Versatile Video Coding for Video-based Point Cloud Compression

Dominik Mehlem
Institut für Nachrichtentechnik, RWTH Aachen University

SVCP 202(0)1 Berlin
Motivation

- Capturing and rendering point clouds becomes more and more popular
  - Compression of point cloud data inevitable [Tul+16]
Motivation

- Capturing and rendering point clouds becomes more and more popular
  - Compression of point cloud data inevitable [Tul+16]
- Video-based Point Cloud Compression (V-PCC) designed to be video coder agnostic [3Dg16]
Motivation

• Capturing and rendering point clouds becomes more and more popular
  – Compression of point cloud data inevitable [Tul+16]
• Video-based Point Cloud Compression (V-PCC) designed to be video coder agnostic [3Dg16]
• Current testing only performed with HEVC (CTC) [3DG20]

⇒ Approach shall also be proven for other video coding standards
Outline

1. Point Cloud Coding
2. Video-based Point Cloud Compression
3. Simulation Setup and Results
4. Conclusion and Outlook
**Point Cloud Coding**

- **Point cloud**
  - 3D collection of points representing objects, e.g. people, rooms or geographic maps
  - Typically captured at 30 fps
  - Approx. 700k to 2 million points per frame
  - Geometry with 10-11 bits and colour attributes with 8-10 bits

⇒ Raw data has high amount of bandwidth demand
Point Cloud Coding

- **Point cloud**
  - 3D collection of points representing objects, e.g. people, rooms or geographic maps
  - Typically captured at 30 fps
  - Approx. 700k to 2 million points per frame
  - Geometry with 10-11 bits and colour attributes with 8-10 bits
  
  ⇒ Raw data has high amount of bandwidth demand

- ISO/IEC WG7 currently developing two different point cloud compression standards [Sch+18]
  - Video-based Point Cloud Compression (VPCC)
    - mostly suitable for dynamic objects
  - Geometry-based Point Cloud Compression
    - mostly suitable for static scenes and dynamically acquired content
1. Point Cloud Coding

2. Video-based Point Cloud Compression

3. Simulation Setup and Results

4. Conclusion and Outlook
• General idea:
  – Project 3D point cloud onto 2D video frames
  – Compress video data with conventional video coder
• General idea:
  – Project 3D point cloud onto 2D video frames
  – Compress video data with conventional video coder

• Four streams generated during process
VPCC

• General idea:
  – Project 3D point cloud onto 2D video frames
  – Compress video data with conventional video coder

• Four streams generated during process
  – Atlas
    ▪ List of all candidate patches and their positions
    ▪ Not compressed by a conventional video coder
General idea:
- Project 3D point cloud onto 2D video frames
- Compress video data with conventional video coder

Four streams generated during process
- Atlas
  - List of all candidate patches and their positions
  - Not compressed by a conventional video coder
- Occupancy map
  - Binary representation of patch boundaries within 2D representation
• General idea:
  – Project 3D point cloud onto 2D video frames
  – Compress video data with conventional video coder

• Four streams generated during process
  – Atlas
    ▪ List of all candidate patches and their positions
    ▪ Not compressed by a conventional video coder
  – Occupancy map
    ▪ Binary representation of patch boundaries within 2D representation
  – Geometry video
    ▪ Greyscale video, specifying depth of patches
VPCC

• General idea:
  – Project 3D point cloud onto 2D video frames
  – Compress video data with conventional video coder

• Four streams generated during process
  – Atlas
    ▪ List of all candidate patches and their positions
    ▪ Not compressed by a conventional video coder
  – Occupancy map
    ▪ Binary representation of patch boundaries within 2D representation
  – Geometry video
    ▪ Greyscale video, specifying depth of patches
  – Attribute video
    ▪ Attribute information (e.g. color) of patches
VPCC

input point cloud frame

3D patch generation

Patch packing

Smoothing

Attribute image generation

Image padding

Video compression

Geometry image generation

Image padding

Video compression

Video compression

Patch sequence compression

Compressed bitstream

VPCC TMC2 encoder [3DG19]
Patch generation and packing

• Patch generation
  – Project every point on closest surface of bounding cube
  – Separate into patches
  – Bounding box around patches
Patch generation and packing

