SPIE Optics + Photonics (invited paper), San Diego, CA, Aug. 2015
Published in: Applications of Digital Image Processing XXXVIII, Proc. SPIE, vol. 9599. @SPIE

Objective Video Presentation QoE Predictor for Smart Adaptive Video Streaming

Zhou Wang, Kai Zeng, Abdul Rehman, Hojatollah Yeganeh, and Shiqi Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Emails: {zhou.wang, kzeng, abdul.rehman, hyeganeh, s269wang}@uwaterloo.ca

ABSTRACT

How to deliver videos to consumers over the network for optimal quality-of-experience (QoE) has been the central goal of modern video delivery services. Surprisingly, regardless of the large volume of videos being delivered everyday through various systems attempting to improve visual QoE, the actual QoE of end consumers is not properly assessed, not to say using QoE as the key factor in making critical decisions at the video hosting, network and receiving sites. Real-world video streaming systems typically use bitrate as the main video presentation quality indicator, but using the same bitrate to encode different video content could result in drastically different visual QoE, which is further affected by the display device and viewing condition of each individual consumer who receives the video. To correct this, we have to put QoE back to the driver's seat and redesign the video delivery systems. To achieve this goal, a major challenge is to find an objective video presentation QoE predictor that is accurate, fast, easy-to-use, display device adaptive, and provides meaningful QoE predictions across resolution and content. We propose to use the newly developed SSIMplus index (`https://ece.uwaterloo.ca/~z70wang/research/ssimplus/`) for this role. We demonstrate that based on SSIMplus, one can develop a smart adaptive video streaming strategy that leads to much smoother visual QoE impossible to achieve using existing adaptive bitrate video streaming approaches. Furthermore, SSIMplus finds many more applications, in live and file-based quality monitoring, in benchmarking video encoders and transcoders, and in guiding network resource allocations.

Keywords: quality-of-experience, video streaming, adaptive video streaming, video quality assessment, visual communication, video compression, SSIM, SSIMplus

1. INTRODUCTION

In recent years there has been a tremendous growth of Internet-based video-on-demand (VoD) services based on over-the-top (OTT) technologies, exemplified by the remarkable success of Netflix and Youtube. An increasingly popular approach for such OTT services is the adoption of adaptive video streaming techniques.[1] In adaptive video streaming, each source video content is encoded/transcoded into multiple variants (also referred to as streams or representations) of different bitrates and resolutions in the video stream preparation stage. The video streams are divided into time segments in the order of seconds and all streams are stored at the video hosting server. When a client application, such as a video player in a mobile device, requests to play the video content remotely, it can adaptively pick one of the multiple streams instantly for each time segment to be transmitted from the hosting server based on the current network conditions and the buffer size, playback speed, power consumption and other instant states of the receiving device. The adaptive video streaming framework puts the burden at the video server side in terms of increased CPU power for repeated encoding/transcoding demand and increased storage space to store many representations of the same content. On the other hand, it allows to serve users of large variations in terms of their connections to the network without relying on dedicated networks or changing the existing network delivery infrastructure. It is also attractive for its potential to provide the best possible service to each individual user on a per-user and per-moment basis.

Nevertheless, a major problem with the current implementation and deployment of adaptive streaming techniques is that the viewers' quality-of-experience (QoE) is not properly taken into account. Video quality assessment (VQA) has been an active research topic in recent years.[2,3] The simplest and most widely used VQA measure is the mean squared error (MSE) and peak signal-to-noise ratio (PSNR), which are simple to calculate and are mathematically convenient in the context of optimization, but they are not well matched to perceived visual quality.[4] State-of-the-art VQA methods that have been drawing significant attention in recent years include the structural similarity index (SSIM),[5–7] the multi-scale structural similarity index (MS-SSIM),[8] the video quality metric (VQM)[9] (which is recommended by ITU and adopted by ANSI as a US national standard), and the motion-based video integrity evaluation index (MOVIE).[10] All of them have achieved better quality prediction performance than MSE/PSNR. However, there are significant limitations when applying VQA models to VoD applications, because VQA models assess the perceptual quality of the video stream only, without considering the perceptual quality variations when the video is undergoing network transmission and displayed on different devices, at different resolutions, and under different viewing conditions.[11] What we need are video QoE predictors that take into account as much such variations as possible. Since the ultimate goal of video delivery services is to provide the clients with the best possible video in terms of their visual QoE, properly assessing visual QoE and using such assessment as the key factor in the design and optimization of the video delivery systems is highly desirable.

