MPEG-Immersive 6DoF Standard Work: Related to MPEG Immersive Video

Presenter: Eun-Seok Ryu (esryu@skku.edu)
Jong-Beom Jeong, Soonbin Lee, Eun-Seok Ryu

Multimedia Computing Systems Lab. (MCSL)
http://mcsl.skku.edu
Department of Computer Education
Sungkyunkwan University
Compliance with IEEE Standards Policies and Procedures

Subclause 5.2.1 of the *IEEE-SA Standards Board Bylaws* states, "While participating in IEEE standards development activities, all participants...shall act in accordance with all applicable laws (nation-based and international), the IEEE Code of Ethics, and with IEEE Standards policies and procedures."

The contributor acknowledges and accepts that this contribution is subject to


**MPEG-Immersive 6DoF Standard Work: Related to MPEG Immersive Video**

**Date:** 2021-07-19

**Author(s):** Jong-Beom Jeong, Soonbin Lee, Eun-Seok Ryu

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Phone [optional]</th>
<th>Email [optional]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jong-Beom Jeong</td>
<td>Sungkyunkwan Univ.</td>
<td></td>
<td><a href="mailto:uof4949@skku.edu">uof4949@skku.edu</a></td>
</tr>
<tr>
<td>Soonbin Lee</td>
<td>Sungkyunkwan Univ.</td>
<td></td>
<td><a href="mailto:soonbinlee@skku.edu">soonbinlee@skku.edu</a></td>
</tr>
<tr>
<td>Eun-Seok Ryu</td>
<td>Sungkyunkwan Univ.</td>
<td></td>
<td><a href="mailto:esryu@skku.edu">esryu@skku.edu</a></td>
</tr>
</tbody>
</table>
MPEG Immersive Video (MIV) Overview

- Step-by-step objective of ISO/IEC MPEG immersive video
  - MPEG-I is responsible for standardizing immersive media in MPEG and specifies the goals of step 3
  - Goal of Revitalizing VR Commercial Service by 2020
  - Goal of 6DoF media support by 2022 after completing 3DoF standard by 2017

Definitions of 3DoF, 3DoF+, and 6DoF
MPEG Immersive Video (MIV) Overview (-Cont’d)

- MPEG-I requested call for proposals (CfP) on 3DoF+
- Philips, Intel & Technicolor, Nokia, PUT & ETRI, ZJU submitted responses
  > test model for immersive video (TMIV)
- Removes correlation between multi-view
- VR streaming on 6DoF natural-content
MPEG Immersive Video (MIV)-related Activities (-Cont’d)

- Mandates and Input Contributions in MPEG134 MPEG-I Visual
  - 1. Prepare draft international standard of MPEG Immersive Video  2
  - 2. Study profiles, tiers, and levels for MPEG Immersive Video  1
  - 3. Carry on core experiments of MPEG Immersive Video  3
  - 4. Update test model and reference software for MPEG Immersive Video  3
  - 5. Define CTC for MPEG Immersive Video  8
  - 6. Carry out exploration experiments on MPEG Immersive Video  10
  - 7. Investigate dense light field content in immersive video applications  10
  - 8. Study performance improvements under CTC conditions  2
  - 9. Promote MPEG Immersive Video outside MPEG  0
  - Total  39 doc.
**MIV Common Test Conditions**

- **Reference softwares**

<table>
<thead>
<tr>
<th>Software name</th>
<th>Location</th>
<th>Tag/branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMIV</td>
<td><a href="https://gitlab.com/mpeg-i-visual/tmiv">https://gitlab.com/mpeg-i-visual/tmiv</a></td>
<td>v8.0</td>
</tr>
<tr>
<td>VVenC</td>
<td><a href="https://github.com/fraunhoferhhi/vvenc">https://github.com/fraunhoferhhi/vvenc</a></td>
<td>v0.2.0.0</td>
</tr>
<tr>
<td>VVdeC</td>
<td><a href="https://github.com/fraunhoferhhi/vvdec">https://github.com/fraunhoferhhi/vvdec</a></td>
<td>v0.2.0.0</td>
</tr>
<tr>
<td>VMAF</td>
<td><a href="https://github.com/Netflix/vmaf">https://github.com/Netflix/vmaf</a></td>
<td>v1.3.14</td>
</tr>
<tr>
<td>IV-PSNR</td>
<td><a href="https://gitlab.com/mpeg-i-visual/ivpsnr">https://gitlab.com/mpeg-i-visual/ivpsnr</a></td>
<td>v3.0</td>
</tr>
<tr>
<td>IVDE</td>
<td><a href="https://gitlab.com/mpeg-i-visual/ivde">https://gitlab.com/mpeg-i-visual/ivde</a></td>
<td>v3.0</td>
</tr>
</tbody>
</table>
[MIV] ZJU response to CE2 viewport culling (m56415)