- Patch generation
  - Project every point on closest surface of bounding cube
  - Separate into patches
  - Bounding box around patches

Patches in 3D

Atlas patches
Patch generation and packing

- **Patch generation**
  - Project every point on closest surface of bounding cube
  - Separate into patches
  - Bounding box around patches

- **Patch packing**
  - Sort patches according to size
  - Pack from largest to smallest
  - Smaller patches fill "gaps"
  - 8 patch orientations possible
    - 4 rotations
    - 4 respective mirror images
VPCC video frame examples

Occupyancy map

Geometry video

Attribute video
Contents

1. Point Cloud Coding

2. Video-based Point Cloud Compression

3. Simulation Setup and Results

4. Conclusion and Outlook
Simulation Setup

- Implementation based on TMC2-14.1 [3DG] and VTM-13.0 [JVE] reference softwares
  - Coding of occupancy map, geometry video and attribute video with VVC
  - All-intra and random access lossy configurations
  - 7 sequences with 5 rate points each

Cat2 test set description

<table>
<thead>
<tr>
<th>Sequence</th>
<th>N</th>
<th># of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>loot</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>redandblack</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>soldier</td>
<td>10</td>
<td>250</td>
</tr>
<tr>
<td>longdress</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>basketball_player</td>
<td>11</td>
<td>64</td>
</tr>
<tr>
<td>dancer</td>
<td>11</td>
<td>64</td>
</tr>
</tbody>
</table>
Simulation Setup

- Implementation based on TMC2-14.1 [3DG] and VTM-13.0 [JVE] reference softwares
  - Coding of occupancy map, geometry video and attribute video with VVC
  - All-intra and random access lossy configurations
  - 7 sequences with 5 rate points each

<table>
<thead>
<tr>
<th>Sequence</th>
<th>geometry bits</th>
<th># of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>loot</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>redandblack</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>soldier</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>queen</td>
<td>10</td>
<td>250</td>
</tr>
<tr>
<td>longdress</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>basketball_player</td>
<td>11</td>
<td>64</td>
</tr>
<tr>
<td>dancer</td>
<td>11</td>
<td>64</td>
</tr>
</tbody>
</table>
Simulation Setup

- Implementation based on TMC2-14.1 [3DG] and VTM-13.0 [JVE] reference softwares
  - Coding of occupancy map, geometry video and attribute video with VVC
  - All-intra and random access lossy configurations
  - 7 sequences with 5 rate points each

<table>
<thead>
<tr>
<th>Sequence</th>
<th>geometry bits</th>
<th># of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>loot</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>redandblack</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>soldier</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>queen</td>
<td>10</td>
<td>250</td>
</tr>
<tr>
<td>longdress</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>basketball_player</td>
<td>11</td>
<td>64</td>
</tr>
<tr>
<td>dancer</td>
<td>11</td>
<td>64</td>
</tr>
</tbody>
</table>
Simulation Setup

• Implementation based on TMC2-14.1 [3DG] and VTM-13.0 [JVE] reference softwares
  – Coding of occupancy map, geometry video and attribute video with VVC
  – All-intra and random access lossy configurations
  – 7 sequences with 5 rate points each

• Anchor implementation based on TMC2-14.1 and HM16.20+SCM8.8 [Bos13] reference softwares

<table>
<thead>
<tr>
<th>Sequence</th>
<th>geometry bits</th>
<th># of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>loot</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>redandblack</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>soldier</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>queen</td>
<td>10</td>
<td>250</td>
</tr>
<tr>
<td>longdress</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>basketball_player</td>
<td>11</td>
<td>64</td>
</tr>
<tr>
<td>dancer</td>
<td>11</td>
<td>64</td>
</tr>
</tbody>
</table>
Table shows BD-rate savings [Bjo01] for geometry and attributes

