AN ENHANCED REFERENCE STRUCTURE FOR REFERENCE PICTURE RESAMPLING (RPR) IN VVC

Tianliang Fu¹, Kai Zhang², Li Zhang³, Shanshe Wang¹, Siwei Ma¹

¹Department of Computer Science and Technology, Peking University.
²Bytedance Inc., San Diego CA. 92122 USA.
³{futl, sswang, swma}@pku.edu.cn, {zhangkai.video, lizhang.idm}@bytedance.com

ABSTRACT

Reference picture resampling (RPR) is an important feature for real-time communication in Versatile Video Coding, which supports adaptive resolution change without introducing an instantaneous decoder refresh (IDR) or intra random access picture (IRAP). With RPR, considerable coding gain is achieved while bitrate leap still exists. In this paper, an enhanced reference structure for RPR is proposed. First, the high-resolution (HR) references are used to replace the low-resolution (LR) references to provide more accurate prediction. Furthermore, an adaptive decision on the number of HR references kept in RPL is applied to achieve more performance, which utilizes the history information in the current period. Simulation results show about 9% and 5% BD-rate savings on average can be achieved for scaling ratio of 2:1 and 1.5:1 compared to VTM-10.2 under RPR configurations. For sequences with slight motions, the coding gain is up to 20%. Besides, the bitrate explosion in RPR is reduced generally, where over 50% reductions of bitrates at explosion point is observed.

Index Terms— RPR, reference structure, resolution change

1. INTRODUCTION

With the development of emerging mobile Internet era, the demand for immersive video experiences is growing fast. The video data is exploding with the usage of higher spatial resolutions, frame rates and dynamic ranges, which dramatically increases the bitrates required to store and transmit videos. The increasing video data volume poses a great challenge to current bandwidth restrictions and requires greater compression efficiency than that offered by the current generation of video coding standard, formally designated as ITU-T H.266 and ISO/IEC 23090-3, promises a major improvement in video compression relative to its predecessors. It can offer roughly double the coding efficiency.

One of the target applications of VVC is real-time communication, such as video conferencing and live video. These applications can benefit from the reference picture resampling (RPR) feature for adaptive video transmission under different network conditions in VVC, supporting spatial resolution changes within a video sequence. More specifically, when the network condition gets worse with lower available bandwidth, the encoder can adapt to it by encoding low-resolution (LR) frames. Meanwhile, the encoder could transmit the original full-resolution video when the network condition gets better.

The basic idea, down-sampling the video frame before compression and getting reconstructed frame by up-sampling after decompression, is first introduced and exploited for image compression standard [4]. Ilgin and Chaparro et al. [5] extends this idea to DCT domain for low bitrate video coding. Based on the previous work of adaptive down-sampling to improve image compression [6], Nguyen et al. [7] implements an adaptive spatial resampling method for better video compression at low bitrate. Dong et al. [8] proposes a practical algorithm to find the optimal down-sampling ratio that balances the distortions caused by down-sampling and coding. Hu et al. [9] performs rate-distortion optimization (RDO) to determine the coding structure between regular coding and down-sampling coding. In [10], state-of-the-art video coding and super-resolution (SR) techniques are combined to improve video compression both in terms of coding efficiency and complexity. In [11], the dependency between resampling cost and the quantization parameter (QP) threshold determining whether to resample is studied, based on which a spatial resolution adaptation scheme for video compression was proposed. An adaptive resolution change (ARC) method [12] is adopted in VVC to provide better adaptation of the video transmission under dynamic network environments.

In the literature, most researches regarding adaptive resolution changing concentrate on the devise of down-sampling or up-sampling procedures. Less efforts have conducted on the reference structure for RPR. In this article,
an enhanced reference structure is proposed for RPR in VVC. First, HR references are used to replace the LR references in RPL without enough HR references to provide more accurate prediction and reduce the bitrate surge. Second, an adaptive decision on the number of HR references kept in RPL is proposed to enhance the compression performance based on the ratio of blocks referring from HR references.

The rest of this paper is organized as follows: Section 2 reviews the technical details of RPR in VVC. Section 3 describes the proposed enhanced reference structure for RPR. Simulation results are reported in Section 4 and conclusions are drawn in Section 5.

2. REFERENCE PICTURE RESAMPLING IN VVC

In the previous video compression standards, such as H.264/AVC and H.265/HEVC, changing picture resolution can only be implemented by inserting an instantaneous decoder refresh (IDR) or intra random access picture (IRAP) with a new resolution setting. An IDR or IRAP picture with reasonable quality tends to consume more bits than an inter-coded picture and correspondingly take more time to decode. The low-latency buffer conditions are broken as the end-to-end delay will increase, providing poor user experience.

To enable adaptive resolution changes without the insertion of an IDR or IRAP, the RPR feature is supported in VVC. In the sequence parameter set (SPS), a list of picture resolutions is signaled and an index to the list is signaled in the picture parameter set (PPS) to specify the size of an individual picture. Therefore, an arbitrary scaling ratio for horizontal and vertical directions is supported.

