Quality-of-Experience Monitoring, Optimization and Management: A Unified End-to-End Solution

A Technical Paper prepared for SCTE•ISBE by

Zhou Wang, PhD, FIEEE
Chief Science Officer, SSIMWAVE Inc.
Professor, University of Waterloo
zhou.wang@uwaterloo.ca

Abdul Rehman, PhD
Chief Executive Officer
SSIMWAVE Inc.
abdul.rehman@ssimwave.com

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

Abstract

Traditional quality assurance methods for large-scale video distribution networks operate independently at different points along the video delivery chain, reporting partial and incoherent measurements, leading to poor and fragmented understanding about how multiple stages of quality degradations affect the final quality-of-experience (QoE) of end users. We propose a framework that uses a unified end-to-end solution to produce consistent QoE scores at all points along the delivery chain under the same evaluation criterion. The novel solution produces a clear and complete picture instantaneously about how video QoE degrades over the network, allows immediate issue identification, localization and resolution, enables quality and resource optimization, and provides reliable predictive metrics for long-term strategic resource and infrastructure allocations. The main challenge in the implementation of the solution is to create a unified QoE metric that not only accurately predicts human perceptual QoE, but is also light-weight and versatile, readily plugged into multiple points in the video delivery chain. The QoE metric should produce real-time QoE scores across a wide range of bitrates, resolutions, frame rates and dynamic ranges, and combine presentation picture quality with the perceptual impact of video freezing and adaptive streaming events. We show that the SSIMPLUS metric offers the best promise to meet all the challenging demands.

Keywords

Quality-of-experience, video distribution system, video delivery chain, video quality assessment, video streaming, end-to-end quality assessment, video encoding, adaptive streaming

Introduction

There has been a remarkable growth of video distribution services in the past few years [1]. While common consumers are enjoying the video streams delivered to their TVs, smart phones and tablets, they often complain about the quality of the video they are experiencing [2]. Meanwhile, content producers are concerned about whether their creative intent is properly preserved during the video distribution process [3], [4]. Quality assurance (QA) is an essential component to warrant the service of video distribution systems. Traditionally, QA has been network-centric, focusing on the quality-of-service (QoS) [5] provided to the users, where the key metrics are determined by the network service level parameters such as bandwidth, package drop rate, and network delay. However, QoS metrics have fundamental problems in tracking what the users are actually experiencing. Recently, Quality-of-Experience (QoE) [6], which measures "the overall acceptability of an application or service as perceived subjectively by the end-user" [7], has been set to replace the role of QoS. In practice, the actual meaning of "QoE" measurement could vary significantly from one solution to another. For example, simple device playback behaviors such as statistics on the duration and frequency of video freezing events, may be employed to create a crude estimate of visual QoE. Such simple measures only provide a rough idea about how certain components of the video delivery system perform, but are distance away from what we really need in terms of accuracy, comprehensiveness and versatility. Moreover, the perceptual artifacts that affect picture quality are not properly measured, and the large perceptual differences due to viewing conditions are not properly taken into consideration. Consequently, they are at best "pseudo-QoE measures" or "QoS measures at the client", and are difficult to be used to localize quality problems, to optimize system performance, and to manage the visual QoE of individual users.

We propose a unified end-to-end framework for QoE monitoring, optimization and management. The general philosophy is to align all measurements with the visual QoE of end users. Keeping this in mind, any design and resource allocation in the video distribution system, regardless of if it is for the whole system or for any individual component at the head-end, media data center, network, access server, or user device, should be evaluated, compared and optimized for one criterion, i.e., the impact on end users' QoE. To make such a system work properly, the most challenging task is to find a highly accurate, efficient and versatile QoE metric. Such a QoE metric, deployed throughout the video distribution system, establishes the basis for unified QoE monitoring, optimization and management.