- Previous culler removed some patches from the viewport
- This is because of the patches which has FoV >= 180°
- Divided such large patches into sub-blocks
- AhG recommendations:
  - Integrate into TMIV
  - Enable in the CTC anchors

Frame in SQ P01 anchor (left) and modified subblock culler (right)
Current TMIV normalizes depth values into 10-bits.

Proposed method conducts:
- Divide depth into predefined number of intervals (proposed: 16)
- Adaptively scale depth values
- Important close depth - disparity is large

Example of piecewise linear scaling:

\[ d' = \frac{(b_i^2 - b_i^1) \times (d - a_i^1)}{a_i^2 - a_i^1} + b_{1i} \]

The figure illustrates the piecewise linear scaling with intervals defined by \( a_i^1 \) and \( a_i^2 \), and \( b_i^1 \) and \( b_i^2 \), where \( d \) is the original depth value, and \( d' \) is the scaled depth value. The scaling function is depicted across the range of original depth values, with different accuracy levels indicated for different depth ranges.
3.4% BD-rate gain compared to the anchor
Subjective quality advantages at the edges
Proposed new syntax for depth quantization
Pose trace issues with TMIV8 (m56774)

- Block artifacts are generated when using TMIV8 anchor (VVenC). Maybe because of optimization in VVenC..? Blocky artifacts in high QP.

AhG recommendations:
- 1. Better crosschecking: Evaluate the pose traces, study the code, apply a strict expert viewing process.
- 2. Focus more on the pose trace generation: Viewing pose traces when releasing TMIV9, objective result is less important.
- 3. Tool evaluation: Decision with only subjective results, encode 17 frames and render 97.
Frame Packing Implementation in TMIV (m56827)

- Frame packing implementation in TMIV to:
  - No empty spaces in the packed video
  - Not restricted to VVC (sub-pictures) or HEVC-specific (MCTS) solutions
  - Frame packing across different atlas indices is not available
    - current syntax does not allow it

<table>
<thead>
<tr>
<th>packing_information()[]</th>
<th>Descr</th>
<th>Value for jth atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>pin codec id[1]</td>
<td>u(8)</td>
<td>1 for HEVC main10, 3 for VVC main10</td>
</tr>
<tr>
<td>pin regions count minus[1]</td>
<td>u(1)</td>
<td>2</td>
</tr>
<tr>
<td>for(i = 0; i &lt; pin regions count minus[1]; i++)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pin region tile id[i][1]</td>
<td>u(8)</td>
<td>Looping over regions</td>
</tr>
<tr>
<td>pin region type id minus2[i][1]</td>
<td>u(2)</td>
<td>2 (for attribute)</td>
</tr>
<tr>
<td>pin region top left x[i][1]</td>
<td>u(8)</td>
<td>0</td>
</tr>
<tr>
<td>pin region top left y[i][1]</td>
<td>u(8)</td>
<td>0</td>
</tr>
<tr>
<td>pin region width minus[i][1]</td>
<td>u(16)</td>
<td>1919</td>
</tr>
<tr>
<td>pin region height minus[i][1]</td>
<td>u(16)</td>
<td>4640</td>
</tr>
<tr>
<td>pin region unpack top left x[i][1]</td>
<td>u(16)</td>
<td>0</td>
</tr>
<tr>
<td>pin region unpack top left y[i][1]</td>
<td>u(16)</td>
<td>0</td>
</tr>
<tr>
<td>pin region map index[i][1]</td>
<td>u(4)</td>
<td>0</td>
</tr>
<tr>
<td>pin region rotation flag[i][1]</td>
<td>u(1)</td>
<td>0</td>
</tr>
</tbody>
</table>

