tests the hypothesis that the samples are drawn from populations with the same mean against the alternative hypothesis that the population means are not all the same. The achieved result: p_{value} = 0.5849, indicates the perceptual quality is not significantly different. In other words, the perceptual quality of the asymmetric coded and alternate coded videos are comparable with the symmetric encoded videos, that is, we can exploit the same level of perceptual quality even for higher QPs by using the asymmetric coding.

Group B: As shown in Figure 5, from the achieved result from ANOVA test (p_{value} = 0), we can conclude that, for high values of QPs, the perceptual quality varies significantly.

TABLE II
PERCEPTUAL QUALITY DIFFERENCE ANALYSIS WITH RESPECT TO SYMMETRIC (SYM) CODING. THE $\simeq$ INDICATES THE PERCEPTUAL QUALITY IS NOT SIGNIFICANT DIFFERENT AND SIGNIFICANT DIFFERENCE ON THE PERCEPTUAL QUALITY IS INDICATED AS $\neq$.

	Group A (QP < 30)	*Group B (QP > 30)*
ASYM	$\simeq$	$\neq$
alternate	$\simeq$	$\neq$

The analysis of the achieved results is summarized in Table II. It shows that the asymmetric and/or alternate coding results similar perceptual quality as the one provided by symmetric encoding even at higher level of QPs; however, when QP value exceeds 30, the quality decreases significantly. Therefore, QP = 30 can be considered as a Just Noticeable Distortion (JND) threshold for asymmetric encoding.

2) 3D-HEVC encoded video: In general, the results achieved in Section V-A1 are confirmed for 3D-HEVC encoding as well. The perceptual quality of HRCs is shown in Figure **??** and the difference on collected opinion scores is analyzed with the help of significance analysis (ANOVA test: achieved results are shown in Table III). Figure **??** also shows that the perceptual quality at QP30 S is even higher than QP25 S, this is due to the use of test videos from different SRCs (content) (as mentioned in Section IV-B3). Since the content has a significant impact on the video QoE [89]. From Table III,

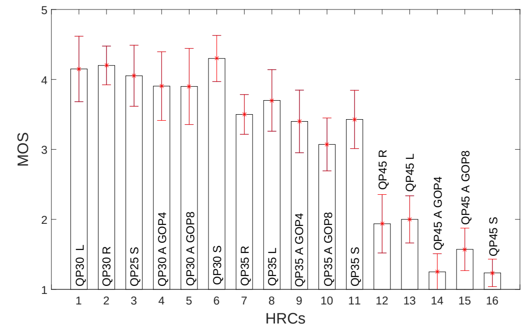


Fig. 6. Subjective test results. In the plots S = Symmetric encoding, A= Asymmetric encoding, R = Only right frames coded, L = Only left frames coded, QP = Quality parameter.

TABLE III
RESULTS OF ANOVA TEST. GROUP A1=[QP30 L, QP30 R, QP25 S, QP30 A GOP4, QP30 A GOP8], GROUP B1 = [QP30 S, QP35 L, QP35 R, QP35 A GOP4, QP35 A GOP8], AND GROUP B2 = [QP45 L, QP45 R, QP45 A GOP4, QP45 A GOP8, QP45 S]

Test Conditions	p_{value}	comment
Group A1	0.8610	$\simeq$
Group B1	0	$\neq$
Group B2	0	$\neq$
All HRCs	0	$\neq$

we can notice that the difference on the opinion scores for the test videos at lower QPs (Group A1) is not significant, and the contrary is true for group B1, B2, and all HRCs. Moreover, it is also noticed that in asymmetric encoding with alternation the GOP size does not evidence any significant variation in the perceptual quality.

It is worth mentioning that, as expected, the achieved results show that the perceptual quality of the symmetrically and asymmetrically encoded streams are significantly different, and thus, the quality scores of symmetrically coded video (QP25 S) and asymmetrically coded videos at higher compression level, QP30 L, QP30 R, and QP30 A, are not significantly different. This observation is in-line with the conclusion drawn in [57], but it contradicts the conclusion drawn in [90]: there is no significant difference between asymmetric and symmetric stereoscopic stimuli.

TABLE IV
RESULTS OF ANOVA ON THREE CATEGORIES: ALL OBSERVERS, RIGHT-EYE DOMINANT OBSERVERS, AND LEFT-EYE DOMINANT OBSERVERS.

SRCs	Test Conditions	p_{value}
3D_22	all HRCs	0.5991
	ASYM R	0.9984
	ASYM L	0.7616
3D_27	all HRCs	0.5464
	ASYM R	0.8775
	ASYM L	0.8782
3D_28	all HRCs	0.6445
	ASYM R	0.9487
	ASYM L	0.9179

An interesting observation concerns the ranking of the asymmetrically coded pairs with alternation, compared to the ranking of the symmetrically coded video sequences, when encoded with the same QP: asymmetrically coded pairs with alternation are always preferred over symmetrically coded video sequences. This can be explained by the fact that symmetrically coded video sequences induce a larger number of mismatching regions, compared to the asymmetrically coded pairs with alternation. This may lead to uncomfortable viewing conditions, in particular binocular rivalry [91] conditions. Indeed, binocular rivalry occurs when the two views are too different. This leads to a non-converging percept and thus to visual discomfort if the differences are too strong. In the *Alternate* coded video sequences, since one of the two views is always encoded with a good quality, it is possible that the masking effect occurs. Indeed, there is still one view with significantly less contrast than the other view. In this case, binocular suppression [92], [93] may occur, leading to a better overall comfort quality than that of the symmetrically coded pairs. In other words, the results show that binocular suppression involved by the asymmetrically coded pairs with alternation is less uncomfortable than binocular rivalry involved by symmetrically coded pairs, when considering the same QP.

