MPEG-Immersive 3DoF+ Standard Work:
Related to 3DoF+ Call for Proposal

Presenter: Eun-Seok Ryu (esryu@gachon.ac.kr)
Jong-Beom Jeong, Dongmin Jang, Ju Hyeong Kim, SoonBin Lee, Eun-Seok Ryu

Multimedia Communications and Systems Lab. (MCSL)
http://mcsl.gachon.ac.kr
Department of Computer Engineering
Gachon University
Test Materials for 3DoF+ - w17726

❖ Philips
  ▪ ClassroomVideo

❖ Technicolor
  ▪ TechnicolorMuseum
  ▪ TechnicolorHijack

Class A
ClassroomVideo (4096x2048), 360° x 180° FOV ERP, 30fps, 120frames, 15 source views

Class B
TechnicolorMuseum (2048x2048), 180° x 180° FOV ERP, 30fps, 300frames, 24 source views

Class C
TechnicolorHijack (4096x4096), 180° x 180° FOV ERP, 30fps, 300frames, 10 source views
Common Test Conditions (CTC) for 3DoF+ - w17726

❖ Software

<table>
<thead>
<tr>
<th>Software name</th>
<th>Location</th>
<th>Tag/branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVS</td>
<td><a href="http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/RVS">http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/RVS</a></td>
<td>v2.0.1</td>
</tr>
<tr>
<td>ERP-WS-PSNR</td>
<td><a href="http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/ERP_WS-PSNR">http://mpegx.int-evry.fr/software/MPEG/Explorations/3DoFplus/ERP_WS-PSNR</a></td>
<td>v1.0</td>
</tr>
<tr>
<td>HDRTools</td>
<td><a href="https://gitlab.com/standards/HDRTools/tree/0.18-dev">https://gitlab.com/standards/HDRTools/tree/0.18-dev</a></td>
<td>v0.18</td>
</tr>
<tr>
<td>360Lib</td>
<td><a href="https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/branches/360Lib-5.1-dev">https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/branches/360Lib-5.1-dev</a></td>
<td>5.1-dev</td>
</tr>
<tr>
<td>HM</td>
<td><a href="https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.16">https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.16</a></td>
<td>16.16</td>
</tr>
</tbody>
</table>
Common Test Conditions (CTC) for 3DoF+ - w17726

❖ Anchor definition

Definition of the anchor

Technical proposal with pre- and post-processing
Common Test Conditions (CTC) for 3DoF+ - w17726

❖ Quantization Parameter and class

<table>
<thead>
<tr>
<th></th>
<th>QP1</th>
<th>QP2</th>
<th>QP3</th>
<th>QP4</th>
<th>QP5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>12</td>
<td>17</td>
<td>22</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td>Texture</td>
<td>22</td>
<td>27</td>
<td>32</td>
<td>37</td>
<td>42</td>
</tr>
</tbody>
</table>

QPs used for depth and texture

<table>
<thead>
<tr>
<th>Test class</th>
<th>Sequence Name</th>
<th>No. of source views</th>
<th>No. of anchor-coded views</th>
<th>Anchor-coded views</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>ClassroomVideo</td>
<td>15</td>
<td>15</td>
<td>All</td>
</tr>
<tr>
<td>A2</td>
<td>ClassroomVideo</td>
<td>15</td>
<td>9</td>
<td>v0, v7…v14</td>
</tr>
<tr>
<td>B1</td>
<td>TechnicolorMuseum</td>
<td>24</td>
<td>24</td>
<td>All</td>
</tr>
<tr>
<td>B2</td>
<td>TechnicolorMuseum</td>
<td>24</td>
<td>8</td>
<td>0, 1, 4, 8, 11, 12, 13, 17</td>
</tr>
<tr>
<td>C1</td>
<td>TechnicolorHijack</td>
<td>10</td>
<td>10</td>
<td>All</td>
</tr>
<tr>
<td>C2</td>
<td>TechnicolorHijack</td>
<td>10</td>
<td>5</td>
<td>1, 4, 5, 8, 9</td>
</tr>
</tbody>
</table>