When the resolutions of current picture and reference picture are different, the resampling process is necessary to enable motion compensation (MC) as resolution changes are performed within a coding video sequence. Block-based resampling method is adopted in VVC and reference pictures are stored in the decoded picture buffer (DPB) with the coded size only.

In VVC, the common test conditions (CTC) for RPR configurations [13] are established to evaluate the coding performance. Real-time communication with high requirements of low delay is the target application of RPR. Therefore, the same low delay (LD) configurations as in SDR CTC [14] are used. In terms of resolution changes, two scaling ratios, 2:1 and 1.5:1 in both dimensions, are included. Besides, the resolutions of pictures change every 0.5 seconds. For performance evaluation, PSNR1 and PSNR2 as shown in Fig. 1 should be computed. In Fig. 1, PSNR1 is computed between the resampled source picture and reconstructed picture; PSNR2 is computed between picture with original resolution and the resampled version of reconstructed picture whose resolution is the same as original resolution.

3. PROPOSED METHOD

In the past years, discussions for RPR are focused on the resampling procedures and methods of referring from pictures with different resolutions efficiently. The benefits of designing targeted reference structure for RPR are ignored. Also, although RPR can change the resolution of any frames without introducing an IDR or IRAP, reducing the bitrate surge brought by IDR or IRAP, the heavy burden on bits cost still exists. As shown in Fig. 2, the bits cost is no longer a heavy burden with the introduction of RPR for resolution decreased pictures (RDP). While for resolution increased pictures (RIP), the bits cost shoots up, 2 to 3 times as much as that of normal inter-coded pictures with the same resolution. In order to solve bitrate surge in RPR and further improve the coding performance, an enhanced reference structure for RPR is proposed, including replacing LR references in RPL with HR references and adaptive decision on the number of HR references kept in RPL.

3.1. Replacing LR References with HR References

It can be observed in Fig. 3 that the references of RIP, whose picture order count (POC) is 48, are all LR pictures. Compared with HR references, LR references cannot provide accurate enough prediction for RIP. In particular, the original HR picture is shown in Fig.4-(a) and the difference between original HR reference and up-sampled picture from LR reference is shown in Fig. 4-(b). It is observed that up-sampled picture loses much detail both for luma and chroma components, which straightforward degrades the accuracy of prediction samples. HR references are introduced to replace the original LR references in the RPL for predicting to-be-code pictures more accurately.

Replacing LR references in RPL with HR references benefits not only RIP but also pictures with few HR references. The introduction of HR references is applied in all frames consequently. First, the number of HR references in the original RPL is counted. Second, the temporal nearest HR references are used to replace the temporal farthest LR references in original RPL when the counted number is less.
25% > ThreNumHRRef > 75%, ThreNumHRRef ++

Y

N

Fig. 6 Flowchart of adaptive adding HR reference.
4. EXPERIMENTAL RESULTS

To evaluate coding performance of the enhanced reference structure, the proposed method is implemented on the VVC test model VTM 10.2. The simulations are performed under the CTC for RPR configurations [13]. Tab. 1 and Tab. 2 illustrate the overall results in terms of BD-rate [15] with PSNR1 and PSNR2 for scaling ratio setting to 2:1 respectively. Tab. 3 and Tab. 4 illustrate the coding performance with scaling ratio 1.5:1. It can be observed that the proposed method performs well and can achieve significant coding performance improvement and save some encoding and decoding time. For 2:1 scaling ratio, the BD-rate savings with PSNR1 and PSNR2 are 8.70%, 11.23%, 10.81% and 9.13%, 11.54%, 11.12% for Y, U and V components, respectively; For 1.5:1 scaling ratio, the BD-rate with PSNR1 and PSNR2 is reduced by 5.27%, 6.16%, 6.00% and 5.46%, 6.22%, 6.12% for Y, U and V components, respectively. The largest gain comes from the sequences of Class E, along with less and slow motions. By contrast, lower performance gains are observed in the sequences of Class B, especially those with complex and severe motions. Meanwhile, comparing Tab. 1 with Tab. 2, it can be found that coding gain increases with the scaling ratio. The reason is that less low frequency information is retained as the scaling ratio increases. As a result, the prediction accuracy of LR reference provides becomes lower and the benefits from replacing LR reference with HR reference increase.

Regarding the computational complexity, we can observe that the encoding time is reduced and the decoding time is almost unchanged. In generally, the introduced HR references can predict more accurately than LR references. Thus, the motion estimation and interpolation of fraction pixel can be skipped such that the encoding time saving can be observed. Conversely, the proposed reference structure has little impact on the decoding process. As a result, no decoding time changing is observed.