Content

1. End-to-End Visual QoE Monitoring, Optimization and Management

Figure 1 illustrates a general framework of modern video distribution systems. When the source video content is received, it passes through a sophisticated video delivery chain consisting of many processing, encoding, transcoding, packaging, routing, streaming, decoding, and rendering stages before it is presented on the screen of individual users' viewing devices. To ensure the video is faithfully and smoothly delivered to the consumer device, the ideal quality assurance method would be to have human inspectors placed at all transition points along the chain, so that any quality issue can be identified instantaneously, and all measurements can be compared directly. In practice, however, this is infeasible because it requires thousands of source video streams and millions of derivative streams to be evaluated continuously by human inspectors, a non-scalable resource in the real-world. A viable solution is to replace humans with objective QoE monitoring probes, as illustrated in Fig. 1, which constantly predict human QoEs based on objective QoE metrics at the corresponding inspection spots.

There are two essential properties of such QoE monitoring probes. First, they should "see" and "behave" like human inspectors. More specifically, they should "perceive" all the actual pixels of all video frames like humans, and they should produce QoE scores just like what humans would say about the video quality when seeing the same video streams. Second, they should provide a "unified end-to-end" monitoring solution in the sense that the QoE evaluation methods at all transition points along the video delivery chain are designed under the same evaluation framework and compatible methodology to produce consistent quality scores that are directly comparable.

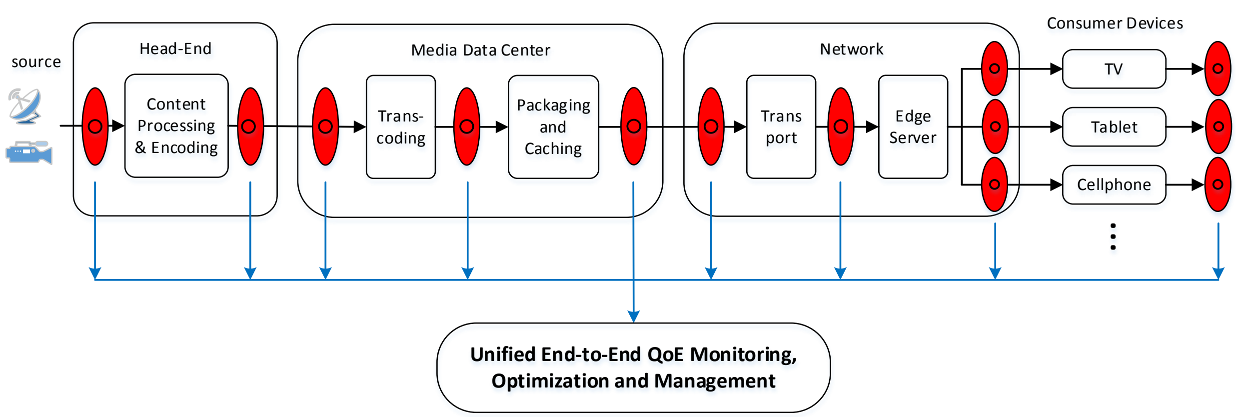

Figure 1. Unified end-to-end QoE monitoring, optimization and management framework in a video distribution system.

Once QoE monitoring probes are deployed throughout the video delivery chain, QoE data can be collected instantaneously and continuously. Subsequently, statistics can be computed at different time-scales (minutes, hours, days, weeks, months, years). These lead to many valuable benefits, as described in Fig. 2. More specifically,

- Operation engineers are able to gain immediate awareness about how video QoE degrades along the video delivery chain. As such, quality problems can be immediately identified, localized and resolved.
- Design engineers are able to closely observe the QoE variations between the input and output of individual components or the whole video delivery system as a whole. This helps them perform better design and optimization that target at improving and stabilizing the QoE of end users.

Managing executives are able to obtain a clear picture about how video quality evolves throughout the video distribution system and over long time scales. When long-time, large-scale data has been collected, big data analytics can be performed to help make intelligent strategic decisions on the operations of the system.

Figure 2 - Applications of unified end-to-end QoE monitoring, optimization and management system.