Anchor-coded views per class
## Common Test Conditions (CTC) for 3DoF+ - w17726

### 3DoF+ objective evaluation

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coded view position</td>
<td>The position of a source view which is coded by the anchor</td>
</tr>
<tr>
<td>Non coded source view position</td>
<td>The position of a source view which isn’t coded by the anchor</td>
</tr>
<tr>
<td>Intermediate view position</td>
<td>The position of a view which is out of any source view</td>
</tr>
</tbody>
</table>

**Definitions for test class on 3DoF+**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coded view position</td>
<td>The position of a source view which is coded by the anchor</td>
</tr>
<tr>
<td>Non coded source view position</td>
<td>The position of a source view which isn’t coded by the anchor</td>
</tr>
<tr>
<td>Intermediate view position</td>
<td>The position of a view which is out of any source view</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coded BD rate</td>
<td>All frames</td>
</tr>
<tr>
<td>Non-coded BD rate</td>
<td>All frames</td>
</tr>
<tr>
<td>Intermediate BD rate</td>
<td>All frames</td>
</tr>
</tbody>
</table>

**WS-PSNR**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coded BD rate</td>
<td>All frames</td>
</tr>
<tr>
<td>Non-coded BD rate</td>
<td>All frames</td>
</tr>
<tr>
<td>Intermediate BD rate</td>
<td>All frames</td>
</tr>
</tbody>
</table>

**For the anchor RD curve**

- All frames
- Coded view & source view

**For the proponent’s RD curve**

- All frames
- Proponent’s view & source view
- Non-coded source view & source view
- Proponent’s intermediate view & Same view made by all source views with 16-bit depth maps
Reference View Synthesizer - w17759

- Reference software for view synthesis of 3DoF+

Conceptual Diagram

Virtual View

View Blending
by dividing High Frequency and Low Frequency

View & Depth Synthesis using Warping

> Texture & Depth

Input View 1

View & Depth Synthesis using Warping

> Texture & Depth

Input View 2

View & Depth Synthesis using Warping

> Texture & Depth

Input View 3

View & Depth Synthesis using Warping

> Texture & Depth

Input View 4

Simple Example

Virtual View

View Blending

View & Depth Synthesis

View & Depth Synthesis

Input View L

Input View R
Reference View Synthesizer - w17759

- 3D Warping
  - Pixel movement between viewpoints of the same world coordinate
  - Input view is divided into triangles with the pixels centers as vertices
  - Use affine transformation
  - Being filled with interpolated colors after affine transformation

(a) Input view  (b) Obtained view after synthesizing view and depth
View synthesis

\[
M_{Ci} = (X_{Ci}, Y_{Ci}, Z_{Ci}) \\
M_{Cs} = (X_{Cs}, Y_{Cs}, Z_{Cs}) \\
M_W = (X_W, Y_W, Z_W)
\]

Camera

\[
M_C = RM_w + t \quad M_C = \tilde{m}A^{-1}
\]

Camera coordinate move by \(t\) and rotate by \(R\) from world coordinate

\[
M_w = R_i^{-1}A_i^{-1}\tilde{m}_i - R_i^{-1}t_i \quad \text{(Input view)}
\]

\[
R_i^{-1}A_i^{-1}\tilde{m}_i - R_i^{-1}t_i = R_s^{-1}A_s^{-1}\tilde{m}_s - R_s^{-1}t_s
\]

\[
\tilde{m}_s = A_sR_sR_i^{-1}A_i^{-1}\tilde{m}_i - A_sR_sR_i^{-1}t_i + A_s
\]
Reference View Synthesizer - w17759

❖ Good quality: low depth and triangle with a regular shape
❖ Taking the pixel with the maximal quality would give a sharper result
❖ Taking the weighted mean is more resistant to errors
❖ High and low frequencies are separated with a mean blur
❖ Low frequencies are blended with the weighted mean
❖ High frequencies are blended by choosing the pixel of highest weight

(a): Blending by argmax
(b): Weighted mean
(c): Multi-spectral blending: argmax - high frequencies
(weighted mean - low frequencies)

$w_i = \left( \frac{q_i}{d_i} \right)^\alpha$  
depth at pixel for synthesized view $i$