B. Eye-Dominance Analysis

1) HEVC encoded video: We plot the preference scores for the left-eye dominant participants (Figures 7(a), 8(a), and 9(a)) and for the right-eye dominant participants (Figures 7(b), 8(b), and 9(b)). We expected to observe that the right-eye dominant participants would prefer the video sequences whose left view is more coarsely quantized and that the left-eye dominant participants would prefer the video sequences whose right view is ore coarsely quantized. However, this is not actually observed from the figures.

To obtain a more accurate observation, we ran statistical tests on the subjective scores. ANOVA was performed considering the three groups: the whole set of participants, the right-eye-dominant participants only, and the left-eye-dominant participants only, for three test conditions (all, ASYM R, and SSYM L videos). Table IV shows that the perceptual quality does not vary significantly to the users based on the eye-dominance.

The Table V shows that whatever the eye dominance, asymmetrically coded pairs are always significantly superior to symmetrically coded pairs. This is in agreement with the observation above: binocular suppression involved by the asymmetrically coded pairs with alternation coded video sequences is less uncomfortable than binocular rivalry involved by symmetrically coded pairs. No significant statistical difference was found between symmetrically coded pairs and asymmetrically coded pairs with alternation, neither between asymmetrically coded pairs and asymmetrically coded pairs with alternation.

We also investigated the difference of behaviors between left-eye dominant participants and right-eye dominant participants by considering the subjective scores provided by the two groups separately. A Student's t-test revealed that there was no statistical significant difference between the two groups. In other words, in this experiment no difference between left-eye dominant observers and right-eye dominant observers was proven.

TABLE V
RESULTS OF STUDENT'S T-TEST ON THE- LEGEND:↑: SUPERIOR, ↓: INFERIOR, O: STATISTICALLY EQUIVALENT. READING: LINE "1" IS STATISTICALLY SUPERIOR TO COLUMN "2". ASYM: ASYMMETRICALLY CODED PAIRS, SYM: SYMMETRICALLY, *Alternate*: ASYMMETRICALLY CODED PAIRS WITH ALTERNATION

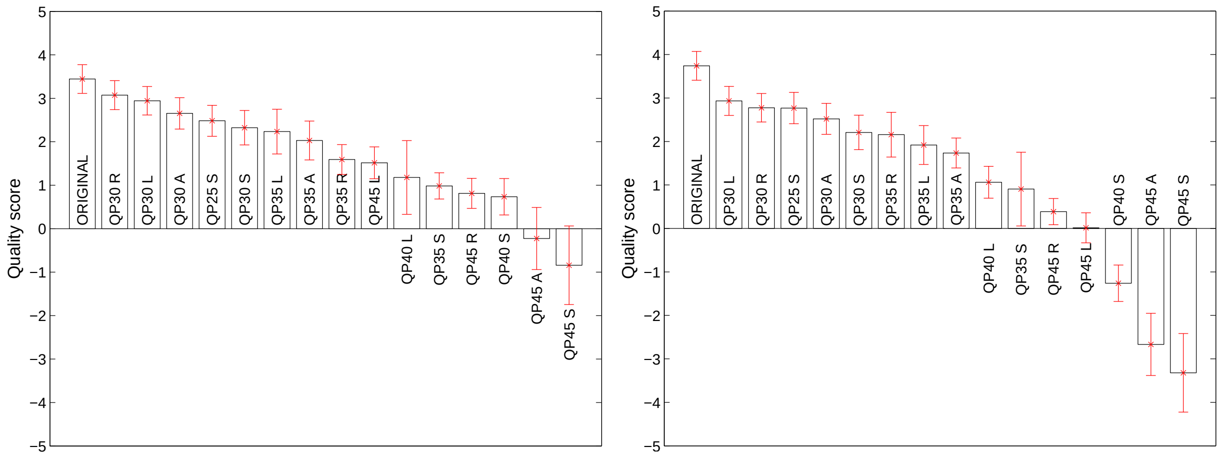
		ASYM	*SYM*	*Alternate*
all scores	*ASYM*		↑	**o**
	SYM	↓		**o**
	Alternate	**o**	**o**	
left-eye	*ASYM*		↑	**o**
	SYM	↓		**o**
	Alternate	**o**	**o**	
right-eye	*ASYM*		↑	**o**
	SYM	↓		**o**
	Alternate	**o**	**o**	

2) 3D-HEVC encoded video: The results presented in Section V-B1: the perceptual quality of the 3D-HEVC encoded videos are not varies noticeably to the left/right-eye dominance observers, is also confirmed for 3D-HEVC encoded videos. Moreover, to confirm the result, a statistical test (ANOVA) is performed for the scores given by all observers, left-eye dominance observer, and right-eye dominance observer. The achieved results shown in Table VI, indicate that the difference on perceptual quality is not significant based on the eye dominance.