Indeed, consumers' expectations for better QoE nowdays have been higher than ever before. Based on a recent viewer experience study,[12] "In 2012, global premium content brands lost $2.16 billion of revenue due to poor quality video streams and are expected to miss out on an astounding $20 billion through 2017". The poor video quality keeps challenging the viewers' patience and becomes a core threat to the video service ecosystem. According to the same study,[12] roughly 60% of all video streams experienced quality degradation in 2012. In another recent study,[13] 90.4% interviewers reported "end-user video quality monitoring" as either "critical", "very important", or "important" to their video initiatives, and almost half of the customer phone calls is related to video quality problems in VOD services and HDTV. Therefore, effective and efficient objective video QoE assessment tools can play a critical role in current video delivery systems. Unfortunately, this is exactly what is lacking in the decision making process of the current adaptive video streaming implementations. Real-world systems are essentially *bitrate-driven* (rather than *quality-driven*) where bitrate is used as the key factor, erroneously equated to a visual quality indicator. Equating bitrate and quality is an extremely poor assumption because using the same bitrate to encode different video content could result in dramatically different visual quality, possibly ranging between the two extremes on a standard five-category subjective rating scale (Excellent, Good, Fair, Poor, Bad). Besides, the actual user QoE varies depending on the device being used to display the video (among a number of viewing conditions that could change viewer experiences), another factor that are not taken into consideration in bitrate-driven streaming strategies.

There are two major focus points of the current work. The first is to find an objective video presentation QoE predictor that is suited for practical usage in video delivery services, especially for OTT-based adaptive video streaming applications. The second is to make use of such a QoE predictor in the decision making process of adaptive video streaming. We call such a *QoE-driven* adaptive streaming approach *smart streaming*, which is shown to significantly change the decision making process in adaptive streaming strategies, and may lead to a number of highly desirable benefits, including improved video quality, smoothed video quality, reduced bandwidth, reduced probability of video freezing, and reduced data usage and power consumption at the client receiving devices.

2. OBJECTIVE VIDEO PRESENTATION QOE PREDICTOR

Objective video presentation QoE predictor aims to predict from the hosting server side the perceptual QoE of the video stream presented at the receiver device, where the focus is on the presentation quality of the video without considering the errors occurred during transmission or the video stalling events during playback. The overall QoE of an end receiver should be a combined effect of presentation QoE and the instant transmission QoE, where in OTT applications, the latter is mainly determined by the severeness of the video playback delay and stalling events. The overall QoE cannot be fully predicted at the hosting server side and is a much more complicated problem. Here we are mainly concerned about the presentation QoE, which serves as the basis of a good overall QoE estimator. A list of desired properties of a presentation QoE predictor is given below.

- *High accuracy*: Since the ultimate receivers of the video streams are human eyes, the presentation QoE must predict with high accuracy a quality score that an average consumer would give to a video stream.

- *High speed*: The volume of video data is increasing exponentially and in OTT applications multiple video streams are created and stored for each video content. The QoE predictor needs to quickly evaluate the video streams. Faster than real-time computation is required in many application scenarios.

- *Reliable, easy-to-understand & easy-to-use*: The QoE predictor must provide reliable quality predictions for video streams of a large variety of video content, complexity and resolution, and displayed on a variety of viewing devices. It is desired to produce easy-to-understand scores that directly tells what a typical consumer would say about the video quality (bad, poor, fair, good, excellent). It is also desirable to be easily embedded into existing video encoding, decoding and transmission systems for monitoring and optimization purposes.

- *Cross-content assessment*: Real-world video sources exhibit great variations in terms of content types and spatial and motion complexities. A useful QoE predictor is expected to produce meaningful quality scores across large variations of video content.

- *Cross-device assessment*: The visual QoE of the same video stream displayed on different receiving device (e.g., HDTV versus smart phone) could be drastically different. The presentation QoE predictor should be able to take the device properties as well as the viewing conditions into account and produces meaningful scores across different devices.

