**INTERNATIONAL ORGANISATION FOR STANDARDISATION**
**ORGANISATION INTERNATIONALE DE NORMALISATION**
**ISO/IEC JTC 1/SC 29/WG04**
**MPEG VIDEO CODING**

**ISO/IEC JTC 1/SC 29/WG 04 m58046**
**October 2021, Online**

**Title:** Texture prefiltering in depth estimation
**Source:** PUT: Adrian Dziembowski, Dawid Mieloch, Marek Domański
ETRI: Gwangsoon Lee, Jun Young Jeong

## Abstract

This document presents a description of the experiments on texture filtering performed before depth estimation. The results show a significant improvement of the proposal over the G17 anchor. The recommendation is to include the proposal into IVDE 5.0 and to enable such filtering in the G17 anchor.

# 1 Proposed modification

In the proposal, input views are additionally filtered just before depth estimation. Each view is filtered independently. In the DSDE case, the decoded views are being filtered (at the decoder side).

The proposed solution is supposed to remove an influence of two factors decreasing the quality of depth estimation:

- noise in input views,
- coding artifacts (incl. block artifacts) in DSDE scenario.

The proposal was implemented in IVDE 4.0, tested in the DSDE scenario, and compared against the G17 anchor.

# 2 Results

The proposal was tested on perspective content only (mandatory + optional). Sequence R was excluded, as the IVDE with automatic depth range calculation is not able to produce depth maps of reasonable quality because of invalid depth range.

In the experiments, two types of filtering were tested: median filtering and simple low-pass (average) filtering with blocks of size 3x3, 5x5, and 7x7.

## 2.1 Average vs. median filtering

Table: G17 with average texture filtering (3x3 block) vs. G17 anchor.

| Mandatory content - Proposal vs. Low/High-bitrate Anchors | | | | | | | Runtime ratio (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| Sequence | | High-BR BD rate Y-PSNR | Low-BR BD rate Y-PSNR | Max delta Y-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR | Atlas encoding | Video encoding | Decoding & Rendering |
| Fan | O | -7.4% | -5.6% | 10.67 | -8.3% | -4.7% | 100.0% | 100.0% | 89.0% |
| Kitchen | J | -5.2% | -5.5% | 11.80 | 4.6% | 1.7% | 100.0% | 100.0% | 108.7% |
| Painter | D | -0.8% | 0.2% | 8.94 | -1.5% | 0.1% | 100.0% | 100.0% | 99.6% |
| Frog | E | -2.5% | -1.9% | 7.59 | 1.6% | 0.6% | 100.0% | 100.0% | 104.7% |
| Carpark | P | -12.2% | -7.8% | 10.90 | -9.1% | -7.1% | 100.0% | 100.0% | 81.5% |
| **MIV** | | **-5.6%** | **-4.1%** | **9.98** | **-2.5%** | **-1.9%** | 100.0% | 100.0% | 96.7% |
| **Optional content - Proposal vs. Low/High-bitrate Anchors** | | | | | | | | | |
| Fencing | L | -37.2% | -22.5% | 13.14 | -1.9% | -2.5% | 100.0% | 100.0% | 89.8% |
| Hall | T | -70.5% | --- | 15.70 | -71.8% | --- | 100.0% | 100.0% | 101.4% |
| Street | U | -9.8% | -5.2% | 6.96 | -3.0% | -0.9% | 100.0% | 100.0% | 63.8% |
| Mirror | I | -6.9% | -4.9% | 13.34 | -5.0% | -3.5% | 100.0% | 100.0% | 97.6% |
| **MIV** | | **-31.1%** | --- | **12.29** | **-20.4%** | --- | 100.0% | 100.0% | 88.1% |

Table: G17 with median texture filtering (3x3 block) vs. G17 anchor.

