**Source:** **China Mobile Com. Corporation**

**Title: [FS\_Beyond2D] Quality aspects of stereoscopic video content**

**Agenda Item: 9.9**

**Document for: Discussion and Agreement**

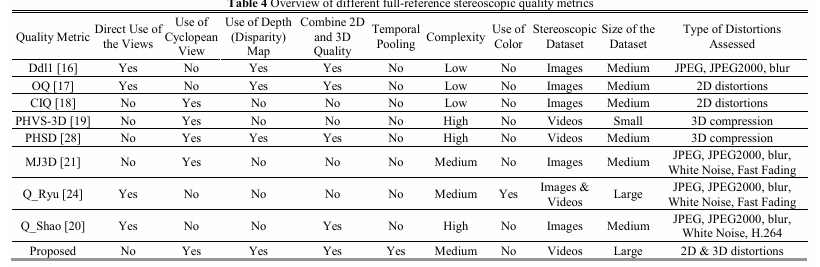
# Introduction

This document provides the quality evaluation aspects of stereoscopic video content and is applicable to Scenario 5: *UE-to-UE Beyond 2D Video Streaming*, of the FS\_Beyond2D study [1].

# Proposed Updates

### 2.1 Overview of Different Full Reference Sterescopic Quality Metrics

In general, the quality of stereoscopic 3D video includes the spatial-temporal visual quality, depth perception and visual comfort. The follow table provides an overview of existing stereoscopic quality metrics.



The Mean Square Error (MSE) is used as a quality measure, which is known to have low performance in accurately representing the human visual system. Moreover, this method utilizes disparity information instead of the actual depth map, which may result in inaccuracies in the case of occlusions.

Similarly, in the study by Jin et al. the quality of cyclopean view was taken into account in the design of a full-reference 3D quality metric for mobile applications, which is called PHVS-3D [X1]. In this study the information of the left and right channels are fused using the 3D DCT transform to generate a cyclopean view. Then, a map of local block dis-similarities between the reference and distorted cyclopean views is estimated using the MSE of block structures. The weighted average of this map is used as the PHVS-3D quality index. Although the proposed schemes take into account the quality of cyclopean view, they ignore the depth effect of the scene.

### 2.2 Human-Visual-System-based 3D (HV3D) Quality Metric

A full-referenced human visual-system-based quality metric for 3D videos called HV3D had been proposed in ITU-T [3]. It takes into account the quality of individual views, the quality of the cyclopean view (fusion of the right and left view, what the viewer perceives), as well as the quality of the depth information.

The HV3D quality metric is taking values between 0 and 1 (because it is optimized to be correlated with MOS/10), and higher than 1 in case quality is improved.