It is worth mentioning that a similar conclusion, the impact of eye dominance in the perception of asymmetrically compressed stereoscopic video is insignificant, was reached in [11], [94]. Our observation is consistent with the concept

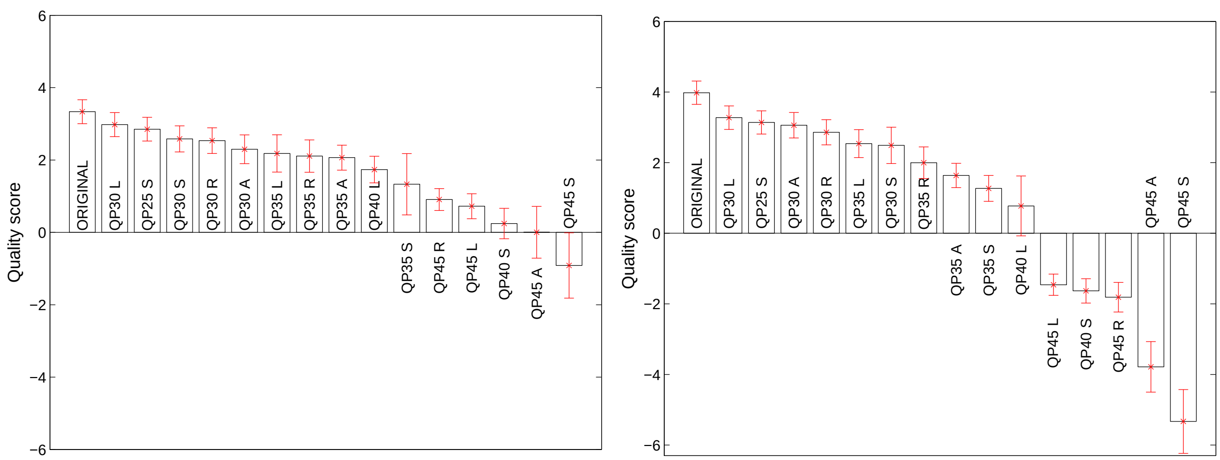
TABLE VI
RESULTS OF ANOVA ON THREE CATEGORIES: ALL OBSERVERS, RIGHT-EYE DOMINANT OBSERVERS, AND LEFT-EYE DOMINANT OBSERVERS.

Test Conditions	p_{value}
all HRCs	0.9876
ASYM R	0.8577
ASYM L	0.9990



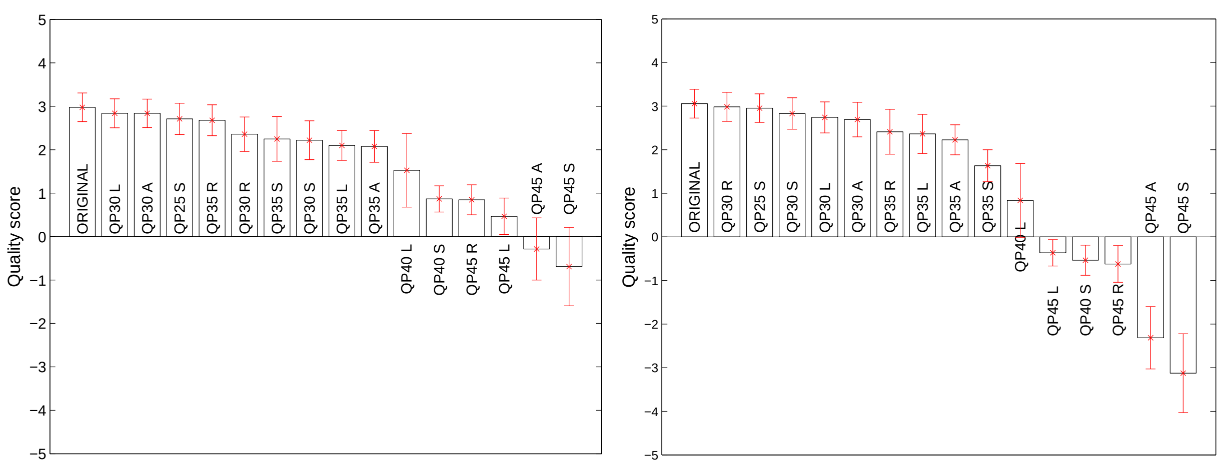
(a) (b)

Fig. 7. Subjective test results for the sequence 3D_22. In the plots S= Symmetric encoding, A= Asymmetric encoding, R= Only right frames coded, L= Only left frames coded, QP= Quality parameter. (a) shows the results for the left eye dominant group and (b) shows the results for the right eye dominant group.



(a) (b)

Fig. 8. Subjective test results for the sequence 3D_27: (a) shows the results for the left eye dominant group and (b) shows the results for the right eye dominant group.



(a) (b)

Fig. 9. Subjective test results for the sequence 3D_28: (a) shows the results for the left eye dominant group and (b) shows the results for the right eye dominant group.

of binocular rivalry that is widely accepted in the field of visual science [95].

C. Preference of Experience Analysis

As further analysis, we have analyzed the behavior of PoE [12] versus the bit rate achieved by using the different compression schemes. The PoE concept has been introduced to specify the subjective scores collected by means of pair comparison based test. Basically, it represents the preference of the QoE of the observers between two videos rather than an absolute scale value for each sequence.