- *Cross-resolution assessment*: In OTT applications, the video source needs to be transcoded into video streams of different resolutions. As a result, a full-reference presentation QoE predictor needs to handle the case when the reference and test videos have different resolutions (reduced- and no-reference VQA methods[3, 14] may not have the problem but are inferior in quality prediction accuracy). It is also expected to provide useful predictions when the resolution of the display window on the viewing device differs from that of the video stream.

Unfortunately, none of the existing popular VQA models (PSNR, SSIM,[5–7] MS-SSIM,[8] VQM,[9] MOVIE[10]) demonstrates a satisfying coverage of these properties. In particular, all of them evaluate video quality based on the video streams and thus produce the same quality score of a given video stream, regardless of the viewing device on which the video stream is displayed. Moreover, none of them can handle the cross-resolution cases properly.

In order to overcome the limitations of existing VQA models, we proposed a novel video presentation QoE model named SSIMplus,[11] which is built upon the basic philosophy of SSIM,[6] but is a much more advanced model that takes into account various human visual system characteristics as well as display device and viewing condition properties. In addition, SSIMplus also has several other attractive features that promotes its practical usage. First, it can be implemented faster than real-time, even for 4K videos when equipped with GPU. Second, it produces easy-to-understand and easy-to-use scores between 0 and 100, evenly divided into five regions corresponding to bad, poor, fair, good and excellent quality, respectively. Third, it generates in real-time a quality map at per-pixel precision that indicates how local quality varies across space and time, a property that is especially useful for video engineers to optimize their algorithms. More details about the SSIMplus model can be found in[11] or at `https://ece.uwaterloo.ca/~z70wang/research/ssimplus/`.

Table 1: Display Devices & Viewing Conditions employed in the subjective study

Display Device	Diag. Screen Size (in)	Resolution	Brightness (cd/m^2)	Viewing Distance (in)
iPhone 5S	4″	1136×640	556	10
iPad Air	9.7″	2048×1536	421	16
Lenovo Laptop	15.6″	1920×1080	280	20
Sony TV	55″	1920×1080	350	90
Sony TV (TV-Expert)	55″	1920×1080	350	40

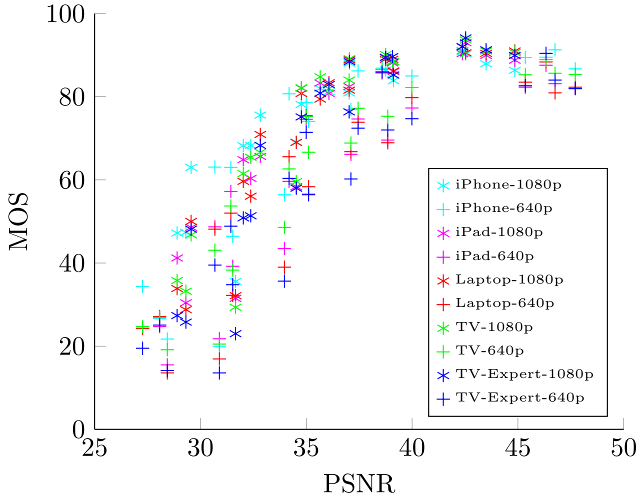


Figure 1: Scatter plot of MOS versus PSNR.

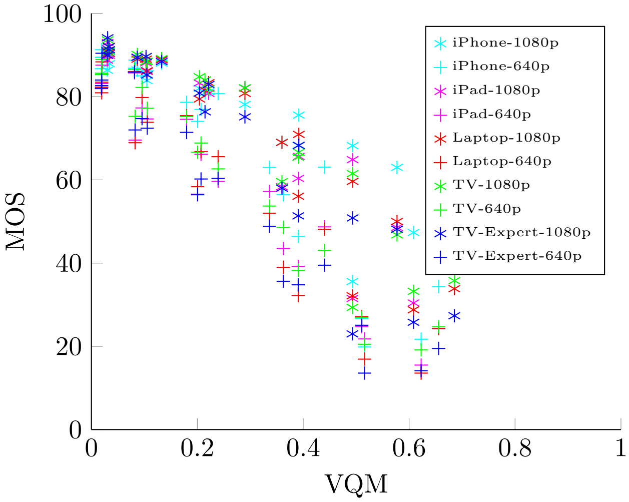


Figure 2: Scatter plot of MOS versus VQM.