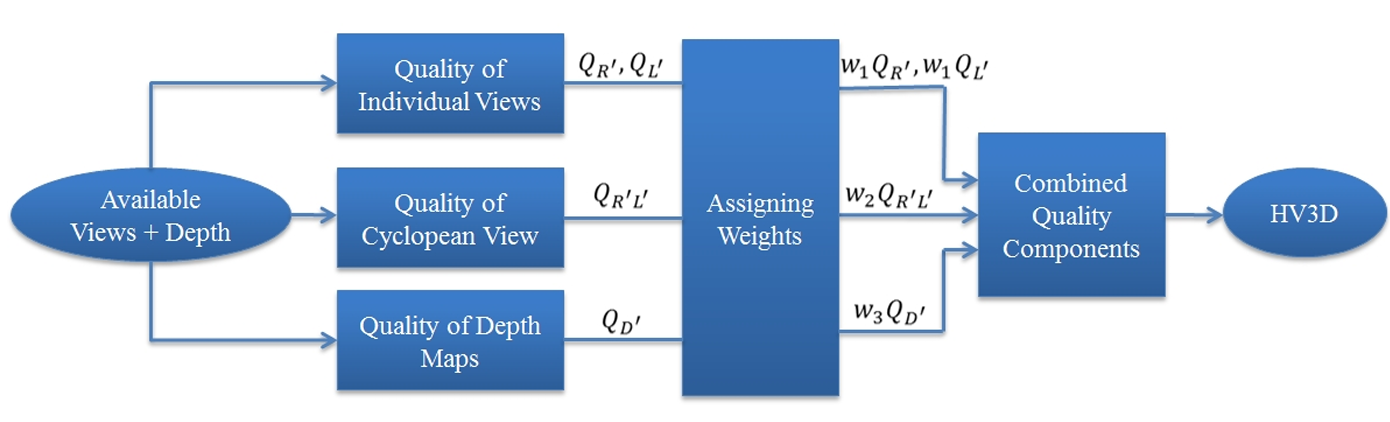


Figure 2.2-1 HV3D Flowchart

The performance of the proposed HV3D metric is validated by subjective tests, using 4 reference and 40 modified videos and 19 subjects, following the ITU-R BT.500-11 recommendation [x2].

Performance evaluations showed that HVS 3D quality metric quantifies quality degradation caused by several representative types of distortions very accurately, with Pearson correlation coefficient of 90.8 %, a competitive performance compared to the state-of-the-art 3D quality metrics. (Ref: <https://arxiv.org/pdf/1803.04832>)

#### 2.2.1 Quality of individual views

The metrics (see below) are computed for the quality of individual views that form the stereo pair. The quality of the distorted right view with respect to its matching reference view is calculated as followings. The quality of the left view is calculated in the same fashion..

#### 

Where *YR* and *YR’*are luma information of the reference and distorted right views respectively, *UR* and *VR* are the chroma information of the reference right-view, *UR’*and *VR’*are the chroma information of the distorted right-view, *w1*and *w4* are weighting constants.

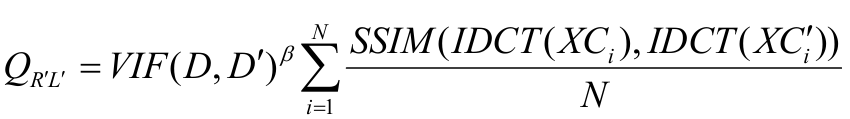
#### 2.2.2 Quality of cyclopean view

In order to measure the quality of the cyclopean view, first the cyclopean view is constructed by combining the corresponding areas from the left and right views.This is done by finding the matching blocks between right and left views. To this end, the luma information of each view is divided into 16×16 blocks. Here the 3D-DCT transform is applied to each pair of the matching blocks (left and right views) to generate two 16×16 DCT-blocks, which contain the DCT coefficients of the fused blocks. Since the human visual system is more sensitive to the low frequencies of the cyclopean view, only the first level of the coefficients is considered (which is a 16×16 DCT-block) and the rest of them are discarded.



Where *XC* is the cyclopean-view model for a pair of matching blocks in the right and left views, *Xi,j* are the low frequency 3D-DCT coefficients of the fused view, *i* and *j* are the horizontal and vertical indices of coefficients, and *Ci,j* is the CSF modeling mask *Ci,j* is derived based on the JPEG quantization table (see[X3] for more details).

Once the cyclopean-view model for all the blocks within the distorted and reference 3D views is obtained, the quality of the cyclopean view is calculated as follows:



Where *D* is the depth map of the reference 3D view, *D'* is the depth map of the distorted 3D view, *XCi*is the cyclopean-view model for the *ith* matching block pair in the reference 3D view,*XC'i* is the cyclopean-view model for the *ith* matching block pair in the distorted 3D view, *IDCT* stands for inverse 2D discrete cosine transform,*N* is the total number of blocks in the each view, *β* is a constant, empirically assigned to 0.7 (resulted from a series of subjective tests).

#### 2.2.3 Quality of depth map

The quality of depth map is formulated as follows:



Where di is the variance of block *i* in the depth map of the 3D reference view and the local disparity variance is calculated over a block size area of 64x64.

