INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG4/m68710
July 2024, Sapporo, JP

Source	**Electronics and Telecommunications Research Institute (ETRI)**
Status	**Input Contribution**
Title	**Summary of MIV-DSDE anchor generation for 3D-INVR EEs**
Author	**Jun Young Jeong, Gun Bang (ETRI), Adrian Dziembowski (PUT), Jong-Beom Jeong (SKKU), Yiyi Liao (Zhejiang Univ.)**

1 Introduction

3D-INVR is considering supporting a 6DoF service through view synthesis as one of the main use cases. However, the latest MPEG standard, MIV (MPEG Immersive Video) already supports a similar use case. To justify the development of a brand-new standard, it's necessary to demonstrate that 3D-INVR methods outperform the MIV standard in terms of compression efficiency and rendering quality.

For this reason, during the MPEG 146 meeting, EE3 was launched to generate the MIV-DSDE (Decoder-side Depth Estimation) anchor, which will serve as the baseline for evaluating the various outcomes of 3D-INVR EEs. Following MPEG 146, EE3 was carried out in collaboration with multiple organizations, and this contribution summarizes the results and discussion points.

2 CTC and task allocations

TMIV 18 [N0409] and its' corresponding CTC [N0406] were selected as a baseline, but the below modifications were made specifically for the 3D-INVR activity:

- The total number of frames for encoding increased to 97 from 65 (GOP size is identical to MIV CTC, which is 32)
- Selected test views are excluded from the encoding process. Then, these excluded views are synthesized at the decoder-side based on the transmitted views and their estimated depths. Unlike MIV CTC, which uses both transmitted (training) and non-transmitted (test) views for objective evaluation, in the case of INVR, only the non-transmitted views are considered for objective evaluation
- The reporting template used in MIV is simplified according to the 3D-INVR conditions, but the current version is preliminary and definitely needs improvements, such as adding metrics for measuring encoding and decoding times
  - ✓ The initial version of the reporting template is provided as an attachment. This template accepts bitrates of encoded data across four rate points, along with corresponding metric values (RGB-PSNR, SSIM, LPIPS) for rendered test views. Then, it automatically generates RD-curves and calculates BD-rate between the anchor and proposed techniques

The initial attempt of anchor generation was conducted with the following task allocation. Though the results of some contents were not provided, several discussion points were raised.

Name	Gun Bang	Jong-Beom Jeong	Adrian Dziembowski	Yiyi Liao	Jun Young Jeong
Content	Garage	VRoom1D	Mirror, Choreo HauntedLamp	KITTI-360	Pierrot, Musical

3 Summary of discussion points

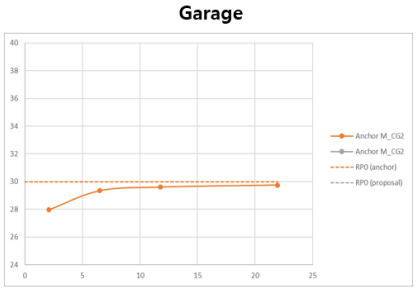
The following is a summary for each discussion point

