**3GPPSA4 131-bis-e MeetingS4-250568**

Online, 11 – 17 April 2025

**Source: China Mobile**

**Title: [FS\_Beyond2D] Objective Metric for Generated Stereoscopic Video**

**Agenda item: 9.7**

**Document for: Agreement**

**1. Introduction**

This document extends the current objective evaluation metrics in TR 26.956 Section 7.2.5, which currently applies only to captured stereoscopic sequences. The proposed content specifically addresses quality assessment of AI-generated stereoscopic video sequences in UE-to-UE Live Streaming Scenario.

**2. Proposal**

It is proposed to agree the following changes to the 3GPP draft TR 26.926 V0.3.0.

\* \* \* First Change (all new Text)\* \* \*

### 7.2.5 Performance Metrics

#### 7.2.5.1 Objective Metrics

##### 7.2.5.2 Objective Metric for Generated Stereoscopic Video

Current AI-based stereoscopic video generation methods often experience artifacts such as edge sharpness mismatch, cardboarding effects and crosstalk. These artifacts are praticularly prominent along foreground objective edges [S7,], as shown in Figure1. Therefore the objective metric for generated stereoscopic video emphasize edge region evaluation.

**Figure 1 Eample of AI-generated stereoscopic video**



The objective evaluation of generated stereoscopic video sequences generally involves the following procedures:

**Predicted Image and Depth Map Acquisition:**

- It first takes a source monocular image as input into a stereo video generation model (e.g., AI-based algorithm) to generate the predicted image. For example, if the source image is a left-view image, the model generates the right-view image, and vice versa.

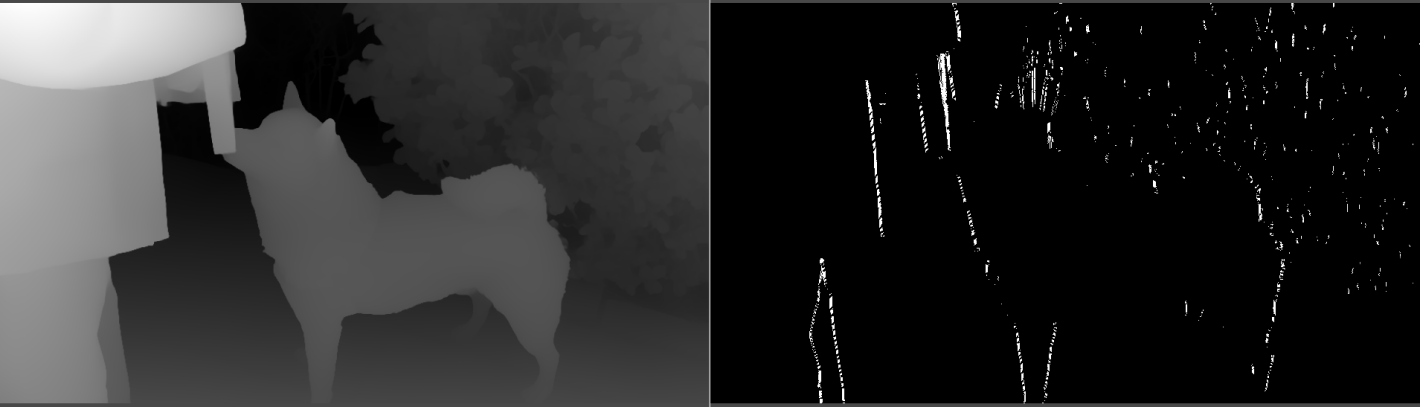
- After generation, a depth estimation algorithm to obtain the depth map of the predicted image.

**Depth Map Processing and Edge Expansion:**

- The depth values in the predicted image's depth map are divided into *N* ranges (where *N* is a positive integer)

- Edge detection and pre-processing such as filtering are the applied to extract vertical edge lines from the depth map (As shown in Figure 2).

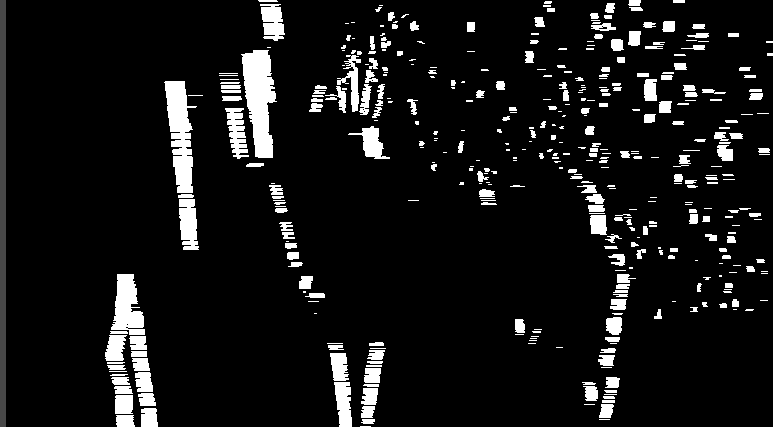
**Figure 2 Eample of Edge detetion in depth maps**



- The expansion ratio for each vertical edge line is determined based on the depth region it belongs to. Regions with greater depth values (i.e., farther from the viewer) are assigned smaller expansion ratios, while regions closer to the viewer has larger deformation, reflecting the greater disparity typically found in close objects.

- Expand the vertical edge lines horizontally according to their respective expansion ratios, and finally obtaining the segmentation mask (Figure 3) of the edge region in the predicted image’s depth map.

**Figure 3 Eample of segmentation mask**



**Evaluation Function Calculation**

- Obtain the edge region in the predict image based on the segmentation mask.

- Calculate the first sub-evaluation function by comparing the edge region of the predicted image with the edge region of the Ground Truth Image.

- Calculate the second sub-evaluation function by comparing the non-edge regions of the predicted image (i.e., regions excluding the second edge region) with the non-edge regions of the Ground Truth Image.

The final evaluation function Q is calculated through the following weighted summation formula:  
 Q = α·subQ1 + β·subQ2

Here, α and β represent the weights assigned to the first and second sub-evaluation functions, respectively. Since artifacts are significantly more pronounced in edge regions compared to non-edge regions, α is significantly smaller than β. For example, α and β may be assigned values of 0.2 and 0.8, respectively.

\* \* \* Second Change \* \* \*

# 2 References

[S7] S. Khan and S. S. Channappayya, "Estimating Depth-Salient Edges and Its Application to Stereoscopic Image Quality Assessment," IEEE Transactions on Image Processing, vol. 27, no. 12, pp. 5892–5903, 2018, doi: 10.1109/TIP.2018.2860279.

\* \* \* End of Changes \* \* \*