[2.2.4 Weighting constants

Weighting constants are found using least mean square such that the difference between the mean opinion scores (MOS) values and the HV3D values is minimized as follows:



The authors [X] has conduct experiment to to determine the best values for the weighting constants *w1, w2, w3*and *w4* which result in the minimum mean of square errors between our HV3D index and the MOS.

TABLE X shows the resulting values for these constants from their experiment.



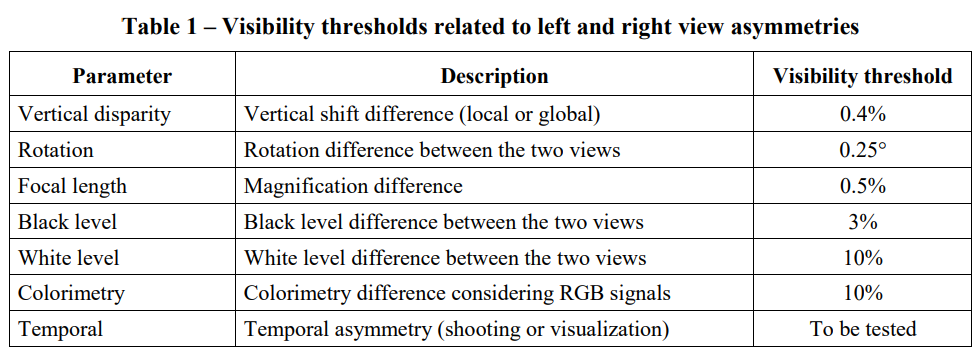
]

### 2.3 Subjective evaluation

1. Subjective Assessment Methods for 3D Video Quality, document ITU-T P.3D-sam, International Telecommunication Union, Geneva,Switzerland, Jul. 2015.

*“In stereo 3D systems, a binocular 3D image is formed by presenting the left and right image to their respective eye. If discrepancies arise between these two images, they can cause psychophysical stress, and in some cases 3D viewing can fail. For example, when shooting and displaying stereoscopic 3DTV programmes, there may be geometrical, optical, electrical or temporal asymmetries, such as size inconsistency, vertical shift, rotation error, and luminance or colour levels between the left and right images. For the production of natural scene content using two independent video cameras, the main issue is to guarantee that the asymmetries of the views are under perceptual limits.”*

*Table 1 illustrates visibility thresholds obtained from subjective experiments using an impairment scale and for a viewing distance of 4.5 times the display height.*



1. Assessment Methods of Visual Fatigue and Safety Guideline for 3D Video, document ITU-T J.3D-fatigue, International Telecommunication Union, Geneva, Switzerland, 2015.

# Proposal

We propose to document section 2 to PD as the methodology for evaluating the quality stereoscopic 3D video content.

# References

[1] 3GPP TR 26.956 v0.0.2 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Evaluation and Characterization of Beyond 2D Video Formats and Codecs (Release 19), S4-240947, Jeju (Korea), May 2024.

[2] Xie, Junyuan et al. “Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks.” European Conference on Computer Vision (2016).

[3] ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 - JCT3V-C0032: A human visual system based 3D video quality metric

[4] ITU-R Recommendation BT.2021, “Subjective methods for the assessment of stereoscopic 3DTV systems,” Aug. 2012.

[X1] L. Jin, A. Boev, A. Gotchev, K. Egiazarian, “3D-DCT based perceptual quality assessment of stereo video,” 18th International Conference on Image Processing, ICIP, 2011.

[X2] Recommendation ITU-R BT.500-11, “Methodology for the subjective assessment of the quality of the television pictures”.

[X3] A. Banitalebi-Dehkordi, Mahsa T. Pourazad, P. Nasiopoulos, “A human visual system based 3D video quality metric,” 2nd International Conference on 3D Imaging, IC3D, Dec. 2012, Liege, Belgium.