$C_{low} = \sum_{i=0}^{n} w_i c_i^l \bigg/ \sum_{i=0}^{n} w_i$

$C_{high} = argmax_{c_i^h}(w_i)$

$c_i^l, c_i^h$: color of the pixel in view $i$ for the low and high frequencies
Reference View Synthesizer - w17759

❖ Architecture
Pipeline Object (Pipeline.cpp)

Pipeline::compute_views(int Frame)

Routine: Create Virtual View(Renderd View) to Write, For i = 1 to num_virtual_view

- Set Blending Method
  - Type: 1)Simple 2)MultiSpec
  - CTC use (1)

- Set Virtual Camera Parameters
- Inpainting
  - Type: 1)Interpolation 2)Nearest
  - RVS only uses (2)
  - Using Manhattan Not Main Interest

- Downscale
  - Downscale rasterized image

- Write OutputFile & OutMaskedFile

Routine: Load Image per Frame

- Initialize Image Buffer

Routine: Load Image per Frame

- Set Image Params

Routine: Blend Virtual Views from Input Views, For i = 1 to num_input_view

- Set Projection Type
  - Type: 1)Perspective 2)Equirectangular

- Set Input Camera Parameters
- Set Synthesis Method
  - RVS only has Triangle Method

- Load Input View Image

- Synthesis

- Blending

- Compute Views
Fast Color Correction for View Synthesis

- Fast color correction technique for view synthesis
- Reduction of color artifacts in synthesized images
  - Offset calculation
  - Offset addition

<table>
<thead>
<tr>
<th>Test sequence</th>
<th>PSNR [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VSRS</td>
</tr>
<tr>
<td></td>
<td>No color correction</td>
</tr>
<tr>
<td>Ballet</td>
<td>31.45</td>
</tr>
<tr>
<td>Breakdancers</td>
<td>31.89</td>
</tr>
<tr>
<td>Soccer Linear</td>
<td>34.80</td>
</tr>
<tr>
<td>BBB Butterfly</td>
<td>32.06</td>
</tr>
<tr>
<td>BBB Flowers</td>
<td>22.71</td>
</tr>
</tbody>
</table>

SVS, no color correction (31.7 dB)  SVS, with color correction (32.6 dB)

View Synthesis with Proposed Color Correction
(oC* – offset calculation)
3DoF+ Software Platform Description - w17761

- 3DoF+ Software platform
Outperforming 3DoF+ Anchors – m43504

- Removing overlap between anchors reduces data
- Send a central view and sparse views

What we did today:
- Coder (25 x2 streams)

What it would mean conceptually:
- Coder (2 x2 streams)
Outperforming 3DoF+ Anchors – m43504

- 1 central ERP + 24 semi ERP views
Source View Pruning & Packing

❖ Based on m43504, build source view pruning module with RVS
❖ Packing module divides the sparse views into small size blocks
❖ If the block contains information more than threshold, it is added to packing view

\[ M_{Wi} = (X_{wi}, Y_{wi}, Z_{wi}) \]
\[ M_{Wc} = (X_{wc}, Y_{wc}, Z_{wc}) \]

Pixel-by-pixel comparison

Remove redundancy
\[ sv = iv \setminus (iv \cap cv) \]
Source View Pruning & Reconstruction

❖ Architecture

Source views → Anchor-coded views → Bitstream → Decoded views

Source views

Anchor-coded views

Bitstream

Decoded views

Sequences in Source Format

Central & Source View Selection

Source View Pruning with RVS

HEVC Encoding

HEVC Decoding

HEVC Encoding

HEVC Decoding

Source View Reconstruction

Measure WS-PSNR with Source Views
Source View Pruning & Reconstruction

❖ RD-curve
Conclusion

❖ Motivation
  ▪ 3DoF+ requires high resolution, large amount of videos
  ▪ Multi-view video transmission needs multiple decoders

❖ Proposed method
  ▪ Generate sparse view by removing overlap between views with RVS
  ▪ Divide sparse view into blocks and add them to packing view
  ▪ Reconstruct source view with central view and sparse views

❖ Results
  ▪ Source view pruning saves bitrate with small loss of PSNR
  ▪ Packing reduces the size of view to transmit

❖ Future work
  ▪ Color correction technique with illumination compensation
  ▪ Extensive experiment for optimal parameters