2. Objective QoE Metric

At the core of the end-to-end QoE monitoring framework is the QoE quality metric, which mimics human behaviors in evaluating video quality, and is the most challenging technical problem to solve. A good objective QoE metric combines deep understanding of the human visual system with advanced computational models and algorithms. It also requires smart design and efficient implementation of the algorithms and systems. Traditional approaches such as peak signal-to-noise-ratio (PSNR) have been shown to have poor correlations with perceptual video quality. More advanced perceptual video quality assessment (VQA) methods such as the structural similarity index (SSIM) [8], [9], multi-scale SSIM (MS-SSIM) [10], video quality model (VQM) [11] and video multi-method assessment fusion (VMAF) [12] improve upon PSNR but are still limited in prediction accuracy. More importantly, these traditional VQA approaches have fundamental limitations in their application scopes, functionalities and/or computational cost. These limitations largely impede them from being deployed broadly in real-world video distribution systems. When they are faced with the unified end-to-end QoE monitoring challenge we are targeting here, these disadvantages become even more pronounced.

To meet the challenge in a unified end-to-end QoE monitoring system, an objective QoE metric requires to have a number of must-have features. These include:

- *Accurate and light-weight.* The QoE metric must produce quality scores that accurately predict human visual QoE. The metric should be verified using independent, large-scale subject-rated video databases with diverse content and distortion types, and show high correlations with the opinions of an average human subject, as demonstrated by the scatter plot produced by the SSIMPLUS metric [13], [14] shown in Fig. 3. Meanwhile, the metric needs to be light-weight, allowing for real-time computations of high resolution videos (e.g., full high definition (HD), ultra-high definition (UHD) and 4K videos) with moderate hardware configurations. Such light-weight and speed requirement is critical in large-scale video distribution systems to reduce the overall cost and to maximize the flexibilities in terms of deployment, integration, scaling, and customization.

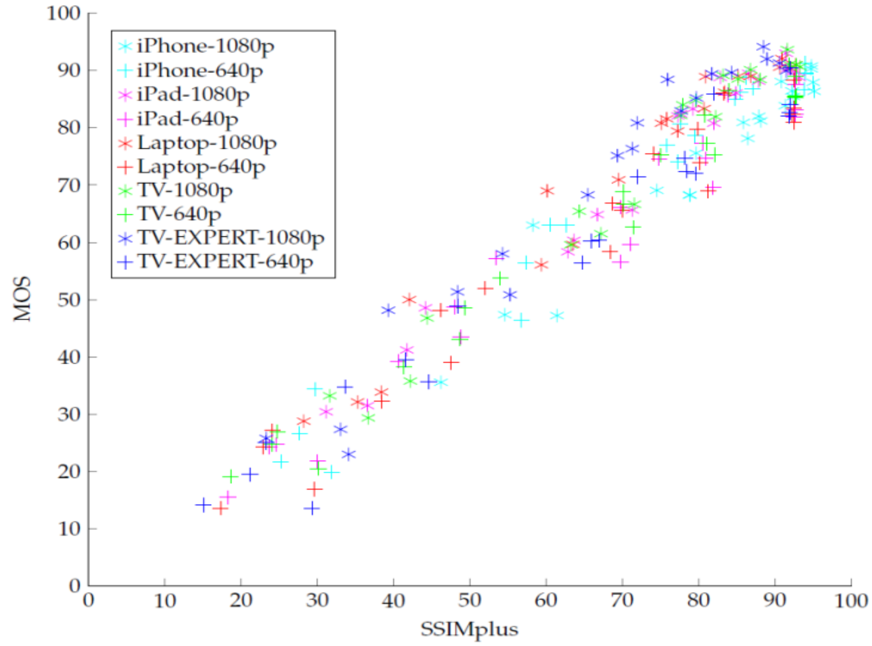

Figure 3 - Quality prediction accuracy performance evaluation of objective QoE metric

Each point in the scatter plot represents a test video. The horizontal and vertical axes are the quality prediction from an objective quality metric (in this case the SSIMPLUS metric [13], [14]) and the mean opinion score (MOS) obtained from subjective test, respectively. A good quality metric should produce a narrow-band cluster extending from low to high quality ranges, regardless of the mixed video content, resolution and viewing devices, as exemplified by the SSIMPLUS metric [13], [14] shown in the figure. The spearman rank-order correlation coefficient (SRCC) between SSIMPLUS and MOS is 0.97.