As mentioned in Section IV-B3, as a limitation of the second experiment, all 3D-HEVC encoded test videos (9×16) are not evaluated by all the subjects, and thus the rate distortion analysis is performed only for HEVC encoded videos.

For the analysis, the bit rate (BR) is computed. In case of alternate coding scheme, the BR has been evaluated as the average bit rate among the two considered videos. The results are shown in Figures 10 to 12. The plots show the scores collected by considering all the subjects, the ones collected by subjects belonging to the left dominant eye class, and those obtained by considering the right dominant class. For better understanding, it is useful to analyze the results by considering three bit rate regions corresponding to low (0 Mbit/s - 1.5 Mbit/s), medium (1.5 Mbit/s - 2.5 Mbit/s), and high (2.5 Mbit/s - 3.5 Mbit/s) bit rate values, represented in the figures by vertical dotted lines (at 1.5 Mbit/s and at 2.5 Mbit/s). It can be noticed that:

- low bit rate (<1.5 Mbit/s): in this interval, for all the considered sequences, the *Alternate* coding scheme shows higher values of PoE than the SYM one;
- medium bit rate (>1.5 Mbit/s and <2.5 Mbit/s): for all the considered sequences starting from 1.5 Mbit/s, the ASYM scheme (both left and right) shows higher PoE score with respect to the *Alternate* one;
- high bit rate (>2.5 Mbit/s): the coding scheme to be used is the SYM.

Based on the performed analysis we can conclude that:

- the difference between ASYM L and ASYM R for comparable bit rate values is very small, so there is not a considerable difference in the use of one of the two approaches as demonstrated by the performed Student's t-tests;
- the *Alternate* scheme presents high subjective scores for small bit rates and for this reason it is more convenient to use the ASYM option that in the overall shows better performances;
- the SYM option seems to be more suitable for higher bit rates.

VI. Towards a guideline

Based on the analysis and discussion presented before the following conclusions can be drawn, and it can be used as guideline, for further study.

- Subjective quality of asymmetric encoded video is always higher than symmetric encoded video. However, if higher values of QPs (QP $>$ 30) are used, the quality decrease significantly. That is, the advantage of asymmetric encoding can be exploited only for low level of QPs ($\approx$ QP $<$ 30). Moreover, the asymmetric coding with alternate scheme also provides the same level of quality as of asymmetric coding. During the study, alternate coding was performed using different number of frames, but the quality was the same.
- The impact of eye dominance is not significant in the performed tests: symmetrically, asymmetrically, and alternate encoded videos.
- The concept of the measure of preference of experience has been recommended for the subjective quality assessment of stereoscopic content.

VII. Conclusion

This paper presents a new 3D stereoscopic video sequences database with associated rates of visual discomfort. Among the numerous sources leading to visual discomfort, the study of binocular rivalry has been targeted with a focus on coding asymmetries. Asymmetrically coded pairs, symmetrically coded pairs and asymmetrically coded pairs with quality alternation on both views have been included in the database. It appears that generally, observers felt more comfortable when watching asymmetrically coded video sequences than symmetrically coded video sequences. Another observation concerns the fact that observers provided higher preference rates for asymmetrically coded video sequences with alternation than for symmetrically coded video sequences, when using considering the same quantization parameter. Moreover, the achieved results are equally applicable for HEVC and 3D-HEVC encoded videos. The results were confirmed by exploiting two subjective quality assessment protocols: absolute category rating and forced choice pair comparison.
This database will be useful for further understanding of human visual system and for the investigation of new quality assessment models for detecting visual discomfort conditions in stereoscopic sequences.

References

[1] M. Domanski, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media mpeg-i: 360 video, virtual navigation and beyond," in *Int. Conference on Systems, Signals and Image Processing (IWSSIP)*, May 2017, pp. 1–9.

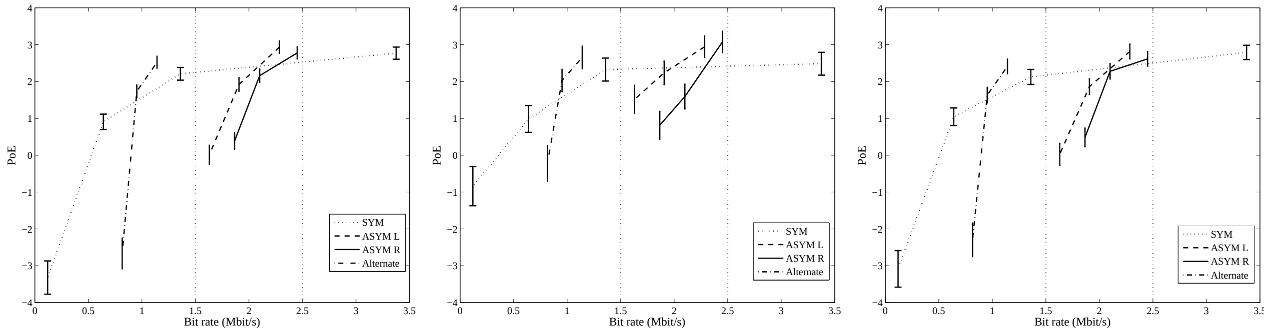
[2] D. Lovell, *Immersive media, virtual reality and a new form of media @ONLINE*, Jun. 2016. [Online]. Available: https://www.iabuk.net/blog/immersive-media-virtual-reality-and-a-new-form-of-media#X82Cq7PSVClhu0zK.99

[3] D. Katzmaier, "Shambling corpse of 3D TV finally falls down dead," *Online(Accessed 15/03/2017): https://www.cnet.com/news/shambling-corpse-of-3d-tv-finally-falls-down-dead/*, 2017.