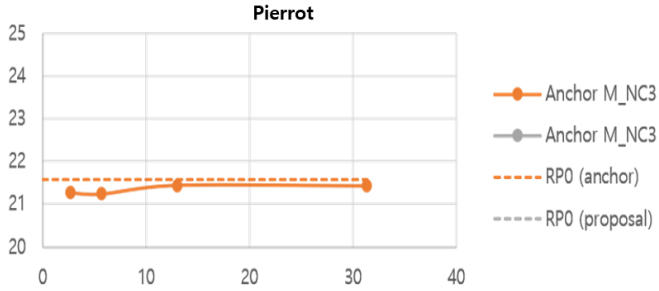
Issue 1. Number of atlases

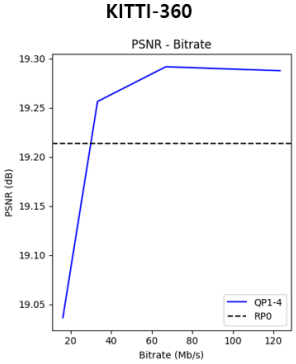
- Considering realistic hardware limitations, MIV CTC restricts the resolution and number of atlases as 4K and 2 pairs (or 4). As a result, the characteristics (such as viewing volume, number of cameras, etc.) of test contents used by MIV etc. naturally conforms to these conditions
- However, this condition may not be suitable for some of 3D-INVR contents, such as Garage and VRroom1D, which cover large viewing volume with a significant number of cameras
- Here are the suggested options regarding this issue:
  - ✓ Use the partial cameras that cover a section of the original viewing volume [m68237]
  - ✓ Increase the number of atlases:
    - Until the encoded data sufficiently covers the viewing volume [m68237]
    - To include all training views for a fair comparison with the radiance field models trained using all training views [m68240]

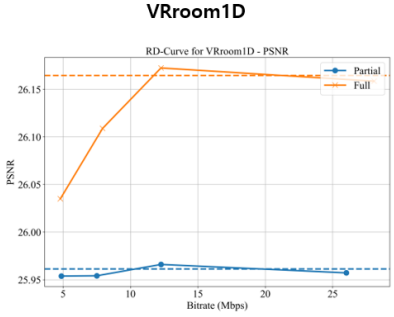
Issue 2. Flat RD-curve

- As shown below, all experimenters observed flat RD-curves across various contents. This is due to fact that the distortion caused by texture compression has little effect on the quality of depth maps and the synthesized views









- According to m68237 and m68245, similar behaviors were also observed in several MIV sequences when only test views are considered. MIV didn't encounter this issue because the degradation of training views prevents the curve being flat. In comparison, 3D-INVR decided not to use training views since objective metrics do not properly reflect subjective quality
- Here are the suggested options regarding this issue:
  - ✓ Modifies IVDE configuration to improve the quality and investigate whether non-flat RD-curve can be obtained [m68240, m68244]

- ✓ Do not calculate BD-rate, but instead, let's show the figure with the RD-curve based on radiance field techniques are located above the one obtained with MIV [m68249]
- ✓ Use NeRF-style methods such as TensoRF + VVC as an anchor [m68241]

Issue 3. Whether to use DSDE anchor at CfP

- During the last AhG meeting, there was a discussion about the anchor of CfP [m68249], but no solutions or alternatives were given

4 ETRI's thoughts on the improvement of anchor generation process

Point 1. Cross-checker

- Cross-checker may be necessary to prevent using distorted results as an anchor

Point 2. Reduce the number of frames

- The most time-consuming part during the generation of the DSDE anchor is the estimation of depth maps using IVDE. Therefore, it is recommended to reduce the current 3GOPs (97 frames) setting to 2GOPs (65 frames)

Point 3. Increase the number of test views

- The current CTC selects only two test views per content. However, some contents such as Garage (36 views) and VRroom1D (30 views) consists of a considerable number of views, and it seems that rendering quality can be maintained even if more views are excluded from encoding. Therefore, it is recommended to increase the number of test views to ensure there are enough test views for drawing stable RD-curves

5 References

N0409 Test model 18 for MPEG immersive video, ISO/IEC JTC 1/SC 29/WG 04 N0409, Hannover, Oct. 2023.

N0406 Common test conditions for MPEG immersive video, ISO/IEC JTC 1/SC 29/WG 04 N0406, Hannover, Oct. 2023.

M68237 Discussion Points on the Generation of MIV DSDE Anchor for 3D-INVR EEs, ISO/IEC JTC 1/SC 29/WG 04 M68237, Sapporo, July. 2024.

M68240 Report on EE3: Thoughts on MIV DSDE Anchor Generation, ISO/IEC JTC 1/SC 29/WG 04 M68240, Sapporo, July. 2024.

M68241 EE3.1 Anchor generation results for Garage; dataset, ISO/IEC JTC 1/SC 29/WG 04 M68240, Sapporo, July. 2024.

M68244 INVR anchor generation on KITTI-360 dataset, ISO/IEC JTC 1/SC 29/WG 04 M68244, Sapporo, July. 2024.

M68245 On the curvature of MIV RD-curves using INVR methodology, ISO/IEC JTC 1/SC 29/WG 04 M68245, Sapporo, July. 2024.