Consistent performance gain for the sequences of the Cat2 test set is observed

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Geo</th>
<th>Att</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1</td>
<td>D2</td>
</tr>
<tr>
<td>loot</td>
<td>−17.3</td>
<td>−15.7</td>
</tr>
<tr>
<td>redandblack</td>
<td>−23.5</td>
<td>−23.3</td>
</tr>
<tr>
<td>soldier</td>
<td>−12.8</td>
<td>−13.2</td>
</tr>
<tr>
<td>queen</td>
<td>−27.9</td>
<td>−27.7</td>
</tr>
<tr>
<td>longdress</td>
<td>−20.3</td>
<td>−20.7</td>
</tr>
<tr>
<td>basketball</td>
<td>−29.1</td>
<td>−27.1</td>
</tr>
<tr>
<td>dancer</td>
<td>−27.7</td>
<td>−25.9</td>
</tr>
<tr>
<td>AVG</td>
<td>−22.6</td>
<td>−22.0</td>
</tr>
</tbody>
</table>
Results - AI

BD-rate change for the lossy all-intra case for simulations performed on the full sequences in %.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Geo D1</th>
<th>Geo D2</th>
<th>Luma</th>
<th>Cb</th>
<th>Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td>loot</td>
<td>−17.3</td>
<td>−15.7</td>
<td>−19.8</td>
<td>−23.4</td>
<td>−15.9</td>
</tr>
<tr>
<td>readandblack</td>
<td>−23.5</td>
<td>−23.3</td>
<td>−25.0</td>
<td>−20.7</td>
<td>−5.0</td>
</tr>
<tr>
<td>soldier</td>
<td>−12.8</td>
<td>−13.2</td>
<td>−18.1</td>
<td>−27.3</td>
<td>−29.0</td>
</tr>
<tr>
<td>queen</td>
<td>−27.9</td>
<td>−27.7</td>
<td>−21.9</td>
<td>−27.0</td>
<td>−21.9</td>
</tr>
<tr>
<td>longdress</td>
<td>−20.3</td>
<td>−20.7</td>
<td>−22.5</td>
<td>−21.2</td>
<td>−11.5</td>
</tr>
<tr>
<td>basketball</td>
<td>−29.1</td>
<td>−27.1</td>
<td>−22.4</td>
<td>−23.5</td>
<td>−23.3</td>
</tr>
<tr>
<td>dancer</td>
<td>−27.7</td>
<td>−25.9</td>
<td>−22.3</td>
<td>−27.0</td>
<td>−23.2</td>
</tr>
<tr>
<td>AVG</td>
<td>−22.6</td>
<td>−22.0</td>
<td>−21.7</td>
<td>−24.3</td>
<td>−18.6</td>
</tr>
</tbody>
</table>

- Table shows BD-rate savings [Bjo01] for geometry and attributes
- Consistent performance gain for the sequences of the Cat2 test set is observed
- Peak D1 performance gain for the “basketball” sequence
Results - AI

BD-rate change for the lossy all-intra case for simulations performed on the full sequences in %.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Geo</th>
<th>Luma</th>
<th>Cb</th>
<th>Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1</td>
<td>D2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>loot</td>
<td>−17.3</td>
<td>−15.7</td>
<td>−19.8</td>
<td>−23.4</td>
</tr>
<tr>
<td>readandblack</td>
<td>−23.5</td>
<td>−23.3</td>
<td>−25.0</td>
<td>−20.7</td>
</tr>
<tr>
<td>soldier</td>
<td>−12.8</td>
<td>−13.2</td>
<td>−18.1</td>
<td>−27.3</td>
</tr>
<tr>
<td>queen</td>
<td>−27.9</td>
<td>−27.7</td>
<td>−21.9</td>
<td>−27.0</td>
</tr>
<tr>
<td>longdress</td>
<td>−20.3</td>
<td>−20.7</td>
<td>−22.5</td>
<td>−21.2</td>
</tr>
<tr>
<td>basketball</td>
<td>−29.1</td>
<td>−27.1</td>
<td>−22.4</td>
<td>−23.5</td>
</tr>
<tr>
<td>dancer</td>
<td>−27.7</td>
<td>−25.9</td>
<td>−22.3</td>
<td>−27.0</td>
</tr>
<tr>
<td>AVG</td>
<td>−22.6</td>
<td>−22.0</td>
<td>−21.7</td>
<td>−24.3</td>
</tr>
</tbody>
</table>

- Table shows BD-rate savings [Bjo01] for geometry and attributes
- Consistent performance gain for the sequences of the Cat2 test set is observed
- Peak D1 performance gain for the “basketball” sequence
- Generally higher performance gain for vox11 sequences
### Results - RA