[4] C. Hewage and M. Martini, "Quality of experience for 3D video streaming," *Communications Magazine, IEEE*, vol. 51, no. 5, pp. 101–107, May 2013.

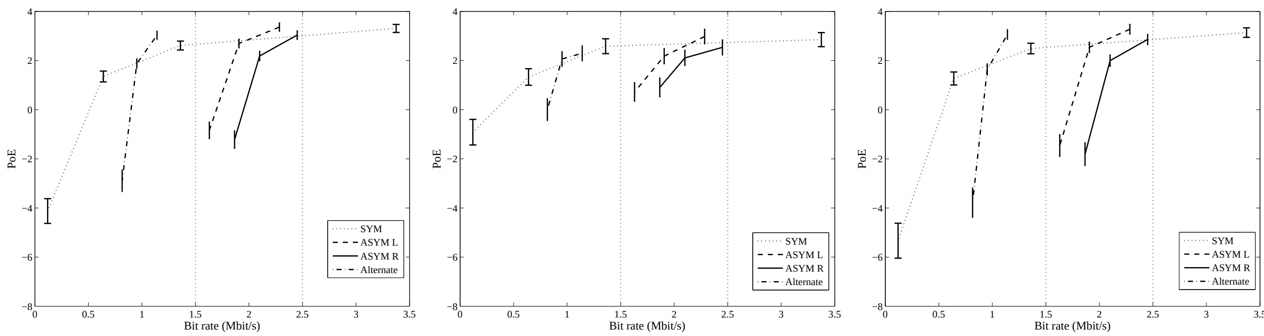
[5] M. Barkowsky, K. Wang, R. Cousseau, K. Brunnstrom, R. Olsson, and P. Le Callet, "Subjective quality assessment of error concealment strategies for 3DTV in the presence of asymmetric transmission errors," in *Packet Video Workshop (PV), 2010 18th Int.*, Dec 2010, pp. 193–200.

[6] I. Recommendation, "ITU-R BT.2021: Subjective methods for the assessment of stereoscopic 3dtv systems," *International Telecommunication Union, Geneva, Switzerland*, 2012.



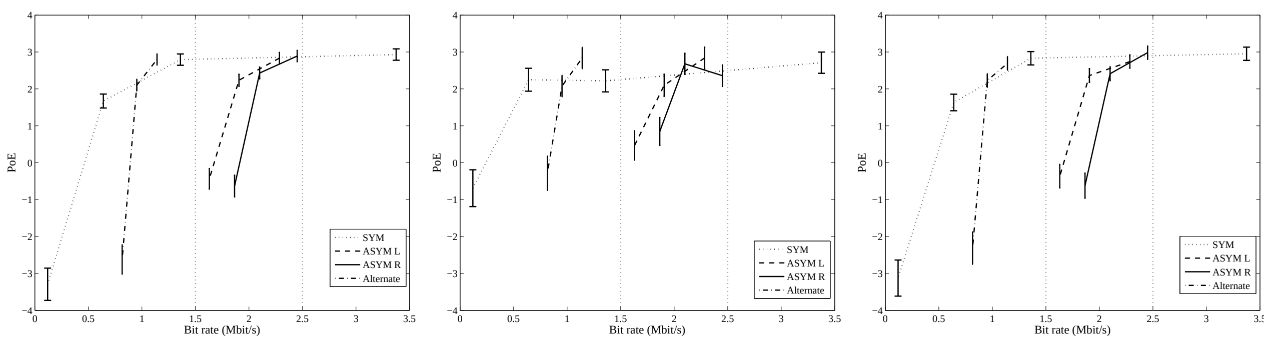
(a) Whole participant set (b) Left-eye dominant subset (c) Right-eye dominant subset

Fig. 10. Results for the sequence 22: a) average PoE computed for all participants to the test, b) PoE computed for left-eye dominant subset, c) PoE computed for right-eye dominant subset.



(a) Whole participant set (b) Left-eye dominant subset (c) Right-eye dominant subset

Fig. 11. Results for the sequence 27: a) average PoE computed for all participants to the test, b) PoE computed for left-eye dominant subset, c) PoE computed for right-eye dominant subset.



(a) Whole participant set (b) Left-eye dominant subset (c) Right-eye dominant subset

Fig. 12. Results for the sequence 28: a) average PoE computed for all participants to the test, b) PoE computed for left-eye dominant subset, c) PoE computed for right-eye dominant subset.

[7] G. Saygili, C. Gurler, and A. Tekalp, "Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 593–601, 2011.

[8] V. De Silva, H. Arachchi, E. Ekmekcioglu, A. Fernando, S. Dogan, A. Kondoz, and S. Savas, "Psycho-physical limits of interocular blur suppression and its application to asymmetric stereoscopic video delivery," in *Packet Video Workshop (PV), 2012 19th Int.*, 2012, pp. 184–189.