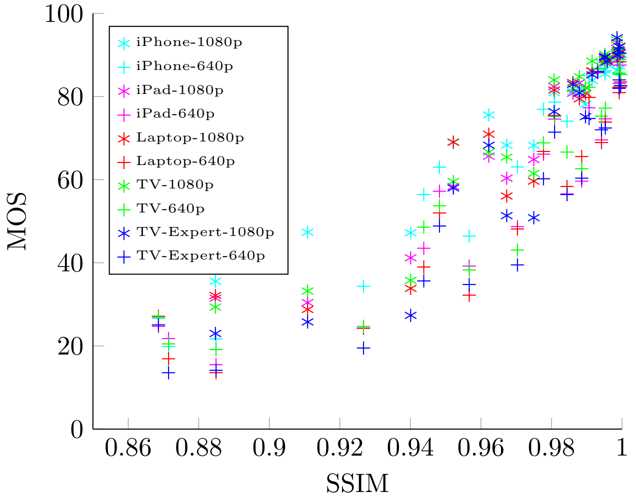


Figure 3: Scatter plot of MOS versus SSIM.

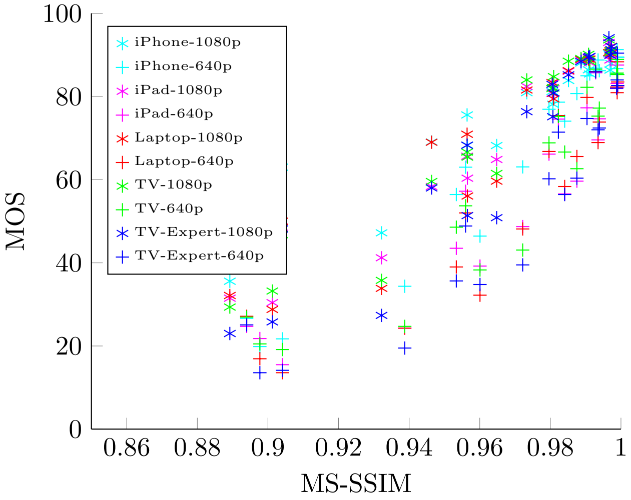


Figure 4: Scatter plot of MOS versus MS-SSIM.

To validate and compare different video quality models, we created a video database and conducted subjective quality evaluations on video streams at different resolutions and displayed on different viewing devices under different viewing conditions. The source video contents are selected to contain indoor and outdoor scenes, simple and complex textures, camera zooming/panning and object motion towards different directions. The source videos are progressive, 10-second long, at a frame rate of 24 frames/second, and in both 1920 × 1080 and 1136 × 640 resolutions. All source video sequences are compressed at five quality levels. The subjective test generally follows the Absolute Category Rating (ACR) methodology, as suggested by ITU-T recommendation P.910.[15] Thirty naïve subjects took part in the subjective test. The first few video sequences were repeated at the end of the test to measure the fatigue factor. We found out that there was no bias or significant statistical difference between the subjective scores obtained for the same set of video sequences at the beginning or the end of the test. The test videos were scored by subjects under the viewing conditions provided in Table 1. Six out of the thirty subjects were found to be outliers.[16] The remaining valid scores were averaged for each test video, resulting in a mean opinion score (MOS) of the video.

In addition to the newly proposed SSIMplus, the other VQA models under comparison include PSNR, SSIM,[6] MS-SSIM,[8] VQM,[9] and MOVIE,[10] which are the most popular VQA models widely recognized in academia. We have also included several other models from industrial commercial products, including the picture quality rating by Tektronix PQA600 (PQR-Tek),[17] the difference of MOS (DMOS) measure by Tektronix (DMOS-Tek),[17] the

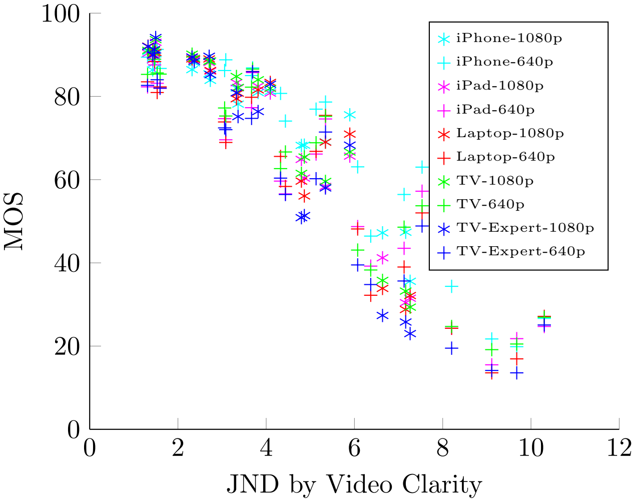


Figure 5: Scatter plot of MOS versus JND-VC.