- *Easy-to-understand and easy-to-use.* The QoE metric must be easy-to-understand, directly producing QoE scores that linearly scale with what an average consumer would say about the video quality. For example, if the quality score range of the metric is between 0 and 100, then the total scale range may be divided into five even segments corresponding to five perceptual QoE categories of bad (0-19), poor (20-39), fair (40-59), good (60-79), and excellent (80-100), respectively. The QoE metric must be deployed with an easy-to-use user interface (UI), where the presentation is simple and intuitive, focusing on the most important trending information. Such an easy-to-understand and easy-to-use QoE metric defines a common language, under which engineers can identify/fix quality problems and optimize system performance, and executives are able to make critical business decisions.
- *Applicable and consistent across resolutions, frame rates, dynamic ranges, user devices and contents.* In addition to accuracy and speed, another critical problem that hinders the wide usage of existing well-known video quality metrics (PSNR, SSIM, MS-SSIM, VQM, VMAF) is their limited applicability. In particular, when videos are of different spatial resolutions, frame rates, and dynamic ranges, these metrics are not applicable, because all of them require pixel-to-pixel correspondence. Moreover, when the same video stream is displayed on different viewing devices (e.g., TV vs. tablet vs. smartphone), the perceptual QoE could be significantly different. However, all traditional metrics fail to make meaningful device-dependent QoE predictions.

Furthermore, these quality metrics often produce inconsistent scores across different content types (e.g., sports vs. news vs. animations), strongly limiting the usefulness of such metrics in large-scale distribution systems that operate on thousands of video service channels to make resource allocation decisions across the whole systems. Therefore, to implement a unified end-to-end quality assurance framework for many real-world video distribution systems (e.g., for multi-screen and adaptive bit rate (ABR) video delivery networks), consistent and cross-resolution, cross-frame rate, cross-dynamic range, cross-viewing device, and cross-content QoE assessments are essential.

- *Versatile for usage in single-ended, double-ended and more sophisticated scenarios.* Single-ended and double-ended video quality assessments refer to the different application scenarios where a reference video may or may not be available when assessing the quality of a test video. Double-ended or full-reference (FR) quality measures assume the reference video is accessible and of perfect quality. They are essentially signal fidelity measures and PSNR, SSIM, MS-SSIM, VQM and VMAF all belong to this category. On the other hand, single-ended or no-reference (NR) measures do not assume access to the reference video. Double-ended quality measures typically have higher quality prediction accuracy than single-ended approaches, but are much more difficult to apply. Very often, the reference videos are completely inaccessible. Even when they are accessible, for example, at video transcoders, the reference videos are often not well aligned with the test videos both in space and time. Moreover, the source videos received from content providers are often distorted themselves, creating even more complex scenarios where the reference videos are already degraded. In order to provide consistent QoE assessment at all points along the video delivery chain, the QoE metric has to be extremely versatile. The QoE metric needs to be easily plugged into single-ended, double-ended and more sophisticated scenarios. It also needs to make the best use of all resources to produce the most accurate QoE predictions. For example, at the transcoder, the QoE metric needs to precisely align the source and test videos before applying double-ended fidelity assessment. It also needs to appropriately handle the case when the reference video quality is already degraded.

All of the above are critical features for a QoE metric to work effectively in a unified end-to-end quality monitoring framework. Conventional and well-known video quality metrics (PSNR, SSIM, MS-SSIM, VQM, VMAF), however, are distant away from meeting these requirements. In practice, their usage is often limited to laboratory-testing environment, restricted to small-scale, non-time-critical use cases, e.g., encoder comparison on videos of the same content, spatial resolution, frame rate, and dynamic range.