[9] M. Azimi, S. Valizadeh, X. Li, L. E. Coria, and P. Nasiopoulos, "Subjective study on asymmetric stereoscopic video with low-pass filtered slices," in *Computing, Networking and Communications (ICNC), 2012 International Conference on*. IEEE, 2012, pp. 719–723.

[10] G. Saygılı, C. G. Gürler, and A. M. Tekalp, "Quality assessment of asymmetric stereo video coding," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 4009–4012.

[11] J. Wang, S. Wang, and Z. Wang, "Asymmetrically compressed stereoscopic 3d videos: Quality assessment and rate-distortion performance evaluation," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1330–1343, 2017.

[12] J. Li, M. Barkowsky, and P. Le Callet, "Subjective assessment methodology for preference of experience in 3DTV," in *IVMSP Workshop, 2013 IEEE 11th*, June 2013, pp. 1–4.

[13] J. Li, O. Kaller, F. De Simone, J. Hakala, D. Juszka, and P. Le Callet, "Cross-lab study on preference of experience in 3DTV: Influence from display technology and test environment," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth Int. Workshop on*, July 2013, pp. 46–47.

[14] P. Aflaki, M. Hannuksela, and M. Gabbouj, "Subjective quality assessment of asymmetric stereoscopic 3d video," *Signal, Image and Video Processing*, pp. 1–15, 2013.

[15] Y. Chang and M. Kim, "Binocular suppression-based stereoscopic video coding by joint rate control with kkt conditions for a hybrid video codec system," *IEEE Transactions on Circuits and Systems for Video*

Technology, vol. 25, no. 1, pp. 99–111, 2015.

[16] T. Na, S. Ahn, H. Sabirin, M. Kim, B. Kim, S. Hahm, and K. Lee, "A hybrid stereoscopic video coding scheme based on mpeg-2 and hevc for 3dtv services," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 9, pp. 1542–1554, 2013.

[17] A. Vetro, A. Tourapis, K. Muller, and T. Chen, "3DTV content storage and transmission," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 384–394, June 2011.

[18] ITU, "H.264 advanced video coding for generic audiovisual services," Int. Telecommunication Union, Recommendation, 2009.

[19] L. Christodoulou, L. M. Mayron, H. Kalva, O. Marques, and B. Furht, "3D TV using MPEG-2 and H.264 view coding and autostereoscopic displays," in *ACM Multimedia Conference*, Santa Barbara, CA, USA, October 2006.

[20] M. Perkins, "Data compression of stereopairs," *Communications, IEEE Transactions on*, vol. 40, no. 4, pp. 684–696, Apr 1992.

[21] M. M. Hannuksela, D. Rusanovskyy, W. Su, L. Chen, R. Li, P. Aflaki, D. Lan, M. Joachimiak, H. Li, and M. Gabbouj, "Multiview-video-plus-depth coding based on the advanced video coding standard." *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3449–3458, 2013.

[22] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *Image Processing, 2007. ICIP 2007. IEEE Int. Conference on*, vol. 1, Sept 2007, pp. I – 201–I – 204.

[23] F. Battisti, E. Bosc, M. Carli, P. Le Callet, and S. Perugia, "Objective image quality assessment of 3d synthesized views," *Signal Processing: Image Communication*, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596514001453

[24] "Committee draft of ISO/IEC 23002-3 auxiliary video data representations, WG 11 Doc. N8038, ISO/IEC JTC 1/SC 29/WG 11, montreux, switzerland, apr. 2006," April 2006.

[25] B. Julesz, *Foundations of Cyclopean Perception*. The University of Chicago Press, 1971.

[26] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, and J. Kim, "Asymmetric coding of stereoscopic video for transmission over T-DMB," in *3DTV Conference*, May 2007, pp. 1–4.

[27] Y. Termin, G. A. Kaminka, S. Semo, and A. Z. Zivotofsky, "Color stereoscopic images require only one color image," *Optical Engineering*, vol. 46, no. 8, 2007.

[28] Y. Chen, Y.-K. Wang, M. M. Hannuksela, and M. Gabbouj, "Picture-level adaptive filter for asymmetric stereoscopic video," in *ICIP'08*, 2008, pp. 1944–1947.

[29] A. Jain, A. Robinson, and T. Nguyen, "Comparing perceived quality and fatigue for two methods of mixed resolution stereoscopic coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 3, pp. 418–429, March 2014.

[30] F. Shao, G. Jiang, M. Yu, and J. You, "Asymmetric multi-view video coding based on chrominance reconstruction," in *Multimedia and Expo (ICME), IEEE Int. Conference on*, July 2010, pp. 962–967.

[31] F. Shao, G. Jiang, X. Wang, M. Yu, and K. Chen, "Stereoscopic video coding with asymmetric luminance and chrominance qualities," *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 4, pp. 2460–2468, 2010.

[32] F. Shao, G. Jiang, M. Yu, K. Chen, and Y.-S. Ho, "Asymmetric coding of multi-view video plus depth based 3d video for view rendering." *IEEE Transactions on Multimedia*, vol. 14, no. 1, pp. 157–167, 2012.

[33] W. J. Tam, L. B. Stelmach, and S. Subramaniam, "Stereoscopic video: asymmetrical coding with temporal interleaving," in *Proc. SPIE*, vol. 4297, 2001, pp. 299–306.