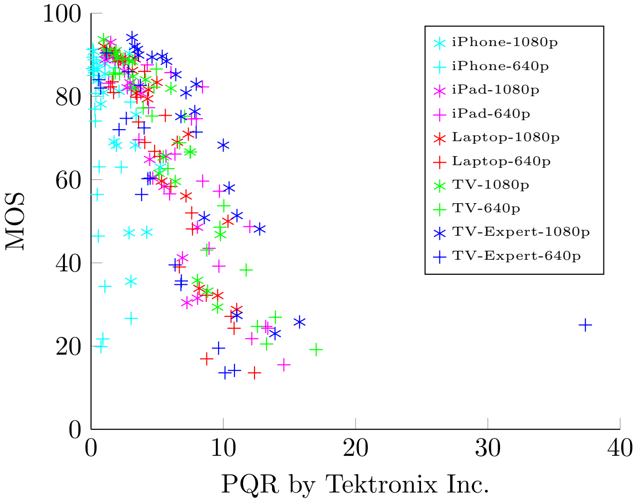


Figure 6: Scatter plot of MOS versus PQR-Tek.

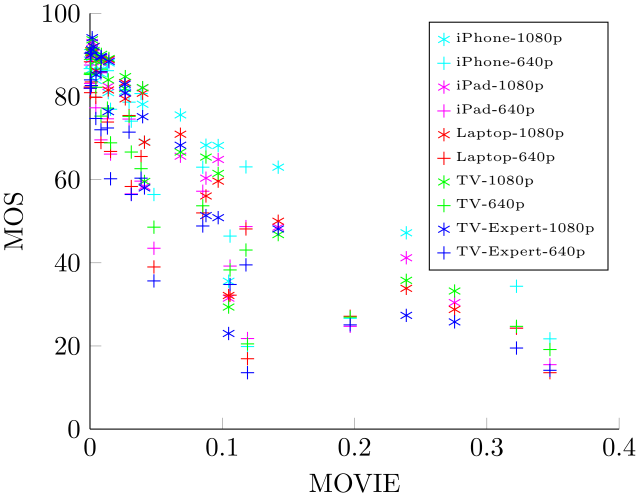


Figure 7: Scatter plot of MOS versus MOVIE.

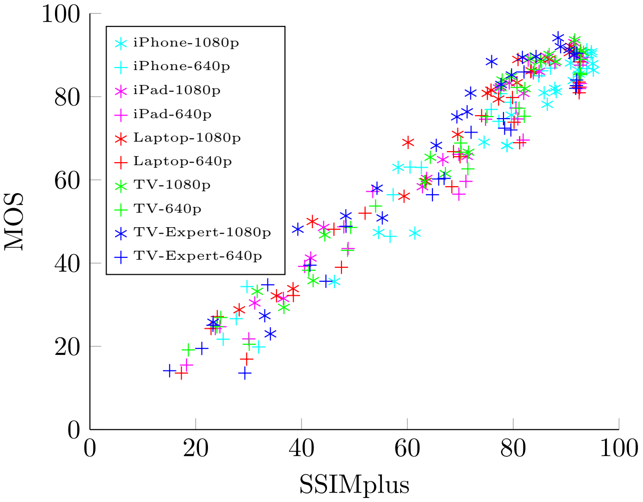


Figure 8: Scatter plot of MOS versus SSIMplus.