| Mandatory content - Proposal vs. Low/High-bitrate Anchors | | | | | | | Runtime ratio (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| Sequence | | High-BR BD rate Y-PSNR | Low-BR BD rate Y-PSNR | Max delta Y-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR | Atlas encoding | Video encoding | Decoding & Rendering |
| Fan | O | -3.1% | -2.5% | 10.75 | -2.8% | -1.2% | 100.0% | 100.0% | 92.0% |
| Kitchen | J | 6.7% | 4.0% | 11.95 | 3.6% | 2.7% | 100.0% | 100.0% | 87.8% |
| Painter | D | 6.1% | 2.0% | 9.17 | 4.9% | 1.6% | 100.0% | 100.0% | 83.7% |
| Frog | E | -0.5% | -0.5% | 7.67 | 2.5% | 1.3% | 100.0% | 100.0% | 84.8% |
| Carpark | P | 14.3% | 11.0% | 11.14 | 9.7% | 7.2% | 100.0% | 100.0% | 87.7% |
| **MIV** | | **4.7%** | **2.8%** | **10.13** | **3.6%** | **2.3%** | 100.0% | 100.0% | 87.2% |
| **Optional content - Proposal vs. Low/High-bitrate Anchors** | | | | | | | | | |
| Fencing | L | -7.4% | -7.1% | 13.23 | 8.2% | 0.3% | 100.0% | 100.0% | 101.9% |
| Hall | T | -50.9% | 22.5% | 19.40 | -4.0% | 8.6% | 100.0% | 100.0% | 92.5% |
| Street | U | 15.6% | 9.4% | 7.06 | 9.1% | 6.2% | 100.0% | 100.0% | 86.6% |
| Mirror | I | 6.8% | 1.9% | 13.77 | 10.5% | 3.2% | 100.0% | 100.0% | 113.1% |
| **MIV** | | **-9.0%** | **6.7%** | **13.37** | **6.0%** | **4.6%** | 100.0% | 100.0% | 98.5% |

Usage of average filtering significantly increases the quality of synthesized views while decreasing the time needed for depth estimation.

Median filtering also decreases the depth estimation time, but the quality is worse than in the G17 anchor.

In the remaining experiments, only the average filtering was tested.

## 2.2 Luma vs. luma + chroma filtering

Table: G17 with average luma filtering (3x3 block) vs. G17 anchor.

**Mandatory content - Proposal vs. Low/High-bitrate Anchors**

| Sequence | | High-BR BD rate Y-PSNR | Low-BR BD rate Y-PSNR | Max delta Y-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR |
|---|---|---|---|---|---|---|
| Fan | O | -3.1% | -3.1% | 10.73 | -6.2% | -3.9% |
| Kitchen | J | -6.6% | -7.4% | 11.61 | -1.1% | -2.6% |
| Painter | D | -1.0% | -0.1% | 9.15 | -1.2% | -0.1% |
| Frog | E | -2.4% | -1.7% | 7.58 | 0.4% | 0.2% |
| Carpark | P | -7.2% | -3.3% | 10.96 | -5.6% | -2.2% |
| **MIV** | | **-4.1%** | **-3.1%** | **10.01** | **-2.7%** | **-1.7%** |

**Optional content - Proposal vs. Low/High-bitrate Anchors**

| | | | | | | |
|---|---|---|---|---|---|---|
| Fencing | L | -38.3% | -18.3% | 13.14 | -1.8% | -1.4% |
| Hall | T | -65.6% | --- | 15.88 | -68.7% | --- |
| Street | U | -8.6% | -2.6% | 6.98 | -2.9% | 1.4% |
| Mirror | I | -3.4% | -3.2% | 13.51 | -0.2% | -1.3% |
| **MIV** | | **-29.0%** | --- | **12.38** | **-18.4%** | --- |

Table: G17 with average luma and chroma filtering (3x3 block) vs. G17 anchor.