[34] A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G. Akar, R. Civanlar, and A. Tekalp, "Temporal and spatial scaling for stereoscopic video compression," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2006.

[35] I.-L. Jung, T. Chung, K. Song, and C.-S. Kim, "Efficient stereo video coding based on frame skipping for real-time mobile applications," *Consumer Electronics, IEEE Transactions on*, vol. 54, no. 3, pp. 1259–1266, August 2008.

[36] S. Liu, F. Liu, J. Fan, and H. Xia, "Asymmetric stereoscopic video encoding algorithm based on subjective visual characteristic," in *Wireless Communications Signal Processing (WCSP) Int. Conference on*, Nov 2009, pp. 1–5.

[37] A. Jain, C. Bal, A. Robinson, D. MacLeod, and T. Nguyen, "Temporal aspects of binocular suppression in 3D video," in *Sixth Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Jan. 2012, pp. 93–98.

[38] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, "Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation," *ACM Trans. Appl. Percept.*, vol. 3, no. 2, pp. 95–109, Apr. 2006.

[39] S. Fezza, M.-C. Larabi, and K. Faraoun, "Asymmetric coding using binocular just noticeable difference and depth information for stereoscopic 3d," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE Int. Conference on*, May 2014, pp. 880–884.

[40] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *Signal Processing Conference, 2007 15th European*. IEEE, 2007, pp. 2110–2114.

[41] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis," in *Proc. of International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA*, 2010.

[42] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Using disparity for quality assessment of stereoscopic images," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 389–392.

[43] J. Yang, C. Hou, Y. Zhou, Z. Zhang, and J. Guo, "Objective quality assessment method of stereo images," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009*. IEEE, 2009, pp. 1–4.

[44] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3379–3391, 2013.

[45] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 591–602, 2014.

[46] Z. Chen, W. Zhou, and W. Li, "Blind stereoscopic video quality assessment: From depth perception to overall experience," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 721–734, 2018.

[47] C. Galkandage, J. Calic, S. Dogan, and J.-Y. Guillemaut, "Stereoscopic video quality assessment using binocular energy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 102–112, 2017.

[48] K. Wang, M. Barkowsky, K. Brunnstrom, M. Sjostrom, R. Cousseau, and P. Le Callet, "Perceived 3d TV transmission quality assessment: Multi-laboratory results using absolute category rating on quality of experience scale," *Broadcasting, IEEE Transactions on*, vol. 58, no. 4, pp. 544–557, Dec 2012.

[49] ITU-T, ""subjective assessment of stereoscope television pictures," in recommendation ITU-R BT.1438," 2000.

[50] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "Methodologies for assessing 3D QoE: Standards and explorative studies," *ZTE Communications*, vol. special issue "QoE Modeling and Applications for Multimedia Systems, no. 1, 2013.

[51] ——, "Quality of experience model for 3DTV," in *Proc. SPIE*, vol. 8288, 2012, pp. 82 881P–82 881P–9.

[52] M. Urvoy, M. Barkowsky, and P. Le Callet, "How visual fatigue and discomfort impact 3DTV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors," *Annals of Telecommunications*, vol. 68, no. 11-12, pp. 641–655, 2013.

[53] C. W. Clifford, M. A. Webster, G. B. Stanley, A. A. Stocker, A. Kohn, T. O. Sharpee, and O. Schwartz, "Visual adaptation: Neural, psychological and computational aspects," *Vision Research*, vol. 47, no. 25, pp. 3125 – 3131, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0042698907003756

[54] J. Wang and Z. Wang, "Perceptual quality of asymmetrically distorted stereoscopic images: the role of image distortion types," in *Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2014.

[55] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013.

[56] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[57] V. De Silva, H. K. Arachchi, E. Ekmekcioglu, and A. Kondoz, "Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video," *IEEE transactions on image processing*, vol. 22, no. 9, pp. 3392–3404, 2013.

[58] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutiérrez, and N. Garcia, "Nama3ds1-cospad1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3d stereoscopic sequences," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth Int. Workshop on*. IEEE, 2012, pp. 109–114.

[59] E. Dumić, S. Grgić, K. Šakić, P. M. R. Rocha, and L. A. da Silva Cruz, "3d video subjective quality: a new database and grade comparison study," *Multimedia tools and applications*, pp. 1–23, 2016.

[60] A. Banitalebi-Dehkordi, M. T. Pourazad, and P. Nasiopoulos, "The effect of frame rate on 3d video quality and bitrate," *3D Research*, vol. 6, no. 1, p. 1, 2015.

[61] S. Jumisko-Pyykkö, T. Haustola, A. Boev, and A. Gotchev, "Subjective evaluation of mobile 3d video content: depth range versus compression artifacts," in *Proc. SPIE*, vol. 7881, 2011, p. 78810C.

[62] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," in *Proceedings of SPIE*, vol. 7526, no. MMSPL-CONF-2009-021, 2010.

[63] ITU-R Rec. BT.2021, "Subjective methods for the assessment of stereoscopic 3dtv systems," 2015.

[64] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk, "Comparison of four subjective methods for image quality assessment," in *Computer Graphics Forum*, vol. 31, no. 8. Wiley Online Library, 2012, pp. 2478–2491.