The large gap between the limited performance and functionality of the well-known video quality metrics and the essential requirements of large-scale unified end-to-end QoE monitoring systems has motivated the development of the SSIMPLUS video QoE metric, which has been set to meet all the requirements throughout its design and implementation phases [13]. A recent study using 10 independent publicly-available subject-rated video databases (created from a collection of hundreds of thousands subjective ratings) evaluates conventional and state-of-the-art video quality metrics (including PSNR, SSIM, MS-SSIM, VMAF, SSIMPLUS and several other metrics), by comparing the quality predictions of these metrics against subjective mean opinion scores (MOS) [14]. The results showed that SSIMPLUS achieves the highest QoE prediction performance in terms of its correlation coefficients against MOS. It appears to be the only QoE metric that achieves an average correlation coefficient higher than 0.9. The same study also found that the SSIMPLUS metric to be 16.4 times faster than the VMAF metric, allowing SSIMPLUS to be computed in real-time in real-world applications [14]. The SSIMPLUS metric is applicable and produces consistent scores across resolutions, frame rates, dynamic ranges and content types. For every single video stream, it generates multiple QoE scores corresponding to a wide spectrum of viewing devices, from small screens on cellphones to large-size TVs. When applied to ABR encoding, SSIMPLUS simultaneously computes single-ended QoE scores of the source video input, together with double-ended scores for all the derivative video output produced by transcoders with different bitrates and resolutions. As well, it provides the absolute QoE scores of the derivative streams considering that the source input does not have perfect quality. At the client side, SSIMPLUS combines picture presentation quality with the impact of switching and stalling events to produce an overall QoE assessment for each individual user on a per-view basis [15], [16]. All of these computations are done at a speed faster than real-time. Due to these features, SSIMPLUS has been successfully deployed in large-scale operational environments, running 24/7 reliably and affecting the viewer experience of millions of users.

3. QoE-Driven Optimization

Many benefits come naturally once a unified end-to-end QoE monitoring solution is in place. The benefits are usually maximized through QoE-driven optimization. Here we use bandwidth optimization as an example. Bandwidth reductions without maintaining the right level of visual QoE makes little sense. Due to the lack of proper QoE assessment tools, currently most bandwidth optimization approaches in the industry result in inefficient and unstable results. The first step to success is to adopt a reliable QoE metric of superior accuracy and speed performance, and broad and powerful functionality. For example, it needs to perform meaningful and consistent video QoE assessment across resolutions, frame rates, dynamic ranges, viewing devices and video content.

Here we use SSIMPLUS as an example to illustrate how the cross-content, cross-resolution and cross-device features of a QoE metric may be employed to produce large bandwidth savings. Figure 4 plots the rate-quality curves of two video content (titles) at the same full-HD (1080p) resolution, assuming they are viewed on the same TV device. The rate-quality curve (or alternatively rate-distortion curve) is a widely used tool in the video coding technical community to evaluate and compare the performance of video encoders. Given a video title, together with its resolution and the quality evaluation criterion, each point on the rate-quality curve represents an operation point of the encoder in terms of a bitrate-quality combination. Thus, when we attempt to encode two titles, we end up with two rate-quality curves, as shown in Fig. 4. The gap between the two curves reveal the difference in encoding difficulty between the titles. To reach a target QoE quality level (e.g., SSIMPLUS = 90) using a fixed bandwidth (e.g., 4Mbps) to encode both videos would be a waste. Indeed, while 4Mbps is necessary for Title 1 to achieve the target quality level, only 3.1Mbps is necessary for Title 2 to achieve the same target, leading to a significant bandwidth saving.

