Super-resolution video coding with additional residual data coding

Byeongdoo Choi*, Andrew Segall, Kiran Misra
Amazon Prime Video, 2021 7th Ave, Seattle, WA, USA 98121

ABSTRACT

Super-resolution video coding describes the process of coding video at lower resolution and upsampling the result. This process is included in the AV1 standard, which ensures the same super-resolution process is employed on all receiving devices. Regrettably, the design is limited to horizontal scaling with a maximum scale factor of two. In this paper, we analyze the benefit of enabling two-dimensional upsampling with larger scale factors. Additionally, we consider the value of sending residual information to correct the super-resolution output. Results show a 6.3% and 5.6% improvement in coding efficiency for UHD SDR and UHD HDR content.

Keywords: Video coding, super-resolution, residual coding, reference frame resampling

1. INTRODUCTION

Super-resolution video coding techniques have gained significant attention in recent years due to the increasing demand for high-quality video content. These techniques aim to compresses video at a lower resolution and then utilize super-resolution concepts to increase the decoded resolution before display. This resolution increase is either achieved with the explicit use of a super-resolution algorithm or with the sequential application of an up-sampling operation followed by a restoration process that recovers high frequency information and removes artifacts. Two approaches are widely researched and adopted for incorporating super-resolution in modern video coding systems. The first is to perform resolution conversion outside of the coding loop. This is used in commercial video streaming services and relies on adaptive streaming protocols such as DASH, HLS and/or Smooth Streaming. While widely deployed, it does have two disadvantages. First, the super-resolution factor can only change on a segment boundary; second, the super-resolution process typically varies between receiving devices. The second approach for performing super-resolution addresses these disadvantages and incorporates resolution conversion directly into the video codec. This is referred to as in-loop processing, and it was included in the AV1 video coding standard. Fig. 1 illustrates the overall in-loop super-resolution framework in AV1. According to this framework, on the encoder side, a source image is first downscaled as a pre-processing step and encoded at a lower resolution. The deblocking filter and the constrained directional enhancement filter (CDEF) are then applied at the lower resolution to remove blocking and ringing artifacts while preserving edges. This is followed by a normative upsampling process to create a rendition at the full source resolution. Finally, the loop-restoration is carried out at the original full resolution to restore the high frequencies lost during the prior resampling and quantization process. A similar process is followed on the decoder side, and the reconstructed frames at the full resolution are used as reference frames for inter-prediction of the subsequence frames. The benefits of super-resolution framework in AV1 are reported in [1]. However, we observe the AV1 super-resolution design may not fully achieve the advantages of in-loop super-resolution, as the super-resolution process is limited to one-dimensional scaling and a maximum scale factor of two.

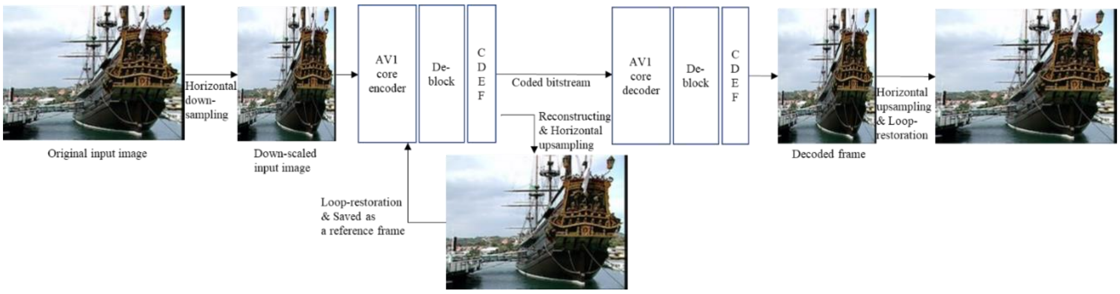

Figure 1.Super-resolution framework in AV1

*{bdchoi, asegall, misrakir}@amazon.com; phone 1 206 740-7776; amazon.com

Several related papers have explored the challenges and possibilities of super-resolution video coding. Khani et al. [2] presented a deep learning-based approach for super-resolution video coding. Lin et al. [3] proposed a video coding framework combining deep learning-based super-resolution with a traditional codec, and Wang et al. [4] investigated rate-distortion optimization for efficient super-resolution video coding. These papers have reported good compression efficiency by utilizing complex neural network models and/or encoding strategies, However, due to the accompanying computational complexity, it is challenging to use these approaches as normative in-loop coding tools in video codecs.