Tab. 1. Coding Performance of the Proposed Method with PSNR1 for 2:1 Scaling Ratio

<table>
<thead>
<tr>
<th>Seq</th>
<th>Y</th>
<th>U</th>
<th>V</th>
<th>Enc</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClassB</td>
<td>-4.05%</td>
<td>-5.96%</td>
<td>-6.81%</td>
<td>98%</td>
<td>101%</td>
</tr>
<tr>
<td>ClassC</td>
<td>-6.31%</td>
<td>-8.27%</td>
<td>-7.93%</td>
<td>98%</td>
<td>96%</td>
</tr>
<tr>
<td>ClassE</td>
<td>-19.66%</td>
<td>-23.96%</td>
<td>-21.30%</td>
<td>91%</td>
<td>95%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>-8.70%</strong></td>
<td><strong>-11.23%</strong></td>
<td><strong>-10.81%</strong></td>
<td><strong>96%</strong></td>
<td><strong>98%</strong></td>
</tr>
</tbody>
</table>

Tab. 2. Coding Performance of the Proposed Method with PSNR2 for 2:1 scaling ratio

<table>
<thead>
<tr>
<th>Seq</th>
<th>Y</th>
<th>U</th>
<th>V</th>
<th>Enc</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClassB</td>
<td>-4.25%</td>
<td>-6.16%</td>
<td>-7.09%</td>
<td>98%</td>
<td>101%</td>
</tr>
<tr>
<td>ClassC</td>
<td>-6.42%</td>
<td>-8.85%</td>
<td>-8.56%</td>
<td>98%</td>
<td>96%</td>
</tr>
<tr>
<td>ClassE</td>
<td>-20.88%</td>
<td>-24.11%</td>
<td>-21.24%</td>
<td>91%</td>
<td>95%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>-9.13%</strong></td>
<td><strong>-11.54%</strong></td>
<td><strong>-11.12%</strong></td>
<td><strong>96%</strong></td>
<td><strong>98%</strong></td>
</tr>
</tbody>
</table>

Tab. 3. Results with PSNR1 for 1.5:1 scaling ratio

<table>
<thead>
<tr>
<th>Seq</th>
<th>Y</th>
<th>U</th>
<th>V</th>
<th>Enc</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClassB</td>
<td>-1.95%</td>
<td>-2.72%</td>
<td>-3.54%</td>
<td>96%</td>
<td>99%</td>
</tr>
<tr>
<td>ClassC</td>
<td>-4.13%</td>
<td>-6.18%</td>
<td>-5.83%</td>
<td>97%</td>
<td>99%</td>
</tr>
<tr>
<td>ClassE</td>
<td>-12.32%</td>
<td>-11.87%</td>
<td>-10.33%</td>
<td>93%</td>
<td>101%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>-5.27%</strong></td>
<td><strong>-6.16%</strong></td>
<td><strong>-6.00%</strong></td>
<td><strong>96%</strong></td>
<td><strong>99%</strong></td>
</tr>
</tbody>
</table>

Tab. 4. Results with PSNR2 for 1.5:1 scaling ratio

<table>
<thead>
<tr>
<th>Seq</th>
<th>Y</th>
<th>U</th>
<th>V</th>
<th>Enc</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClassB</td>
<td>-2.03%</td>
<td>-2.71%</td>
<td>-3.63%</td>
<td>96%</td>
<td>99%</td>
</tr>
<tr>
<td>ClassC</td>
<td>-4.26%</td>
<td>-6.40%</td>
<td>-6.06%</td>
<td>97%</td>
<td>99%</td>
</tr>
<tr>
<td>ClassE</td>
<td>-12.79%</td>
<td>-11.84%</td>
<td>-10.34%</td>
<td>93%</td>
<td>101%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>-5.46%</strong></td>
<td><strong>-6.22%</strong></td>
<td><strong>-6.12%</strong></td>
<td><strong>96%</strong></td>
<td><strong>99%</strong></td>
</tr>
</tbody>
</table>

Fig. 7. Bits cost comparison.

Furthermore, the proposed method makes a great contribution in solving the existing bitrate leap in RPR. As shown in Fig. 7, the bitrate cost in RIP drops tremendously, where more than 50% reduction can be observed. With the proposed method, the bitrate cost of RIP is only slightly larger than normal pictures with the same resolution, which mitigates the transmission burden and is anticipated to provide better quality of experience.

5. CONCLUSION

In this paper, HR references are introduced to the RPL without enough HR references, replacing the original LR references. When the number of HR references is less than the threshold, the temporal nearest HR reference is used to replace the temporal farthest LR reference in the original RPL. Moreover, the adaptive decision on the number of HR references kept in RPL is applied to further improve the performance. The adaptive decision strategy utilizes the ratio of blocks referring from HR reference to all blocks in the current period. Simulation results show 8.71% and 9.13% BD-rate savings with PSNR1 and PSNR2 on average can be achieved for 2:1 scaling ratio. For 1.5:1 scaling ratio, the BD-rate savings with PSNR1 and PSNR2 are 5.27% and 5.46% respectively. Besides, the bitrate cost in RIP drops greatly, where more than 50% reduction can be observed.

6. ACKNOWLEDGMENT

This work was done while the author was a research intern in ByteDance Inc.
7. REFERENCES