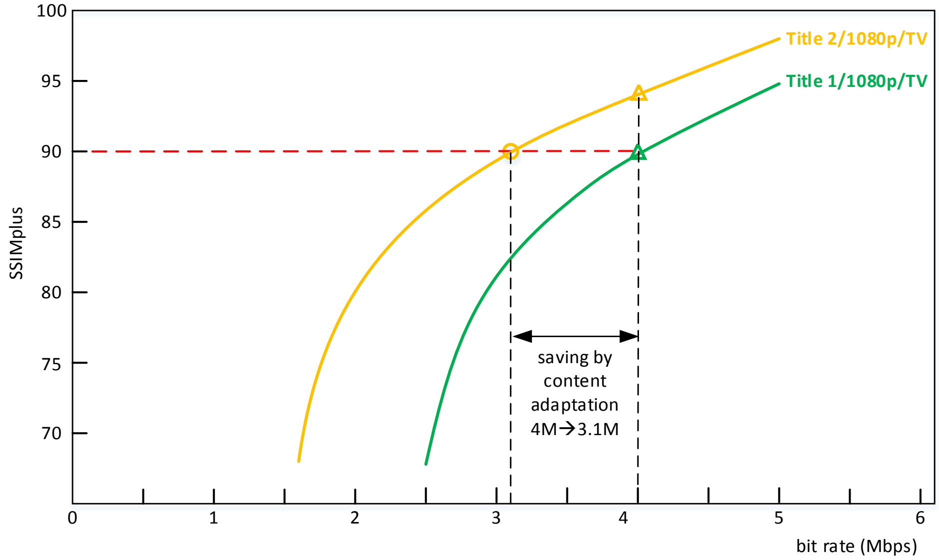

Figure 4 - Illustration of how bandwidth saving is achieved for given target quality (SSIMPLUS=90) by using a QoE metric that is able to provide consistent cross-content evaluation

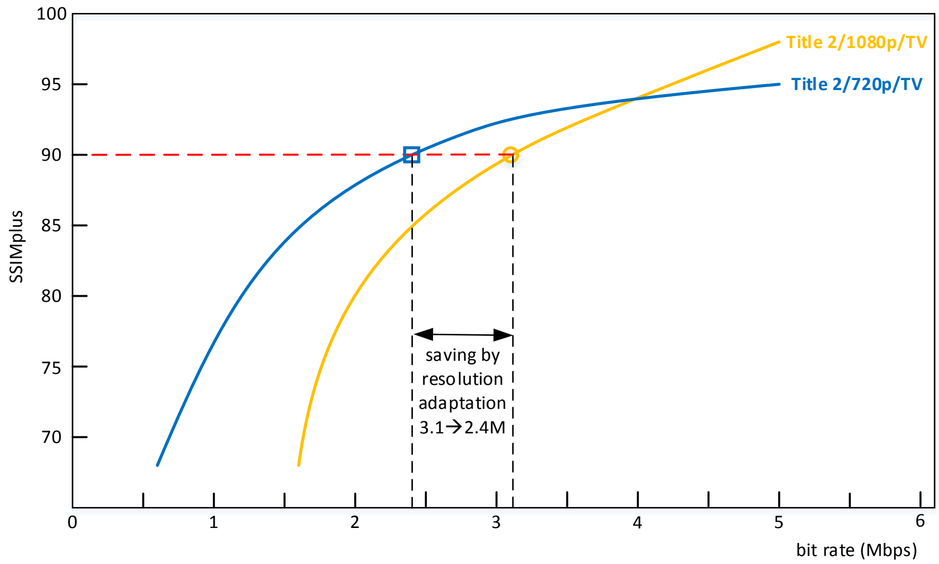

Figure 5 - Illustration of how bandwidth saving is achieved for given target quality (SSIMPLUS=90) by using a QoE metric that is able to provide consistent cross-resolution evaluation

For the same video content (title), when it is converted and then encoded to multiple resolutions, each resolution produces a different rate-quality curve, as exemplified in Fig. 5, where HD (720p) and Full HD (1080p) resolutions are used. It is commonly observed that the rate-quality curves for different resolutions cross at certain bitrate, as illustrated in Fig. 5. This is because when bitrate is high and compression artifacts are hardly visible, higher resolution video produces better sharpness and perceptual fidelity, but when bitrate gets lower, the quality of higher resolution video drops faster due to its high encoding difficulty. A good QoE metric that reflects such trend can help pick the most cost-effective resolution to achieve the target quality while saving large bandwidth. For example, for the same target quality (SSIMPLUS=90), a bandwidth reduction from 3.1Mbps to 2.4Mbps is obtained by switching from 1080p to 720p resolutions, as shown in Fig. 5.