BD-rate change for the lossy random access case for simulations performed on the full sequences in %.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Geo D1</th>
<th>Geo D2</th>
<th>Luma</th>
<th>Cb</th>
<th>Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td>loot</td>
<td>-24.7</td>
<td>-24.4</td>
<td>-26.0</td>
<td>-27.2</td>
<td>-21.2</td>
</tr>
<tr>
<td>redandblack</td>
<td>-25.6</td>
<td>-26.2</td>
<td>-24.8</td>
<td>-23.3</td>
<td>-16.7</td>
</tr>
<tr>
<td>soldier</td>
<td>-21.6</td>
<td>-21.6</td>
<td>-23.0</td>
<td>-27.9</td>
<td>-26.6</td>
</tr>
<tr>
<td>queen</td>
<td>-32.3</td>
<td>-32.1</td>
<td>-26.0</td>
<td>-33.7</td>
<td>-29.9</td>
</tr>
<tr>
<td>longdress</td>
<td>-25.8</td>
<td>-26.3</td>
<td>-27.1</td>
<td>-29.4</td>
<td>-21.0</td>
</tr>
<tr>
<td>basketball</td>
<td>-31.9</td>
<td>-31.0</td>
<td>-25.4</td>
<td>-24.6</td>
<td>-20.3</td>
</tr>
<tr>
<td>dancer</td>
<td>-30.7</td>
<td>-29.6</td>
<td>-25.8</td>
<td>-26.9</td>
<td>-17.4</td>
</tr>
<tr>
<td>AVG</td>
<td>-27.5</td>
<td>-27.3</td>
<td>-25.4</td>
<td>-27.6</td>
<td>-21.9</td>
</tr>
</tbody>
</table>

- Table shows BD-rate savings [Bjo01] for geometry and attributes
- Consistent performance gain for the sequences of the Cat2 test set is observed
- Peak D1 performance gain for the “queen” sequence
- Generally higher performance gain for vox11 sequences
**Results - RA**

BD-rate change for the lossy random access case for simulations performed on the full sequences in %.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Geo</th>
<th>Att</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1</td>
<td>D2</td>
</tr>
<tr>
<td>loot</td>
<td>−24.7</td>
<td>−24.4</td>
</tr>
<tr>
<td>redandblack</td>
<td>−25.6</td>
<td>−26.2</td>
</tr>
<tr>
<td>soldier</td>
<td>−21.6</td>
<td>−21.6</td>
</tr>
<tr>
<td>queen</td>
<td>−32.3</td>
<td>−32.1</td>
</tr>
<tr>
<td>longdress</td>
<td>−25.8</td>
<td>−26.3</td>
</tr>
<tr>
<td>basketball</td>
<td>−31.9</td>
<td>−31.0</td>
</tr>
<tr>
<td>dancer</td>
<td>−30.7</td>
<td>−29.6</td>
</tr>
<tr>
<td>AVG</td>
<td>−27.5</td>
<td>−27.3</td>
</tr>
</tbody>
</table>

- Table shows BD-rate savings [Bjo01] for geometry and attributes
- Consistent performance gain for the sequences of the Cat2 test set is observed
- Peak D1 performance gain for the “queen” sequence
- Generally higher performance gain for vox11 sequences
- Gain for RA is larger than for AI
1. Point Cloud Coding

2. Video-based Point Cloud Compression

3. Simulation Setup and Results

4. Conclusion and Outlook
Conclusion and Outlook

• Conclusion
  – Utilizing VVC instead of HEVC in the VPCC framework shows comparable gain to plain comparison of VVC and HEVC
  – VVC proves to be versatile
⇒ The video coder agnostic approach shows VPCC to be future-proof
Conclusion and Outlook

• Conclusion
  – Utilizing VVC instead of HEVC in the VPCC framework shows comparable gain to plain comparison of VVC and HEVC
  – VVC proves to be versatile
  ⇒ The video coder agnostic approach shows VPCC to be future-proof

• Outlook
  – Implementation of point cloud coding specific tools
    ▪ E.g. Occupancy map based RDO
  – No consistent gain for the lossless approach
    ▪ Further investigation in encoder configurations necessary
    ▪ Might just not be better
Literature


Thank you for your attention

Any questions?
mehlem@ient.rwth-aachen.de