In the paper, we propose three main improvements of the super-resolution process that builds on the AV1 design. First, the super-resolution process is extended to two dimensions and handles both horizontal and vertical scaling factors with extended scaling ranges beyond a factor of two. Second, to further restore the lost details of the super-resolved frame, additional residual data is coded and transmitted. This residual data captures the high-frequency information lost during the compression and resampling process, and it also provides an additional lever to manage visual quality through the assignment of bits to the higher and lower resolutions. Finally, the management of residual frames and motion vector information during the super-resolution process is improved by adopting a frame unit concept. This frame unit enables the efficient storage of super-resolved frames and their associated motion vectors, which are typically used for the prediction of motion vectors in subsequent frames.

The remainder of this paper is organized as follows: Section 2 presents the technical details of the proposed two-dimensional super-resolution process with additional residual coding. Section 3 provides the experimental results, evaluating the coding efficiency improvements achieved by the proposed approach. Finally, Section 4 concludes the paper and discusses future research directions.

2. TWO-DIMENSIONAL SUPER-RESOLUTION FOR VIDEO CODING WITH RESIDAUL CODING

This section introduces the proposed approach, which extends the AV1 design to enable two-dimensional super-resolution with varying scaling factors. Two main structural modifications from the existing AV1 super-resolution framework are described in the following subsections.

2.1 Upsampling and restoration process

Fig. 2 illustrates the decoding process of the proposed super-resolution video coding incorporating two-dimensional upsampling, restoration filtering and high-resolution residual coding. Following the deblocking and CDEF processes, the reconstructed frame is upsampled using a two-dimensional filter. To ensure the preservation of intricate image details while maintaining an acceptable computational complexity within the context of an in-loop coding tool for video codecs, we use a 10-tap Lanczos filter. In prior research [5], this filter has demonstrated remarkable efficacy in minimizing undesirable blurring effects when images are resampled.

The upsampled frame is then processed by a restoration algorithm that further enhances the visual quality by recovering high-frequency information and reducing artifacts. The restoration process significantly contributes to improving both subjective and objective visual quality by allowing for the recovery of detailed texture information that may be lost during the quantization and resampling processes. Finally, based on the observation that filtering-based restoration processes have limits to recover the lost details, additional residual data is coded by reusing the overlay coding mechanism at higher resolution and stored separately in the bitstream. The residual data captures the lost high-frequency details and provides an additional means to manage visual quality by allocating bits between the higher and lower resolutions. By incorporating residual data coding, the subjective and objective visual quality can be noticeably improved.

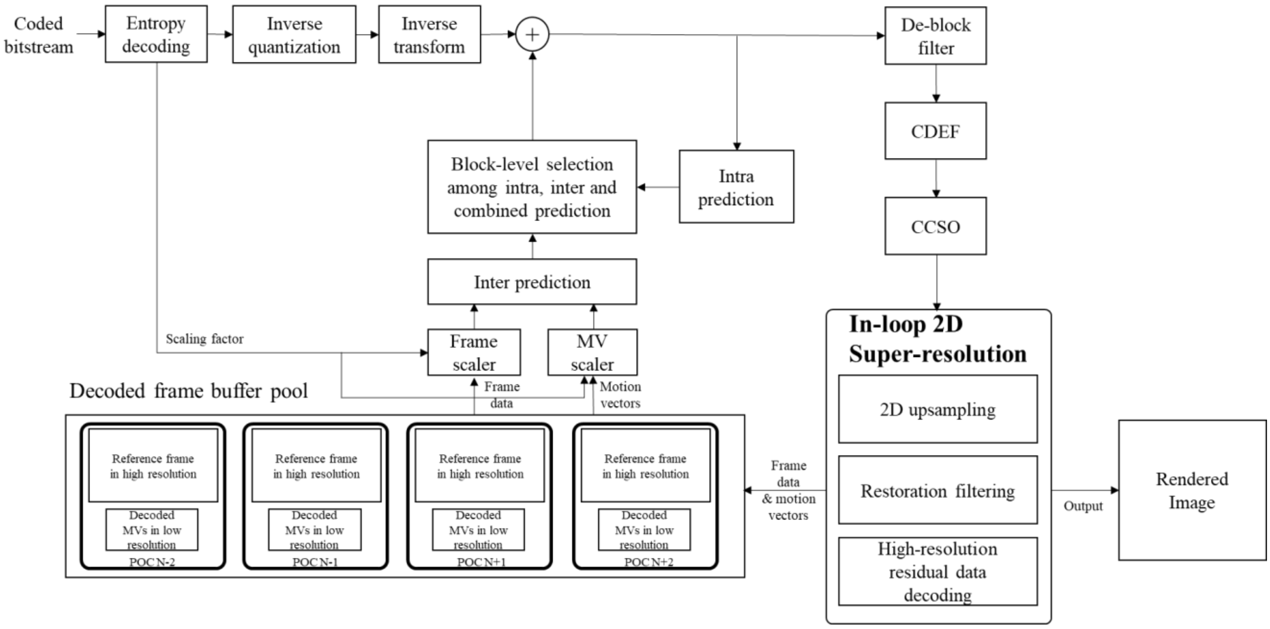

Figure 2. The decoding structure with the proposed 2D super-resolution with a residual frame coding.