**Mandatory content - Proposal vs. Low/High-bitrate Anchors**

| Sequence | | High-BR BD rate Y-PSNR | Low-BR BD rate Y-PSNR | Max delta Y-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR |
|---|---|---|---|---|---|---|
| Fan | O | -7.4% | -5.6% | 10.67 | -8.3% | -4.7% |
| Kitchen | J | -5.2% | -5.5% | 11.80 | 4.6% | 1.7% |
| Painter | D | -0.8% | 0.2% | 8.94 | -1.5% | 0.1% |
| Frog | E | -2.5% | -1.9% | 7.59 | 1.6% | 0.6% |
| Carpark | P | -12.2% | -7.8% | 10.90 | -9.1% | -7.1% |
| **MIV** | | **-5.6%** | **-4.1%** | **9.98** | **-2.5%** | **-1.9%** |

**Optional content - Proposal vs. Low/High-bitrate Anchors**

| | | | | | | |
|---|---|---|---|---|---|---|
| Fencing | L | -37.2% | -22.5% | 13.14 | -1.9% | -2.5% |
| Hall | T | -70.5% | --- | 15.70 | -71.8% | --- |
| Street | U | -9.8% | -5.2% | 6.96 | -3.0% | -0.9% |
| Mirror | I | -6.9% | -4.9% | 13.34 | -5.0% | -3.5% |
| **MIV** | | **-31.1%** | --- | **12.29** | **-20.4%** | --- |

Filtering of the luma component only performs slightly worse than filtering of all color components, but the difference is not huge.

In the next experiment, the luma + chroma filtering was tested.

## 2.3 Different window sizes

Table: G17 with average texture filtering (3x3 block) vs. G17 anchor.

| Mandatory content - Proposal vs. Low/High-bitrate Anchors | | | | | | | Runtime ratio (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| Sequence | | High-BR BD rate Y-PSNR | Low-BR BD rate Y-PSNR | Max delta Y-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR | Atlas encoding | Video encoding | Decoding & Rendering |
| Fan | O | -7.4% | -5.6% | 10.67 | -8.3% | -4.7% | 100.0% | 100.0% | 89.0% |
| Kitchen | J | -5.2% | -5.5% | 11.80 | 4.6% | 1.7% | 100.0% | 100.0% | 108.7% |
| Painter | D | -0.8% | 0.2% | 8.94 | -1.5% | 0.1% | 100.0% | 100.0% | 99.6% |
| Frog | E | -2.5% | -1.9% | 7.59 | 1.6% | 0.6% | 100.0% | 100.0% | 104.7% |
| Carpark | P | -12.2% | -7.8% | 10.90 | -9.1% | -7.1% | 100.0% | 100.0% | 81.5% |
| **MIV** | | **-5.6%** | **-4.1%** | **9.98** | **-2.5%** | **-1.9%** | 100.0% | 100.0% | 96.7% |
| **Optional content - Proposal vs. Low/High-bitrate Anchors** | | | | | | | | | |
| Fencing | L | -37.2% | -22.5% | 13.14 | -1.9% | -2.5% | 100.0% | 100.0% | 89.8% |
| Hall | T | -70.5% | --- | 15.70 | -71.8% | --- | 100.0% | 100.0% | 101.4% |
| Street | U | -9.8% | -5.2% | 6.96 | -3.0% | -0.9% | 100.0% | 100.0% | 63.8% |
| Mirror | I | -6.9% | -4.9% | 13.34 | -5.0% | -3.5% | 100.0% | 100.0% | 97.6% |
| **MIV** | | **-31.1%** | --- | **12.29** | **-20.4%** | --- | 100.0% | 100.0% | 88.1% |

Table: G17 with average texture filtering (5x5 block) vs. G17 anchor.