[65] T. Kawano, K. Yamagishi, and T. Hayashi, "Performance comparison of subjective assessment methods for 3d video quality," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth Int. Workshop on*. IEEE, 2012, pp. 218–223.

[66] ITU-T rec. P.910, "Subjective video quality assessment methods for multimedia applications," 2008.

[67] T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi, "Performance comparisons of subjective quality assessment methods for mobile video," in *Quality of multimedia experience (QoMEX), 2010 second international workshop on*. IEEE, 2010, pp. 82–87.

[68] J.-S. Lee, "On designing paired comparison experiments for subjective multimedia quality assessment," *IEEE Transactions on Multimedia*, vol. 2, no. 16, 2014.

[69] J.-S. Lee, L. Goldmann, and T. Ebrahimi, "Paired comparison-based subjective quality assessment of stereoscopic images," *MULTIMEDIA TOOLS AND APPLICATIONS*, vol. 67, no. 1, pp. 31–48, NOV 2013.

[70] M. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," in *SPIE Video Communications and Image Processing Conference, Lugano, Switzerland*, 2003.

[71] Rec. ITU-R BT.500, "Methodology for the subjective assessment of the quality of television pictures," 2012.

[72] P. Paudyal, F. Battisti, and M. Carli, "Evaluation of the effects of transmission impairments on perceived video quality by exploiting ReTRiEVED dataset," *Journal of Electronic Imaging*, vol. 26 (2), p. 023003, 2017.

[73] E. Cheng, P. Burton, J. Burton, A. Joseski, and I. Burnett, "RMIT3DV: pre-announcement of a creative commons uncompressed HD 3D video database," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth Int. Workshop on*, 2012, p. 212?217.

[74] G. J. Sullivan, J.-R. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard." *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[75] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.

[76] ITU-T Rec. J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," 2008.

[77] S. Gould, J. Arfvidsson, A. Kaehler, B. Sapp, M. Meissner, G. Bradski, P. Baumstarck, S. Chung, and A. Y. Ng, "Peripheral-foveal vision for real-time object recognition and tracking in video," in *Proceedings of the Twentieth Int. Joint Conference on Artificial Intelligence (IJCAI-07)*, 2007.

[78] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, no. 6582, pp. 520–?522, 1996.

[79] K. Amano, N. Goda, S. Nishida, Y. Ejima, T. Takeda, and Y. Ohtani, "Estimation of the timing of human visual perception from magnetoencephalography," *The Journal of neuroscience*, vol. 26, no. 15, p. 3981?3991, 2006.

[80] J. Li, M. Barkowsky, and P. Le Callet, "Boosting paired comparison methodology in measuring visual discomfort of 3DTV: performances of three different designs," in *Proceedings of the SPIE Electronic Imaging, Stereoscopic Displays and Applications, Human Factors 2013*, 2013, pp. 1–12.

[81] ——, "Analysis and improvement of a paired comparison method in the application of 3DTV subjective experiment," in *Image Processing (ICIP), IEEE Int. Conference on*, 2012, p. 629?632.

[82] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, p. 324?345, 1952.

[83] F. Wickelmaier and C. Schmid, "A matlab function to estimate choice model parameters from paired-comparison data," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 1, pp. 29–40, 2004.

[84] M. Barreda-Ángeles, R. Pépion, E. Bosc, P. Le Callet, and A. Pereda-Baños, "Exploring the effects of 3d visual discomfort on viewers' emotions," in *Image Processing (ICIP), 2014 IEEE Int. Conference on*. IEEE, 2014, pp. 753–757.

[85] 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 Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 35–49, 2016. [Online]. Available: {https://hevc.hhi.fraunhofer.de/3dhevc}

[86] B. W. Keelan and H. Urabe, "Iso 20462: a psychophysical image quality measurement standard," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2003, pp. 181–189.

[87] P. Paudyal, F. Battisti, and M. Carli, "Impact of video content and transmission impairments on quality of experience," *Multimedia Tools and Applications*, vol. 75, no. 23, pp. 16 461–16 485, 2016.

[88] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[89] P. Paudyal, F. Battisti, and M. Carli, "Study of the effects of video content on quality of experience," in *SPIE/IS&T Electronic Imaging*. Int. Society for Optics and Photonics, 2015, pp. 93 960W–93 960W.

[90] P. Aflaki, M. M. Hannuksela, J. Häkkinen, P. Lindroos, and M. Gabbouj, "Subjective study on compressed asymmetric stereoscopic video," in *Image Processing (ICIP), 2010 17th IEEE Int. Conference on*. IEEE, 2010, pp. 4021–4024.

[91] L. Lei and C. M. Schor, "The spatial properties of binocular suppression zone," *Vision research*, vol. 34, no. 7, p. 937?947, 1994.

[92] H. Ono, R. Angus, and P. Gregor, "Binocular single vision achieved by fusion and suppression," *Perception & Psychophysics*, vol. 21, no. 6, p. 513?521, 1977.

[93] S. B. Steinman, B. A. Steinman, and R. P. Garzia, *Foundations of binocular vision*. McGraw-Hill, 2000.

[94] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3d images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.

[95] R. Blake, "A primer on binocular rivalry, including current controversies," *Brain and mind*, vol. 2, no. 1, pp. 5–38, 2001.