A Smart Reference Picture Resampling Approach for VVC

Tianliang Fu (PKU), Kai Zhang, Yue Li, Li Zhang (ByteDance)
Shanshe Wang, Siwei Ma (PKU)
Agenda

- Background
- Proposed method
- Simulation results
- Conclusion
Background-Reference Picture Resampling

- Resampling-based coding
  - Down-sampling before compression and up-sampling the decoded video for reconstruction

- Use case
  - Compressing high-resolution videos at low bitrate
  - Bitrate adaption in video telephony and conferencing
  - The frequent active speaker changes in the multi-party video conferencing
  - Fast start for streaming application
  - Adaptive stream switching in streaming application
Background - RPR in VVC

- **Before VVC**
  - Inserting an IDR or IRAP with a new resolution setting
  - Consuming more bits and time than an inter-coded picture

- **RPR in VVC**
  - Support the spatial resolution change without inserting an IDR or IRAP picture
  - The height and width of the current picture are signaled directly in PPS
  - The scaling ratios for horizontal and vertical directions are both arbitrary

- **RPR as a coding tool**
  - Multiple-pass encoded to enable a manual selection between low-resolution coding and full-resolution coding
  - A much higher encoding time and is not desirable in low-delay scenarios
Smart RPR

- Adaptive decision on coding resolution

\[ D_{down} < D_{thre} \]

Downsampling-based coding

\[ D_{thre} = e^{a \cdot QP - b} \]

\[ y = e^{(0.175x-3.5)} \]
Coding parameters setting for Luma component

- Align rate

\[ R = \alpha \times QP_{step}^\beta \]

\[ QP_{step} = 2^{\frac{QP - 4}{6}} \]

\[ QP_{step} = QP_{full} - 6 \]
Coding parameters setting for Chroma component

- **Eliminate the chroma loss**

The critical bitrate of chroma components is lower than that of the luma component when the chroma QP offset increases from 6 to 10, the critical bitrate gradually gets larger.

\[ QP_{chroma\_down} = QP_{chroma\_full} - 9 \]
Smart RPR

- Framework

\[ D_{thre} = e^{0.175 \cdot QP - 3.5} \]

Input

Down

Up

Calculate \( D_{down} \)

\( D_{down} < D_{thre} \)

YES

NO

Downsampling-based

Encoder

QP adjustment

Encoder

Output
Simulation results

- **Experiment configurations**
  - **Platform: VTM-12.0**
  - **Sequences: 4K**

<table>
<thead>
<tr>
<th>Class</th>
<th>Sequence</th>
<th>Width</th>
<th>Height</th>
<th>FrameRate</th>
<th>BitDepth</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Tango2</td>
<td>3840</td>
<td>2160</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>FoodMarket4</td>
<td>3840</td>
<td>2160</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Campfire</td>
<td>3840</td>
<td>2160</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>CatRobot</td>
<td>3840</td>
<td>2160</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>A2</td>
<td>DaylightRoad2</td>
<td>3840</td>
<td>2160</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>ParkRunning3</td>
<td>3840</td>
<td>2160</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>Added</td>
<td>MountainBay2</td>
<td>3840</td>
<td>2160</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>OberbaumSpree</td>
<td>3840</td>
<td>2160</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>RaceNight</td>
<td>3840</td>
<td>2160</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>TiergartenParkway</td>
<td>3840</td>
<td>2160</td>
<td>60</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Y-PSNR</th>
<th>U-PSNR</th>
<th>V-PSNR</th>
<th>EncT</th>
<th>DecT</th>
<th>DeltaBitrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tango2</td>
<td>-7.40%</td>
<td>-16.94%</td>
<td>-22.26%</td>
<td>72%</td>
<td>48%</td>
<td>5.0%</td>
</tr>
<tr>
<td>FoodMarket4</td>
<td>-8.36%</td>
<td>-18.39%</td>
<td>-23.18%</td>
<td>73%</td>
<td>40%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Campfire</td>
<td>2.37%</td>
<td>-0.58%</td>
<td>-1.46%</td>
<td>103%</td>
<td>84%</td>
<td>4.7%</td>
</tr>
<tr>
<td>CatRobot</td>
<td>-1.51%</td>
<td>0.46%</td>
<td>-2.24%</td>
<td>82%</td>
<td>67%</td>
<td>0.9%</td>
</tr>
<tr>
<td>DaylightRoad2</td>
<td>-0.99%</td>
<td>-4.25%</td>
<td>-5.08%</td>
<td>102%</td>
<td>84%</td>
<td>1.5%</td>
</tr>
<tr>
<td>ParkRunning3</td>
<td>-2.37%</td>
<td>32.46%</td>
<td>19.49%</td>
<td>79%</td>
<td>65%</td>
<td>4.2%</td>
</tr>
<tr>
<td>MountainBay2</td>
<td>-2.75%</td>
<td>-4.11%</td>
<td>-29.77%</td>
<td>69%</td>
<td>52%</td>
<td>6.0%</td>
</tr>
<tr>
<td>OberbaumSpree</td>
<td>-3.83%</td>
<td>-27.29%</td>
<td>-29.11%</td>
<td>65%</td>
<td>48%</td>
<td>5.6%</td>
</tr>
<tr>
<td>RaceNight</td>
<td>-2.32%</td>
<td>-11.28%</td>
<td>-12.13%</td>
<td>84%</td>
<td>70%</td>
<td>0.5%</td>
</tr>
<tr>
<td>TiergartenParkway</td>
<td>-0.08%</td>
<td>-2.93%</td>
<td>-2.44%</td>
<td>100%</td>
<td>91%</td>
<td>0.7%</td>
</tr>
<tr>
<td>Average</td>
<td>-2.72%</td>
<td>-5.29%</td>
<td>-10.82%</td>
<td>83%</td>
<td>65%</td>
<td>3.1%</td>
</tr>
</tbody>
</table>
Simulation results

- Subjective quality

VTM12.0 + downsampling-based coding

VTM12.0 anchor

original
We present a smart-RPR approach that enjoys improved compression efficiency and lower encoding/decoding complexity.

- A frame is first down- and up-sampled without encoding and the resampling distortion is calculated by comparing the up-sampled frame with the original frame.
- The RPR for this frame will be enabled if the resampling distortion is less than a distortion threshold conditioning on QP.
- The luma and chroma QPs of the downsampled frame are adjusted to avoid excessive bitrate fluctuation and to achieve a better coding performance.

Smart RPR achieves about 2.72%, 5.29%, and 10.82% BD-rate savings on average for Y, Cb, and Cr components respectively.

Both encoding and decoding time savings are observed and the bitrate fluctuation is minor, i.e. 3.1% on average.
Thanks for your attention!

futl@pku.edu.cn