2.2 Management of decoded frame buffers and motion vectors

A decoded picture buffer pool stores one or more frame units that may be referenced by the following frames in decoding order. The management of super-resolved frames, residual frames, and their associated motion vectors within the decoded picture buffer pool is critical for coding efficiency. By minimizing a buffer size to store multi-layered frames of the same image, we seek to maximize the number of effective reference frames. To address this, we introduce the concept of frame units, which are collections of open bitstream units (OBUs) belonging to the same frame. Fig. 3 provides an overview of our frame unit approach and illustrates the storage of high-resolution output frames, optional residual data and low-resolution motion vectors in a single buffer slot. This approach enables high-resolution pictures to be used for motion compensation in the following frames in decoding order, while lower-resolution motion vectors can also be used for subsequent motion coding in the following frames. This approach offers several advantages over alternative methods. Firstly, it eliminates the need to use separate buffer slots for storing high-resolution and low-resolution decoded frames. Without employing the frame unit concept, the reconstructed frames with additional residual data cannot be stored with the low-resolution motion vectors in a single buffer, since the residual frame is coded by reusing the overlay frame coding mechanism has motion vectors in the higher resolution. This efficient use of buffer space allows for longer temporal prediction, which enhances coding efficiency when the number of frame buffer slots is limited. For example, when the maximum number of frame buffer slots is equal to eight, the proposed approach can store eight reference frames. By contrast, traditional approaches (such as AV1) are limited to store four reference frames with the additional residual coding structure. This means the proposed approach can increase the number of effective reference frames by two under the same prediction structure. Additionally, the coding of residual data frames is carefully constrained to reduce complexity and enable parallel processing. For instance, all motion vectors within the residual data frames are set to zero, simplifying the decoding process and facilitating efficient parallelization. Fig.4 shows an example of layered coding structure with residual frames, which is well compromised between coding efficiency and support of scalable coding features. It offers distinct advantages, including the efficient utilization of buffer slots, longer temporal prediction, association of motion vectors with high-resolution data, and constrained residual data coding. It is worth noting that alternative coding patterns such as a layered coding structure to support a full spatial resolution scalability are also supported and viable within our framework.

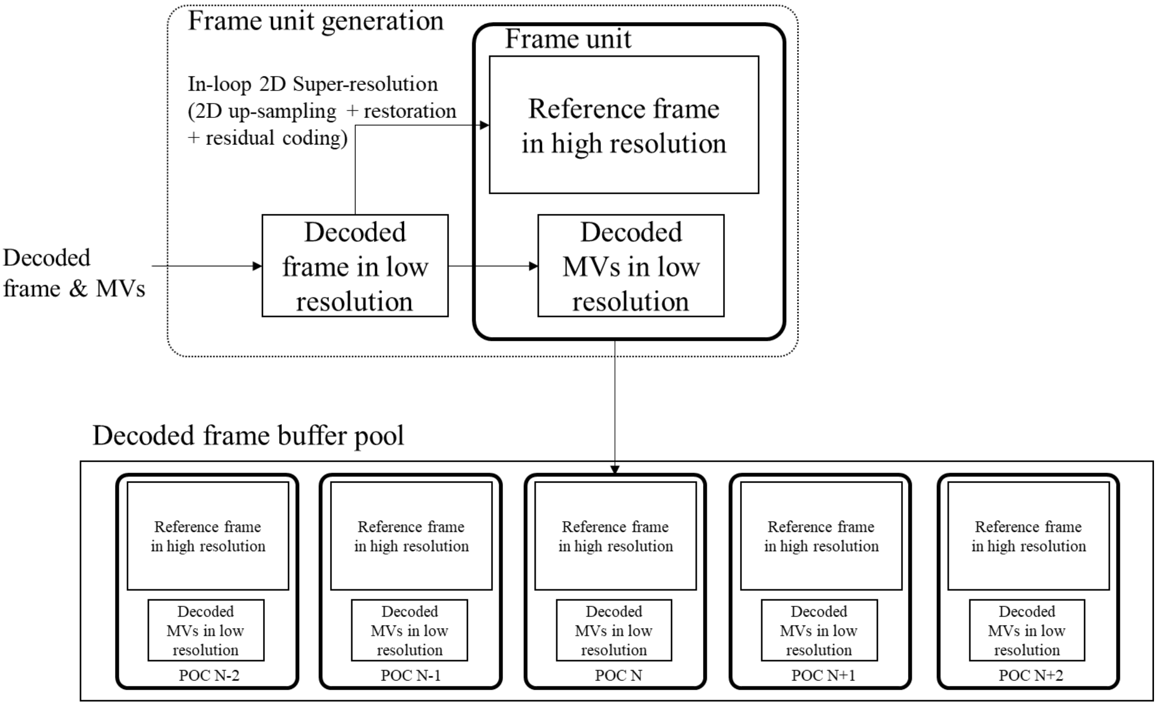

Figure 3. The decoded picture buffer pool with motion vector storage.

3. EXPERIMENTAL RESULTS

The proposed super-resolution approach was implemented based on the reference software for developing the extensions of AV1 and beyond [6], which has been provided by Alliance for Open Media (AOM). The implementation has been tested using all-intra (AI) and random-access (RA) common test conditions [7]. In these conditions, test sequences are divided into multiple categories based on different use cases. We focus our efforts on camera-captured standard dynamic range (SDR) test sequences with resolutions class A1_4K (3840×2160), class A2_2K (1920×1080), class A3_720p (1280×720), class A4_360p (640×360), class A5_270p (480×270), synthesized test sequences class B1_SYN (1920×1080) and high dynamic rage (HDR) test sequences with resolutions class HDR1_4K (3840×2160) and class HDR2_2K (1920×1080).

The proposed approach requires to find the best scaling factor and quantization parameter for each rate point. In the experiments, we have used a full search for determining a scaling factor and residual QP parameter among pre-selected candidates by performing multi-pass coding in each sequence.

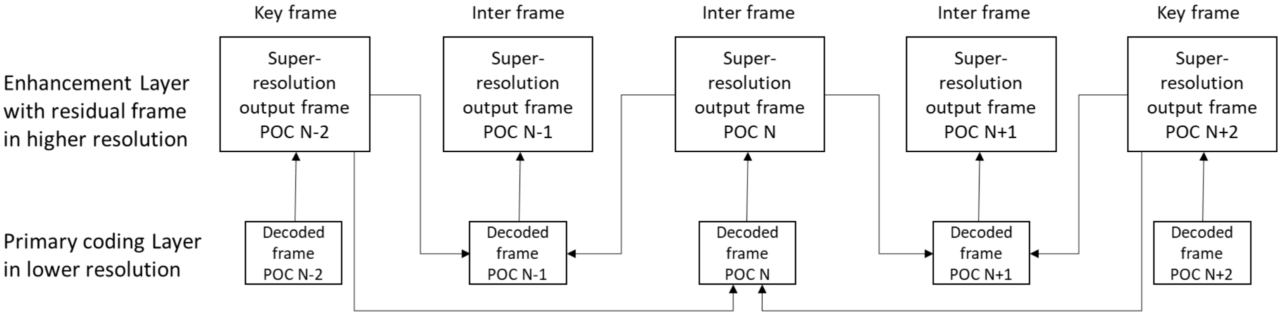

Figure 4. The prediction structure with residual frames.

Table 1 and Table 2 summarize the coding performance of using two-dimensional super-resolution with a maximum scale factor of six for SDR and HDR sequences under the AOM CTC Random Access (RA) configuration. We observe an overall luma BD-rate gain of 2.32% and 3.17% for SDR and HDR sequences, respectively. The gains in chroma are higher, and an estimate of the total gain is provided with the YUV BD-rate estimate that shows a gain of 2.49% and 4.22%, respectively. As can be seen in the Tables, the bit-rate savings are sensitive to the video sequence resolution. We observe YUV BD-rate improvements of 6.33% and 4.58% for the UHD (Class AI_4k) SDR and HDR, respectively. This is impactful in practice, as higher resolution sequences typically have more spatially correlated pixel data, which can be effectively reduced by downsampling. Table 3 and Table 4 provide the coding performance of the proposed method for the All Intra (AI) configuration. The overall test results for the AI configuration have the same tendency as the RA test results.

Table 1. SDR test results of the proposed super-resolution video coding vs. the AVM research ver.3 for RA CTC.

