White Paper

Media & Entertainment
AI Media Enhancement

AWS Compute Video Super-Resolution Powered by the Intel® Library for Video Super Resolution

Authors:

Carlos Salazar, AWS

Arturo Velasco, AWS

Onur Barut, Intel Corp

Christopher A. Bird, Intel Corp

Anand V. Bodas, Intel Corp

Xiaoxia Liang, Intel Corp

Surbhi Madan, Intel Corp

Tahani Trigui, Intel Corp

Abstract

Video Super Resolution (VSR) has become an essential part in the dynamic landscape of video enhancement workloads, with streaming, broadcast and user-generated content (UGC) applications. To date, VSR has been implemented using deep-learning AI algorithms requiring GPU architectures. This paper describes the collaboration between Intel and AWS to deploy VSR on CPUs taking advantage of the acceleration feature in Intel® Xeon® processors available on AWS. This paper provides an overview of the use cases and algorithms for VSR, how Intel works to make VSR available in open-source software for Intel processors, and how to implement VSR using pre-built pipelines from AWS.

Introduction

The rise of free ad-supported streaming TV (FAST) services has surfaced a comprehensive range of personalized content, including old-fashioned movies. However, most of this content is only available in lower-resolution formats (SD) and needs a better viewing experience on a home TV screen. Traditionally, low-complexity methods such as Lanczos[3] and bicubic[6] have been used for spatial upscaling by 2x and 4x. However, these techniques often introduce image artifacts such as blurring and pixelation. On the other hand, deep learning (DL) approaches such as SRCNN[2] and Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR)[7] have shown remarkable results for super- resolution, overperforming classical algorithms in objective quality metrics (VMAF, SSIM, PSNR). Nevertheless, DL super-resolution methods are computationally expensive, which makes them unsuitable for low-cost channels, the most typical case in FAST services. Therefore, we propose a cost-effective architecture for video super-resolution that leverages the benefits of Amazon Spot instance to process video assets following a best-effort strategy. In our approach, mezzanine files are split into smaller pieces (2 to 10 minutes) and independently processed by a super-resolution pipeline using Intel® Advanced Vector Extensions 512 (Intel® AVX-512) and leveraging the Intel® Library for Video.

The Intel® Library for Video Super Resolution (Intel® Library for VSR)[2] is an open-source framework that implements machine learning based video super-resolution, which is trained on a diverse dataset to meet performance and quality for real world scenarios. For the evaluation phase, we use selected Alliance for Open Media (AOM) videos to assess the objective visual quality in terms of state-of-the-art referenced and non-referenced quality metrics, such as VMAF, SSIM, and RAPIQUE.[10] Our approach reported considerable benefits against classical interpolation methods while providing a manageable computational cost utilizing a CPU architecture.

The Intel® Library for Video Super Resolution

The Intel Library for Video Super Resolution harnesses the power of upscaling videos to a higher resolution, to enhance visual experience while meeting industry standards and resolution of end devices in a cost-effective manner. It includes an enhanced implementation of the Rapid and Accurate Image Super Resolution algorithm (RAISR),[9] to meet real-world customer needs and deployment. This implementation takes advantage of machine learning techniques to produce high quality versions of low-resolution video content without adding artifacts, a common hallucination issue caused by artificial intelligence generating unnatural looking details. To apply the algorithm to large-scale industry use, especially in broadcasting and streaming applications, the enhanced RAISR implementation is trained on a highly diverse data set, enables two-step filtering, supports 8-bit and 10-bit processing and is optimized for native CPU instruction sets.

The algorithm is implemented in C++ and the filters are trained with diverse content types encompassing various scenes ranging from natural scenes, animation, action, sports, sitcoms, news conferences and more. The dataset contains pairs of low-resolution and high-resolution images generated by multiple downscaling and degradation techniques, to make it suitable for upscaling real-world content. The implementation supports scaling by factors of 1.5x or 2x, enabling commonly used upscales like 480p to 720p, 540p to 1080p, 1080 to 4K, or all the way up to 8K for both 8-bit and 10-bit content. The two-step filtering option enables the use of an additional pre or post upscaling filter, which could be used to control sharpness or denoising.

Figure 1 shows a high-level implementation overview of the Intel Library for VSR, available in open source under the Open Visual Cloud project. It is available as a plugin to FFmpeg, the most common framework used in the development of multimedia applications to ease development. The Intel Library for VSR has implemented multiple filters based on the content characteristic used in the training, called filters_lowres, filters_highres, filters_denoise. As the name suggests, each filter is trained with a specific type of data to address specific needs. For example, filter_denoise is trained on the video content where different types of noises are present to denoise and upscale the low-resolution input using the two-pass option. The training framework for the Intel Library for VSR[3] enables easy retraining of these filters on custom content. The implementation scales efficiently with the available processing cores, utilizing an x86 instruction set including Intel AVX-512 for the Intel® Xeon® Scalable processor family or the Xe-core for Intel® GPUs. These optimizations enable real-time processing of multiple video streams on Intel architecture. The implementation also supports a quantized FP16 datatype to further boost performance on the latest generations of Intel Xeon Scalable processors[5] and Intel® Data Center GPUs.[4]

AWS implementation

As mentioned previously, implementing the enhanced RAISR algorithm utilizing Intel AVX-512 requires AWS-specific instance types, such as c5.2xlarge, c6i.2xlarge, and c7i.2xlarge. We leverage AWS Batch[4] to compute jobs and automate the entire pipeline rather than dealing with all the underlying infrastructure, including start and stop instances. We also automate the ingress and egress workflow to trigger each job based on an S3 bucket event. Therefore, AWS customers interested in using the benefits of the enhanced RAISR algorithm for super-resolution can continue focusing on the ABR transcoding pipeline and adapt their existing workflow to leverage AWS Batch as a preprocessing stage.

The first step is to create a compute environment in AWS Batch, where CPU requirements are defined, including the type of EC2 instance allowed. The second step regards creating a job queue associated with the proper computing environment. Each job submitted in this queue will be executed using the specific EC2 instances.

The third step involves the definition of a job. At this point, it is necessary to have a container registered in the AWS Elastic Container Register (ECR[5]). Building the container is further detailed in AWS-VSR GitHub repository.[6] The container includes installing the Intel Library for VSR, open-source FFmpeg tool, and AWS CLI to perform API calls to S3 buckets. Once the job is properly defined (image registered in ECR), the job can start being submitted into the queue. This process is automated using a lambda function that receives a trigger from an S3 bucket (source), performs some validation (video codec, resolution, etc.), and sends an API call to AWS-Batch to start the super-resolution job. Once the VSR process has finished,

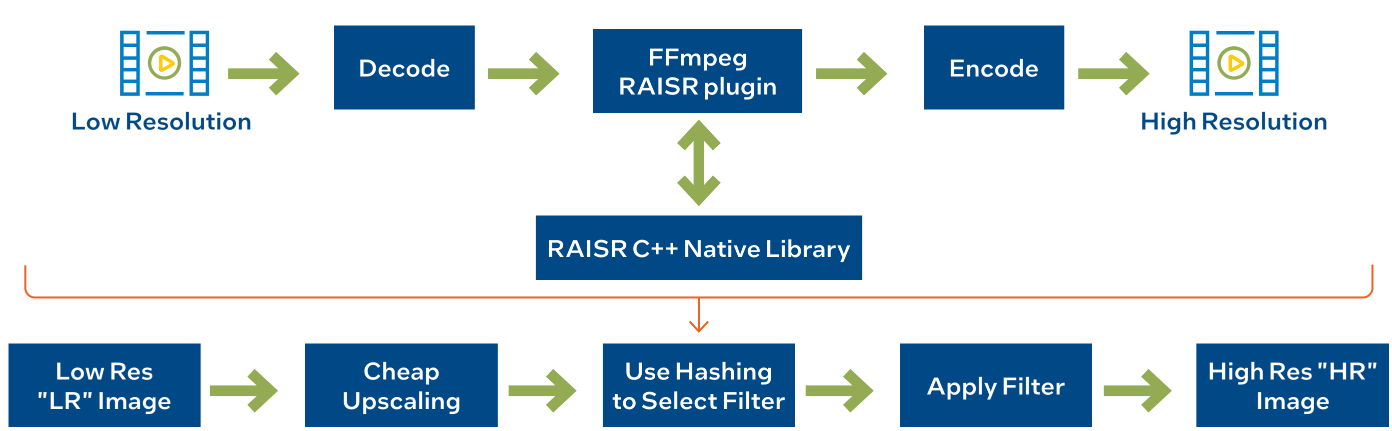

Figure 1. Overview of the Intel® Library for Video Super Resolution.

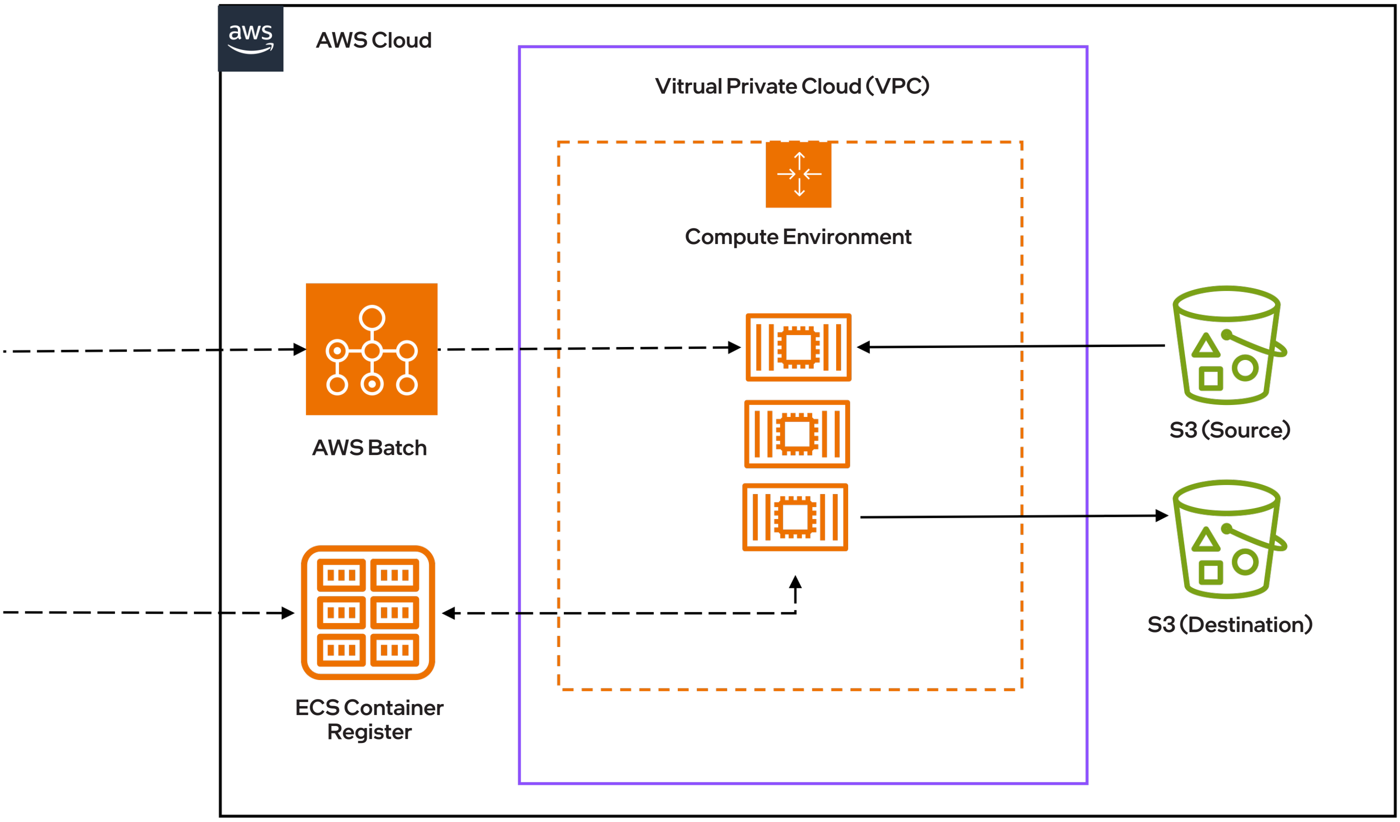

Figure 2. AWS proposed architecture.

the resulting high-dimensional video is uploaded to a destination S3 bucket, becoming the new mezzanine file for the following ABR transcoding stages, which does not depend on an encoder vendor. After that, the job concluded, the underlying infrastructure automatically deployed during the processing is immediately deleted. As a variant, during the computing environment step, the option of spot-instance can be selected, saving up to 90% of the total cost. Therefore, the jobs in the queue will wait until a spot EC2 instance with the specific types is available. Figure 2 describes the general architecture.

Performance and results

Intel evaluation

To verify the effectiveness of the enhanced RAISR implementation, Intel did both objective and subjective visual quality (VQ) assessments against a traditional upscaling approach (Lanczos). 21 video sequences of different resolutions were selected from the Pexels dataset[8] to gather objective quality outcomes. These original videos serve as a ground truth for the objective evaluation. The selection for VQ metrics was made based on industry-standard techniques, including peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and video multimethod assessment fusion (VMAF). The methodology involved creating input videos from the ground truth by downscaling using bilinear interpolation. These low-resolution inputs were then upscaled using Lanczos and the enhanced RAISR implementation and compared with ground truth to measure VQ metrics.

Table 1. Objective quality comparison with different video resolutions.

Lanczos				Enhanced RAISR		
Input Res	PSNR	SSIM	VMAF	PSNR	SSIM	VMAF
sub-720p	31.013	0.897	64.967	31.343	0.902	75.197
sub-1080p	35.759	0.945	63.810	36.319	0.950	75.139
1080p	37.863	0.946	71.865	38.134	0.948	80.162

The quantitative comparison is categorized based on input resolution as shown in Table 1. The results illustrate that all VQ metrics of enhanced RAISR consistently outperform Lanczos. The implementation has shown significant improvements in the perceptual video quality indicator measured by VMAF while demonstrating less significant improvements in PSNR and SSIM.

Further, a subjective quality assessment was done on an extended list of sequences to include videos from the AOM dataset[1] and considerable improvements in the quality of the video were observed as well. Here the ground truth is used as input and upscaled using Lanczos and enhanced RAISR for comparison. As shown in Figure 3, the globe on a high-rise building along with the roof and the windows behind it show better contrast and are sharper on the edges. Figure 4 illustrates the visual improvements from a sports fans sequence where the writing on a T-shirt is visually clearer and sharper.

Table 2. Objective quality comparison with different video resolutions (RQ=RAPIQUE).

	LR 540p	Lanczos 1080p		Enhanced RAISR 1080p	
Sequence	RQ	RQ	VMAF	RQ	VMAF
Motorcycle	0.6530	0.6751	56.3497	0.6861	63.6224
MountainBike	0.6986	0.7287	64.2387	0.7402	70.5487
Skater227	0.4603	0.4872	65.8142	0.4980	69.1595
TreesAndGrass	0.6260	0.6362	64.4056	0.6460	67.9418
WorldCup_far	0.6269	0.6366	67.9202	0.6469	75.7239

AWS evaluation

As described in the introduction, one of the critical drivers of this implementation is running FAST channels, where, in many cases (old movies and series), the original film was produced in SD resolution. Consequently, the target high- resolution video does not exist. In that case, well-known referenced quality metrics such as VMAF and SSIM are unsuitable. This fact motivates the execution of a new set of tests — referenced (VMAF) and non-referenced (RAPIQUE) state-of-the-art quality metrics. Table 2 presents a selected set of sequences from the AOM dataset. For each of those videos, a linear downscaled version is performed using ffmpeg parameter CRF=30, and then this new file is used as input for the Lanczos and Enhanced RAISR algorithm. In this case, we can use VMAF because our environment is controlled (existence of reference high-resolution source). However, our objective is to demonstrate the relation between VMAF and RAPIQUE features, providing the support to use RAPIQUE as a mechanism for quality control in production environments where the reference high-resolution video does not exist. In this case, the RAPIQUE score determines if a video has been improved or not. Figure 5 provides an example of a subjective visual improvement and the relation to RAPIQUE results.

Figure 3. Side by side view of the output of Lanczos (left) and enhanced RAISR (right) upscale of a building.

Figure 4. Side by side view of the output of Lanczos (left) and Enhanced RAISR (right) upscale of a T-shirt.

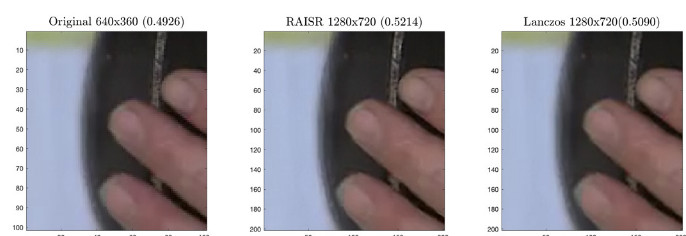

Figure 5. Subjective evaluation and RAPIQUE features score.

Conclusion

The optimized implementation of enhanced RAISR offers an economically efficient means of upscaling videos while improving quality. This approach eliminates the risk of introducing undesirable artifacts into the video, hence facilitating automated enabling at scale for the upscaling pipeline. In the future, the same approach could be applied to video quality enhancement pipelines focusing on improving visual quality without upscaling. Such an enhancement may have a wider impact with a growing number of videos in streaming applications, often captured with low-quality cameras or transcoded several times, impacting visual quality severely.

More Information

For more information on how to use Intel Xeon CPUs for Video Super Resolution, visit github.com/OpenVisualCloud/Video-Super-Resolution-Library.

For more information on AWS services for Video Super Resolution, visit github.com/aws-samples/video-super-resolution-tool.

[1] AOM. 2023. AOM CTC. https://media.xiph.org/video/aomctc/.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," Computer Vision – ECCV 2014: 184-199.

[3] Claude E. Duchon, "Lanczos Filtering in One and Two Dimensions" Journal of Applied Meteorology and Climatology 18, 8 (1979), 1016-1022. https://doi.org/10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2.

[4] Intel. 2023. Intel Data Center GPU. https://www.intel.com/content/www/us/en/products/details/discrete-gpus/data-center-gpu/flex-series.html.

[5] Intel, "Intel Xeon Scalable processors," https://www.intel.com/ content/www/us/en/products/details/processors/xeon/scalable.html.

[6] R. Keys, "Cubic convolution interpolation for digital image processing," IEEE Transactions on Acoustics, Speech, and Signal Processing 29, 6 (1981), 1153-1160. https://doi.org/10.1109/TASSP.1981.1163711.

[7] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced Deep Residual Networks for Single Image Super- Resolution," CoRR abs/1707.02921 (2017), arXiv:1707.02921 https://arxiv.org/abs/1707.02921.

[8] Pexels. 2023. Royalty-free stock footage website. https://www.pexels.com.

[9] Yaniv Romano, John Isidoro, and Peyman Milanfar, "RAISR: Rapid and Accurate Image Super Resolution," CoRR abs/1606.01299 (2016), arXiv:1606.01299 https://arxiv.org/abs/1606.01299.

[10] Zhengzhong Tu, Xiangxu Yu, Yilin Wang, Neil Birkbeck, Balu Adsumilli, and Alan C. Bovik, "RAPIQUE: Rapid and Accurate Video Quality Prediction of User Generated Content," IEEE Open Journal of Signal Processing 2 (2021), 425-440. https://doi.org/10.1109/OJSP.2021.3090333.

0524/DL/MESH/PDF 358259-002US