| Mandatory content - Proposal vs. Low/High-bitrate Anchors | | | | | | | Runtime ratio (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| Sequence | | High-BR BD rate Y-PSNR | Low-BR BD rate Y-PSNR | Max delta Y-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR | Atlas encoding | Video encoding | Decoding & Rendering |
| Fan | O | -5.1% | -3.9% | 10.95 | -9.6% | -4.7% | 100.0% | 100.0% | 114.7% |
| Kitchen | J | -2.9% | -5.7% | 11.86 | 5.6% | -0.1% | 100.0% | 100.0% | 98.0% |
| Painter | D | 1.6% | 0.2% | 9.09 | 0.0% | -0.4% | 100.0% | 100.0% | 84.2% |
| Frog | E | -3.5% | -2.6% | 7.64 | 1.4% | 0.6% | 100.0% | 100.0% | 83.3% |
| Carpark | P | -19.2% | -13.7% | 10.92 | -18.4% | -13.0% | 100.0% | 100.0% | 99.1% |
| **MIV** | | **-5.8%** | **-5.1%** | **10.09** | **-4.2%** | **-3.5%** | 100.0% | 100.0% | 95.8% |
| **Optional content - Proposal vs. Low/High-bitrate Anchors** | | | | | | | | | |
| Fencing | L | -64.7% | -39.3% | 12.99 | -7.3% | -9.4% | 100.0% | 100.0% | 88.2% |
| Hall | T | -63.6% | -88.6% | 16.39 | -63.1% | -85.0% | 100.0% | 100.0% | 93.4% |
| Street | U | -4.9% | -2.1% | 6.98 | 4.9% | 4.6% | 100.0% | 100.0% | 86.4% |
| Mirror | I | -5.9% | -6.0% | 13.86 | -2.3% | -3.7% | 100.0% | 100.0% | 83.5% |
| **MIV** | | **-34.8%** | **-34.0%** | **12.55** | **-16.9%** | **-23.4%** | 100.0% | 100.0% | 87.8% |

Table: G17 with average texture filtering (7x7 block) vs. G17 anchor.

| Mandatory content - Proposal vs. Low/High-bitrate Anchors | | | | | | | Runtime ratio (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| Sequence | | High-BR BD rate Y-PSNR | Low-BR BD rate Y-PSNR | Max delta Y-PSNR | High-BR BD rate IV-PSNR | Low-BR BD rate IV-PSNR | Atlas encoding | Video encoding | Decoding & Rendering |
| Fan | O | -1.0% | -1.7% | 10.96 | -3.2% | -0.6% | 100.0% | 100.0% | 101.0% |
| Kitchen | J | 7.5% | 1.8% | 12.11 | 15.2% | 5.4% | 100.0% | 100.0% | 75.7% |
| Painter | D | 4.9% | 2.3% | 9.02 | 2.7% | 1.5% | 100.0% | 100.0% | 121.7% |
| Frog | E | -2.1% | -1.8% | 7.70 | 2.5% | 1.3% | 100.0% | 100.0% | 88.2% |
| Carpark | P | -10.5% | -8.9% | 10.94 | -13.5% | -11.1% | 100.0% | 100.0% | 69.8% |
| **MIV** | | **-0.2%** | **-1.7%** | **10.15** | **0.7%** | **-0.7%** | 100.0% | 100.0% | 91.3% |
| **Optional content - Proposal vs. Low/High-bitrate Anchors** | | | | | | | | | |
| Fencing | L | -65.2% | -37.7% | 12.98 | -0.7% | -4.3% | 100.0% | 100.0% | 92.4% |
| Hall | T | -25.5% | -71.6% | 19.16 | -22.5% | -48.7% | 100.0% | 100.0% | 125.4% |
| Street | U | 25.6% | 8.6% | 7.48 | 28.7% | 12.2% | 100.0% | 100.0% | 86.7% |
| Mirror | I | -5.9% | -6.8% | 14.04 | 1.7% | -2.8% | 100.0% | 100.0% | 104.4% |
| **MIV** | | **-17.7%** | **-26.9%** | **13.41** | **1.8%** | **-10.9%** | 100.0% | 100.0% | 102.2% |

The influence of the window size can be easily spotted in the tables above. Windows 3x3 and 5x5 perform similarly. When the size of the window is increased even more, the quality of synthesized views is worse in general (but not for all the sequences).

The difference between 3x3 and 5x5 windows is small, but the 5x5 seems to be a better choice.

# 3 Recommendations

We recommend to:

- include the proposal into IVDE 5.0,
- enable 5x5 average filtering in G17 anchor.

# 4 Acknowledgement

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory).