Video Quality for Virtualized Environments

Introduction

The video industry is undergoing an unprecedented amount of change. More video content is being watched across more and more network-connected devices that are increasingly capable of supporting high-quality video. Video experiences are also increasingly becoming more immersive with the advent of 4K TV and virtual reality technology. The Cisco Visual Networking Index™(Cisco VNI™) Forecast 2015-2020 predicts that by 2020, 82 percent of IP traffic by will be video, of which 50 percent will be high definition (HD) or Ultra HD (UHD). The forecast further expects that mobile devices will drive most of the growth, making up 43 percent of IP traffic.

Across these changes, video providers are presented with the opportunity to improve picture quality by using new tools, including emerging technologies that support the High Efficiency Video Coding (HEVC) standard, UHD resolution delivery and high dynamic range (HDR) video. The emergence and adoption of cloud architectures and the virtualization of video processing technologies are key to enabling the flexibility and scalability required to efficiently address these opportunities. Yet, video providers are confronted by the technological challenge of implementing these new tools and technologies across an increasingly software-based infrastructure while maintaining the requisite levels of reliability to successfully deliver high-quality video.

This paper addresses how video providers can pursue video quality enhancements in an increasingly virtualized environment.

Evolution of Video Codecs

The need for superior video compression technology is deemed inevitable as network bandwidth consumption is projected to accelerate with the exponential growth of internet video traffic. This growth is driven by the growing shift in video delivery and consumption through adaptive bitrate-streaming-application clients running on the increasing number of network-connected devices. The network bandwidth crunch is further exacerbated by consumers' affinity for higher video quality, born from the near-ubiquitous presence of broadcast quality HD experiences. Consumers also increasingly expect improvements in video-quality experiences when upgrading from an older to a newer version of the same type of consumer-electronics (CE) device, such as mobile device or a 4K TV.

HEVC is the latest video coding specification jointly developed by ITU-T Video Coding Expert Group (VCEG) and International Standards Organization (ISO)/IEC Moving Picture Experts Group (MPEG). It features a comprehensive suite of coding tools that double coding performance (defined as quality per bits) over the previous video coding standard deemed most efficient, Advanced Video Coding (AVC).

The AVC codec (also known as MPEG-4, Part 10 or H.264) was first published in 2003. AVC achieved rapid and widespread adoption in PC and mobile video applications. The 2005 revision of the AVC specification included the tools required to properly support HD video in broadcast applications. The latest of successive generations of hardware-based AVC technology provides a 40-70 percent reduction in bit rates over MPEG-2 video. AVC was timely for its adoption in IPTV applications where operators new to the service provider market were not faced with an installed base of MPEG-2 decoders. More recently, cable system operators have begun the process of migrating their MPEG-2 video broadcasts to AVC as their deployed populations of set-top boxes have become capable of decoding AVC.

HEVC is likely to see a similar adoption pattern, first in new applications targeting devices that are rapidly refreshed and then coming to legacy applications with their slowly evolving devices. Indeed, as mobile phones, tablets, TVs, laptop computers, and other CE devices evolve, they will feature the benefits of technology advancements for enhanced connectivity and improved media processing capabilities. These connected devices are capable of decoding HEVC-coded content, and thus, extend the benefits of HEVC's superior coding performance and picture quality.

The computation and processing resources required for decoding an HEVC bit stream exceed those required for decoding a comparable AVC bit stream. Despite that requirement, some CE devices exhibit inferior run-time performance when decoding an AVC bit stream that requires double the bit rate of an HEVC bit stream of comparable video quality. This deficiency is because both AVC and HEVC employ context-adaptive binary arithmetic coding (CABAC) as the entropy-coding tool in video streaming and broadcast applications. CABAC processing requires sequential execution and does not exploit any of the parallelization mechanisms in software-based video processing platforms (more on software-based platforms follows). Consequently, CABAC processing is typically one of the main bottlenecks in these CE device implementations. The extent of the required sequential execution tends to be proportional to the bit rate of the coded video bit stream. Thus, in some cases it is preferable to process an HEVC bit stream with half the bit rate of an AVC bit stream of comparable video quality. (CABAC design in HEVC was also simplified to alleviate the decoder-processing burden.)

Evolution of Video Quality in Software-Based Video Processing Platforms

The pursuit of video quality is a constant search to increase the video information density per unit of bandwidth consumed. The evolution from hardware- to software-based video encoding platforms opens the door to a continuous array of new possibilities by adding dimensions such as scalable video processing, virtualization, and microservices. These new dimensions allow more flexibility to optimize video quality for video services.

As CPU capabilities evolve, new sets of instruments become available to implement complex algorithms that can increase information density for video processing. As a result, investments in developing better compression algorithms and next-generation video codes, built on a standardized stack of microservices, are supported by investments in optimizing CPU usage to implement those algorithms.

Software-based and virtualized video platforms extend the desired flexibility and scalability for executing multiple video processing operations in concert, such as video decoding, video encoding, and video transcoding. These virtualized platforms are typically based on architectures encompassing multiple interconnected physical processing devices, each having a number of cores and multilevel caches that support data sharing, data transfers, and data coherency. The multiple interconnected physical processing devices extend the type of high-level–parallelism mechanism, which is useful to achieve density of video-processing operations. (Software-based video processing also supports more agile product development models, which result in shorter product cycles and facilitate the demand to offer new video-quality-enhanced products to the market.)

The flexibility and scalability of software-based video processing platforms provide the foundation for a system architecture to satisfy growing demand for video quality. These new architectures present opportunities for video providers but consequentially mandate a higher level of resource orchestration and platform reliability to effect high-quality video delivery.

Orchestrating Video Quality in Software-Based Video Processing Environments

Today's video workflows operate with a multitude of source formats (codecs, streams, files, etc.), video processing functions (decode, encode, transcode, multiplex, encrypt, etc.) and output formats (live linear, time-shifted, video on demand [VoD], over the top [OTT]). The shift towards virtualizing these video workflows, where video processing applications—once tied to dedicated appliances—are now available as virtual video functions, enables these functions to be automatically configured and wired together based on templates to support any type of workflow. Given the complexity of today's video infrastructure and operating environments, this shift to virtualization introduces a more agile way to configure, manage, and operate video workflows. However, to configure and operate at scale, a future-oriented video processing orchestrator is required to ensure that the video operator can rapidly enable these multisource to multi-output capabilities.

The role of such an orchestrator is to provide the means to configure video processing templates. This role enables intimate aspects of video processing functions to be configured to the optimal settings to achieve the highest possible video quality, given the limitations associated with target outputs, such as bandwidth constraints, client device limitations, and other aspects. A key element of this configurability is that it needs to be agnostic to the video sources and the underlying infrastructure. In other words, it should **not** be bound to the specific video source or the hardware/software components being used to process the video. The templates enable video-savvy users to preconfigure the optimal set of parameters for the highest video quality possible, which are then applied to each video asset in the workflow.

A second role of the orchestrator is to manage and control the hardware devices and software applications that implement the various video processing functions in a workflow. This includes the full lifecycle of deploying, scaling out, scaling in, and upgrading video processing functionality.

Ultimately, the crucial role for the video processing orchestrator is to automatically manage the capacity needs to execute video processing templates. The video processing orchestrator must be able to assess how much resource of what type are needed (software components and number of nodes) to implement various video processing functions and also to manage overhead capacity to ensure system resilience (such as failure and maintenance). In today's high-performance and highly complex video systems, automatic resource assessment and calculation is critical for maintaining video quality.

Defining and Assessing Video Quality

The goal of video-quality assessment is to have an understanding of the quality degradation video experiences when going through a video processing system (for example, HEVC encoder). The result of the assessment can be used to compare different processing algorithms, use cases, or implementations.

Video quality can be assessed in two ways:

I. Subjective video-quality assessment means that a human observer is watching the video (either in real time or at reduced speed) and assigns a score to the video based on the observed quality. In many cases a human observer compares a reference video with a processed video side-by-side on two identical screens. By adding a delay between the reference and the processed signal, the observer can look for impairments in the processed video. When an impairment is detected, the observer has time to see if this impairment is also in the reference system.

II. Objective video-quality assessment makes uses of objective video-quality metrics that are calculated by a computer or measurement system. Some of these metrics can be calculated in real time, while others are very processing intensive and need to be calculated off line.

Objective video-quality metrics can be classified in three major categories:

- Full-reference metrics need access to both the source signal and the output of the video processing system that is being evaluated (for example, the serial digital interface (SDI) source and the HEVC output of an HEVC encoder). Full reference metrics directly compare the processed signal with the source signal to arrive at a quality number. Well-known examples of full-reference metrics are Peak Signal to Noise Ratio (PSNR) and Multi-Scale Structural Similarity (MS-SSIM).
- Reduced-reference metrics still need to access both the source signal and the output of the video processing system under evaluation but only need a reduced part of those signals (for example, a spatially downscaled version of it). This metric can be helpful in situations where the transmission bandwidth towards the measurement system is limited and doesn't permit the complete source and processed signals to be received (for example, a 3 Gbps uncompressed source signal and a 10 Mbps connection to the measurement system).
- No-reference metrics only need the processed output signal to calculate a quality number for the processed video. From a practical point of view, a no-reference metric is the easiest to deploy and can be applied at any point in the video processing chain (for example, at a satellite downlink where the input to the video processing system is not available or a VoD stream received over the Internet).

Evaluating Objective and Subjective Metrics

Subjective video-quality assessments are still the best approach but have a number of disadvantages:

- Subjective video quality assessment is very time consuming. To reach a meaningful conclusion, you must look at many different sources with different content (for example, news, sports, movie content, etc.) and assess the video quality for different operating points and use cases of the video processing system.
- A trained and experienced observer may be needed (so-called "golden eyes") to detect all impairments.
- It's possible that even experienced video-quality specialists might not discover minor video-quality impairments (that is, those impairments that are introduced in a new video-quality algorithm).

Objective video quality assessments have the advantage over subjective video quality assessments, because they can be automated and don't require the expensive time of video-quality specialists (although they sometimes require a significant amount of machine time). Objective video-quality assessments are also very efficient at finding minor video quality differences between consecutive software versions of a certain video processing device and are therefore heavily used for automated video quality regression testing.

One aspect of using objective video-quality metrics is that they don't always agree with each other. This can mimic the real world in that humans also do not always agree on subjective video quality. One approach to address this issue is to combine multiple objective video-quality metrics to arrive at an objective video-quality assessment conclusion. This would mimic a real-world situation where an average or median of multiple human assessments is used.

Using multiple metrics has the advantage that in case not all metrics agree on the video-quality result, you can still derive a meaningful conclusion (for example, if three out of four metrics assign a higher video-quality score to system A than to system B, then system A is most probably better than B).

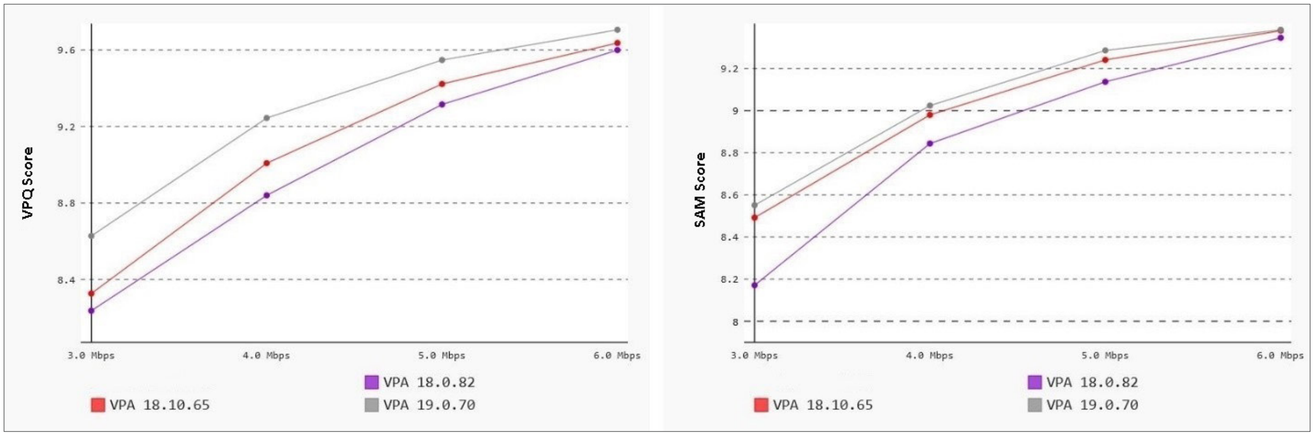
Evolving Approach to Video Quality Assessment at Cisco

At Cisco we use a number of different objective video-quality metrics for this combined approach, including PSNR, MS-SSIM, Scientific Atlanta Metric (SAM) and Video Picture Quality (VPQ). SAM is a full-reference metric developed internally at Cisco and has been used successfully for the past 10 years. VPQ is a new, patent-pending Cisco developed no-reference metric that shows good correlations with subjective video quality assessment.

The Cisco VPQ metric measures video quality by inspecting the coding decisions of the encoder at the codec block levels and assesses those decisions given the characteristics of the video. VPQ performs both pixel-level and codec-specific analysis of the encoded stream. VPQ supports all major broadcast codec formats, including MPEG-2 Video, AVC, and HEVC.

We developed a completely automated objective test system for video quality assessment and video quality regression testing. The test system automatically streams video sources to the video processing equipment under test. The latter is automatically configured by the test system for a specific use case (for example, AVC HD encoding, AVC HD to MPEG-2 SD transcoding, HEVC UHD encoding) and operating parameter (bit rate, GOP structure, latency, density). The processed output of the equipment under test is automatically captured by the test system. The output captures are then analyzed and a number of different objective video-quality metrics are calculated. The results are automatically visualized (see Figure 1 that follows) in different formats (for example, rate distortion curves), which makes it easy to review and compare with previous results.

Figure 1. Comparison of Cisco Objective Video-Quality Metrics



Horizontal axis is encoder output bit rate. Vertical axis is video-quality metric score. Higher values are better video quality. Line colors represent software version for video processing.

The number of databases that include human subjective judgements continues to grow and includes databases such as University of Texas LIVE, Oklahoma State CSIQ, Shanghai Jiao Tong University (SJTU), and IT-IST Lisbon. These databases include human scores for various distortion types representative of content coded with AVC, MPEG-2 Video, and HEVC at various bit rates, facilitating the validation of metrics as they become available.

The performance of SAM and VPQ against the University of Texas LIVE database is shown in Table 1 that follows. What is shown in the table is the Spearman Rank Order Correlation Coefficient (SROCC) of the metrics for the AVC and MPEG-2 Video portions of the LIVE database. An SROCC of unity would indicate perfect correlation. It can be seen that PSNR performs poorly compared to newer metrics, such as SAM and VPQ.

Table 1. Spearman Rank Order Correlation Coefficient (SROCC) for Video Quality Metric on University of Texas LIVE Database

Model	Year	AVC	MPEG-2	Source	Full Reference	Complexity
PSNR	-	0.43	0.36	n/a	Yes	Lowest
SSIM	2004	0.65	0.55	University of Texas	Yes	Low-to-medium
VSNR	2007	0.65	0.59	Oklahoma St University	Yes	Low
SR-SIM	2012	0.64	0.68	Tongji Univ. China	Yes	Low
MS-SIM (DMOS)	2003	0.71	0.66	University of Texas (Video Clarity)	Yes	Medium-To-High
VQM	2004	0.65	0.78	NTIA	Yes	High
SAM	**2005**	**0.72**	**0.74**	**Cisco**	**Yes**	**Medium-To-High**
VPQ	**2014**	**0.74**	**0.76**	**Cisco**	**No**	**Very Low**
MOVIE	2009	0.77	0.77	University of Texas	Yes	Very High
ST-MAD	2012	0.91	0.84	Oklahoma St University	Yes	Very High

Conclusion: Video Quality in Software-based and Virtualized Environments

Video providers are embarking on a journey, replacing existing hardware-based and dedicated infrastructure with software-based virtualized infrastructure. Paramount to this infrastructure advancement is the mandate to provide the end user with the best possible video quality—and to deliver that outcome while optimizing available bandwidth.

One of the advantages of virtualized environments is the ability to automatically configure complex workflows. One such example is the combination of video processing templates to define video quality levels for target applications across available hardware and software resources. These templates provide an intuitive and simplified means to configure workflows. In virtualized environments, a video processing orchestrator is aware of the resources needed to execute these templates and is able to access those resources to produce the video quality targets required by the workflow.

The virtualization of video processing and video workflow orchestration are key to meeting the mandate for video quality improvements by simplifying the support for tools and technologies, such as HEVC, UHD, and HDR, while simplifying the application of video quality in virtualized templates and workflows.

C11-737936-00 10/16