For the same video content (title) encoded at the same resolution, the perceptual QoE could still vary significantly when the video is presented on different viewing devices. This is demonstrated by the rate-quality curves for a TV and a cellphone shown in Fig. 6. When the user is known to use a cellphone rather than a TV to watch the video, a bandwidth of 0.8Mbps is sufficient to achieve the same target quality level (SSIMPLUS = 90), down from 2.4Mbps on a TV. With all the content, resolution and device factors are combined (from Fig. 4 to Fig. 6), a total of 80% bandwidth savings may be obtained.

The example given here is for demonstration purposes only. In practice users may be able to explore more or fewer than the three factors above for maximum cost-savings. Our study suggests that for most video content and most common usage profiles, an average cost saving of 20%-60% is typically achieved by properly adopting QoE metric-driven bandwidth optimization. Such bandwidth savings can be obtained in both live and file-based video distribution systems by smart operation of video encoders and transcoders at the server, and may also be incorporated into adaptive streaming frameworks to achieve similar goals in a dynamic way.

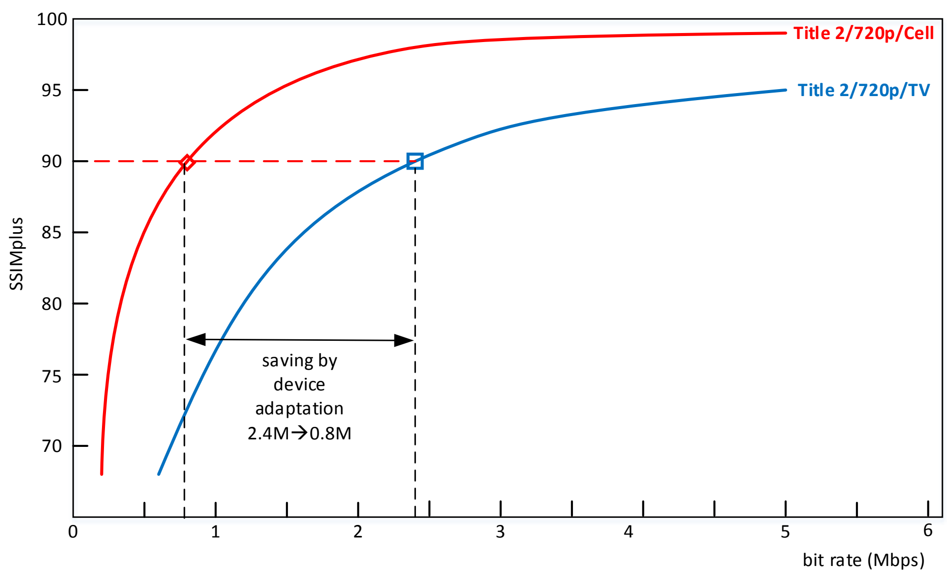

Figure 6 - Illustration of how bandwidth saving is achieved for given target quality (SSIMPLUS=90) by using a QoE metric that is able to provide consistent cross-device evaluation.

Conclusion

We propose a solution for unified end-to-end QoE monitoring, optimization and management in large-scale video distribution systems. The principle behind the solution is to start with end user's QoE in mind, such that all the QoE monitoring points should produce instantaneous scoring that reflects the end user's

QoE up to the monitoring point in the video delivery chain. The QoE scores need to be accurate, consistent and directly comparable, such that the monitoring solutions of the entire video distribution network speaks the same language. Such a unified end-to-end solution laid the groundwork for the subsequent operations for great benefits. Specifically, operation engineers will be able to immediately identify, localize and fix quality problem, design engineers will be able to perform effective and accurate optimizations on the video delivery chain and its individual components, and managing executives will have a clear picture about how video quality evolves throughout the distribution network and over long time scales, so as to make intelligent strategic decisions to manage the QoE of end users.