	PSNR				
Summary	**Y**	**U**	**V**	**YUV**	**wAverage**
Class A1_4K	-6.08%	-8.57%	-8.41%	-6.33%	-6.27%
Class A2_2K	-2.88%	-4.16%	-3.38%	-2.99%	-2.95%
Class A3_720p	-0.10%	-0.85%	-0.48%	-0.15%	-0.14%
Class A4_360p	-0.28%	-3.73%	-4.92%	-0.63%	-0.61%
Class A5_270p	-0.00%	-0.00%	-0.00%	-0.00%	-0.00%
Class B1_SYN	-2.01%	-5.27%	-3.69%	-2.32%	-2.21%
Average	-2.32%	-4.17%	-3.66%	-2.49%	-2.45%

Table 2. HDR test results of the proposed super-resolution video coding vs. the AVM research ver.3 for RA CTC.

	PSNR				
Summary	**Y**	**U**	**V**	**YUV**	**wAverage**
Class HDR1_4K	-4.05%	-11.38%	-24.41%	-5.57%	-5.15%
Class HDR2_2K	-2.67%	-8.66%	-8.88%	-3.45%	-3.16%
Average	-3.17%	-9.65%	-14.53%	-4.22%	-3.89%

Table 3. SDR test results of the proposed super-resolution video coding vs. the AVM research ver.3 for AI CTC.

	PSNR				
Summary	**Y**	**U**	**V**	**YUV**	**wAverage**
Class A1_4K	-3.61%	-10.41%	-11.03%	-4.58%	-4.18%
Class A2_2K	-1.76%	-5.10%	-5.76%	-2.18%	-2.05%
Class A3_720p	-0.25%	-1.65%	-1.71%	-0.41%	-0.36%
Class A4_360p	-0.41%	-3.79%	-5.49%	-0.74%	-0.74%

Class A5_270p	-0.08%	-1.93%	-0.83%	-0.22%	-0.18%
Class B1_SYN	-1.73%	-6.46%	-6.62%	-2.29%	-2.12%
Average	-1.55%	-5.24%	-5.71%	-2.00%	-1.86%

Table 4. HDR test results of the proposed super-resolution video coding vs. the AVM research ver.3 for AI CTC.

	PSNR				
Summary	**Y**	**U**	**V**	**YUV**	**wAverage**
Class HDR1_4K	-5.48%	-17.49%	-33.95%	-8.23%	-7.10%
Class HDR2_2K	-2.48%	-13.91%	-12.39%	-3.80%	-3.34%
Average	-3.57%	-15.21%	-20.23%	-5.41%	-4.70%

4. CONCLUSION

In this paper, we have presented a super-resolution video coding process that addresses a limitation of the AV1 standard and incorporates the benefits of two-dimensional super-resolution with larger scaling factors. Our proposed approach extends the AV1 design to enable higher-quality super-resolution processing and incorporates additional residual coding for improved visual fidelity. Through extensive experiments, we have demonstrated the effectiveness of our approach in enhancing coding efficiency. The results show the efficacy of the approach, including a 6.3% and 5.6% gain in coding efficiency for UHD SDR and UHD HDR content, respectively. These improvements highlight the potential of our approach in achieving higher visual quality while maintaining efficient compression.

REFERENCES

[1] Joshi, U., Mukherjee, D., Chen, Y., Parker, S. and Grange, A., “In-loop frame super-resolution in AV1,” Proc. PCS, 4704-4708 (2019).
[2] Khani, M., Sivaraman, V. and Alizadeh, M., “Efficient video compression via content-adaptive super-resolution,” Proc. ICCV, 4521-4530 (2021).
[3] Lin, H., He, X., Qing, L., Teng, Q. and Yang, S., "Improved low-bitrate hevc video coding using deep learning based super-resolution and adaptive block patching," IEEE Transactions on Multimedia 21(12), 3010-3023 (2019).
[4] Wang, R.-J., Huang, C.-W. and Chang, P.-C., "Adaptive downsampling video coding with spatially scalable rate-distortion modeling," IEEE Trans. Circuits Syst. Video Technol., 24(11), 1957-1968 (2014).
[5] Turkowski, K. and Gabriel, S., “Filters for common resampling tasks,” Graphics Gems /. Academic, 147-165 (1990).
[6] AOM Video Model, Online: https://gitlab.com/AOMediaCodec/avm
[7] Zhao, X., et al., “AV2 Common Test Conditions and Performance Measurement,” Testing Subgroup of AOMedia Codec Work Group.