if(pin region type id minus2[i][1] + 2 == V3C_AVD)
  pin region auxiliary data flag[i][1]
end

if(pin region type id minus2[i][1] + 2 == V3C_GVD)
  pin region auxiliary data flag[i][1]
end

if(pin region attr type id[i][1] != 0)
  pin region attr partitions flag[i][1]
end

if(pin region attr partitions flag[i][1] != 0)
  pin region attr partition index[i][1]
end

if(pin region attr partition_index[i][1] != 0)
  pin region attr partitions minus[i][1]
end

Packing information for V17 with 1 tile
Recommendations for MIV expert viewing sessions (m56321)

- Guidelines for expert viewing session
- Watching anchor/proposed video (10s) two times is enough
  Typical option: A/B/A/B, P.800 for instance
  Lowering frame rate is not acceptable
- Because matching the bitrate is hard, some flexibility is allowed
  Proponents can adjust texture and depth QP

Example of viewing session

Sequence rating
Test Materials (m56450, 56632, 56730)

- Non-Lambertian CG content “MagritteSphere” by m56450
  - Rendered with Blender, two versions (transparent and mirror)
  - 1 frame, 21 × 21 views, 2000 × 2000 size

- NC content “Barn” by m56632
  - 97 frames, 5 × 3 perspective views, 1920 × 1080 size
  - Some holes exist (invalid depth)
  - Estimated depth by proprietary tool

- NC content “Breakfast” by m56730
  - Spec. is same as “Barn”
  - Captured in the dining room of Chateau de la Ballue
Test Materials (m56429, m56652, m56787)

- Natural content “RabbitStamp” by m56429
  - Captured by RayTrix R8 Plenoptic 2.0
  - 1 frame, 7 × 3 views, 2K × 2K size

- Natural content “Clay magic” by m56429
  - Captured by RayTrix R8 Plenoptic 2.0
  - 193-318 frames, 14 × 11 views, 1920 × 1080 size
  - Not parallel cameras -> good for testing MIV..?

- Natural content “Sports Car” by m56787
  - Captured by two plenoptic cameras
  - Recitification (rotate, crop, color matching)
Evaluation of objective quality metrics in MIV context (m56411)

- Current objective metrics are not suitable for MIV
  - Compression, synthesis artifacts
  - Evaluation on source views because GT is not enough (pose trace is needed)
  - Large differences between anchor and proposal
- Crowdsourcing-like subjective evaluation
- No-reference (NR) based on DL, but not re-trained for MIV
- Correlation with MOS: 0.76 RMSE, 0.75 PLCC, 0.73 SRCC

<table>
<thead>
<tr>
<th>Category</th>
<th>Error Visibility</th>
<th>Predicted Quality/ (MOS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>Imperceptible, Not annoying</td>
<td>5</td>
</tr>
<tr>
<td>Good</td>
<td>Perceptible Error, Slightly annoying</td>
<td>4</td>
</tr>
<tr>
<td>Poor</td>
<td>Visible Error, Annoying</td>
<td>3</td>
</tr>
<tr>
<td>Bad</td>
<td>Visible error, Very annoying</td>
<td>1</td>
</tr>
</tbody>
</table>

ACR scale score
Extraction and Merging on Frame Packed Video (m56591)

- SKKU MCSL proposed bitstream-level packing
- Fine-grained system in terms of the number of decoder instances
  - Current TMIV supports pixel-level frame packing -> provides high level two bitstreams
  - Proposed method can provide low level four or high level two bitstreams
  - 0.7%, 3.93% BD-rate saving because of deblocking filter deactivation
- Developed tile bitstream extracting/merging software on HEVC test model
- AhG recommendations
  - Tiling is needed for TMIV
  - Implementation of HLS to enable tiling on TMIV is more than welcome

System overview of packing method by MCSL

Example of bitstream-level packing
Thank You!

http://mcsl.skku.edu/
Questions > esryu@skku.edu