The most challenging task in implementing the proposed solution is to create an objective QoE metric that is not only accurate, fast, easy-to-understand and easy-to-use, but also applicable and consistent across resolutions, frame rates, dynamic ranges, viewer devices and contents. Moreover, it needs to be highly versatile for use in single-ended, double-ended and more sophisticated scenarios. Conventional and well-known video quality metrics such as PSNR, SSIM, MS-SSIM, VQM and VMAF fall short of meeting these requirements. As a result, their usage is limited to lab-testing environment or small-scale use cases. This has motivated the recent development of novel video QoE metrics such as SSIMPLUS [13], [14], which has been deployed in real-world large-scale QoE monitoring systems.

To further demonstrate the benefits of adopting the proposed framework and QoE metric, we use bandwidth optimization as an example, which demonstrates that large bandwidth savings can be obtained by adopting a QoE metric such as SSIMPLUS. With the wide deployment of the proposed solution and QoE metrics in large-scale video distribution networks. The QoE data collected in large and varying space and time-scales constitutes a valuable source for big data analytics and strategic intelligence, which is an interesting direction for future investigations.

Abbreviations

ABR	adaptive bit rate
FR	full-reference
HD	high definition
MOS	mean opinion score
MS-SSIM	multi-scale structural similarity
NR	no-reference
PSNR	peak signal-to-noise ratio
QA	quality assurance
QoE	quality of experience
QoS	quality of service
SRCC	Spearman rank-order correlation coefficient
SSIM	structural similarity
UHD	ultra-high definition
UI	user interface
VMAF	video multi-method assessment fusion
VQA	video quality assessment
VQM	video quality model

References

1. Cisco Inc., “Cisco Visual Networking Index: Forecast and Methodology 2015-2020”, 2016.

2. Limelight Networks, "The state of online video," https://www.limelight.com/video/, 2016.

3. C. Curtis, *et al.*, "American Society of Cinematographers Technology Committee Progress Report 2016," *SMPTE Motion Imaging Journal*, vol. 125, no. 7, pp. 43-58, Sept. 2016.

4. Z. Wang, "New quality-of-experience measurement technologies: streamlining how videos are delivered to consumers," *IEEE Signal Processing Society Blogs*, July 2017.

5. M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, Sept. 2014.

6. O. Oyman, S. Singh, "Quality of experience for HTTP adaptive streaming services," *IEEE Communications Magazine*, vol. 50, Apr. 2012.

7. ITU QoE Recommendation ITU-T P.10/G.100, Amd.1, *New Appendix I Definition of Quality of Experience (QoE)*, 2007.

8. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, Apr. 2004.

9. Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, Feb. 2004.

10. Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conference on Signals, Systems and Computers,* Nov. 2003.

11. M. H. Pinson, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312-322, Sept. 2004.

12. Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy and M. Manohara, *"Toward A Practical Perceptual Video Quality Metric*," *Netflix TechBlog, June, 2016*

13. A. Rehman, K. Zeng and Z. Wang, "Display device-adapted video quality-of-experience assessment," *IS&T/SPIE Electronic Imaging: Human Vision & Electronic Imaging*, Feb. 2015.

14. SSIMPLUS: The most accurate video quality measure, https://www.ssimwave.com/from-the-experts/ssimplus-the-most-accurate-video-quality-measure/

15. Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang "A quality-of-experience index for streaming video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 154-166, Feb. 2017.

16. Z. Duanmu, K. Ma and Z. Wang, "Quality-of-experience of adaptive video streaming: exploring the space of adaptations, " *ACM Multimedia*, Mountain View, CA, Oct. 2017.