Table 2: Performance Comparison between PSNR, SSIM, MS-SSIM, VQM, PQR-Tek, DMOS-Tek, JND-VC, DMOS-VC and SSIMplus including all devices and viewing conditions

Model	PLCC	MAE	RMS	SRCC	KRCC	Complexity (normalized)
PSNR	0.9062	7.4351	9.8191	0.8804	0.6886	1
SSIM	0.9253	6.9203	8.8069	0.9014	0.7246	22.65
MS-SSIM	0.8945	8.1969	10.384	0.8619	0.6605	48.49
VQM	0.8981	8.0671	10.214	0.8703	0.6711	174.53
MOVIE	0.9096	7.4761	9.6493	0.8892	0.7001	3440.27
JND-Tek	0.7615	11.372	15.052	0.6972	0.5241	54.22
DMOS-Tek	0.7568	11.478	15.180	0.6969	0.5236	54.22
JND-VC	0.9289	6.7096	8.5986	0.9206	0.7469	443.15
DMOS-VC	0.8365	9.9292	12.724	0.8090	0.6027	13.78
SSIMplus	**0.9732**	**4.3192**	**5.3451**	**0.9349**	**0.7888**	**7.83**

JND measure by Video Clarity (JND-VC),[18] and the DMOS measure by Video Clarity (DMOS-VC).[18]

The scatter plots of the VQA algorithms under comparison are shown in Figs. 1 - 8, where in each figure, a sample point corresponds to one test video, for which the vertical axis is the MOS value and the horizontal axis indicates the quality prediction by an objective VQA model. An excellent VQA model would provide a monotonic prediction of the MOS scores, and thus the scatter of the sample points is expected to be as tight as possible, indicating small prediction errors. From these plots, the superior performance of the SSIMplus algorithm is evident compared to the other VQA models, which lack the capability of properly differentiating the variations in display devices and viewing conditions.

The objective performance metrics include Pearson linear correlation coefficient (PLCC) after a nonlinear mapping between the subjective and objective scores, mean absolute error (MAE) between the true and predicted MOS values after nonlinear mapping, root mean-squared error (RMS) between the true and predicted MOS values after nonlinear mapping, Spearman rank correlation coefficient (SRCC) between the subjective and objective scores, Kendall rank correlation coefficient (KRCC) between the subjective and objective scores, and the computational complexity estimated by normalized computation time based on PSNR (which is assumed to have unit complexity). Both SRCC and KRCC are independent of any nonlinear mapping attempting to convert the objective scores to match those of the MOS values. A better objective VQA measure should have *higher* PLCC, SRCC and KRCC scores and *lower* MAE, RMS and complexity values. These performance metrics are widely adopted in previous VQA studies.[3, 19, 20] The performance comparison results are summarized in Table 2, where the best results are highlighted in bold face. The results turn out to be a confirmation of what we have observed in Figs. 1 - 8. It can be seen that SSIMplus clearly outperforms all other VQA models by a significant margin. This is not only due to its use of better models for human visual characteristics, but perhaps more importantly, because it takes into account the display device and viewing condition parameters, which may substantially change viewer's perceptual experience (for example, the same video viewed on a smart phone and an HDTV could result in drastically different perceptual experience). Another important observation from Table 2 is that it is much faster than all other models except for PSNR. Such a low complexity adds very little computational overhead on top of existing video coding/transcoding, multiplexing and transmission tasks, and is a critical feature in many real-world application scenarios. In conclusion, our results suggest that SSIMplus is a highly attractive candidate as an effective and well-rounded presentation QoE prediction model that is well suited to network video delivery applications, both as a quality monitoring tool and as a guidance to drive and optimize resource allocations.

3. SMART ADAPTIVE VIDEO STREAMING

The major differences between our QoE-drive smart streaming and the traditional adaptive bitrate streaming are the involvement of QoE predictions throughout the video delivery process and the use of QoE estimation as a key factor in the switching decisions made at the receiving device.[21] A flow diagram that summarizes the smart streaming idea is given in Fig. 9, where in general the QoE prediction could make use of all possible information at the server, inside the network, and/or at the receiver. The decision-making adaptive switching module is at the end receiver, which collects all QoE related information (including that about the video streams, network conditions and device/viewing conditions) and performs instant QoE estimation for each time segment, and based on which, makes the switching decision for the next time segment.

A flow diagram that involves QoE predictions of the video streams at the video stream preparation stage is given in Fig. 10, which gives a general description on where the QoE prediction is performed and used. In particular, if one decides to adopt SSIMplus in the process, the multiple encoded/transcoded video streams are evaluated by SSIMplus for a list of presumed display devices and viewing conditions. One may make use of the predicted QoE scores to adaptively adjust the video encoder/transcoder for better perceptual quality or reduced bandwidth. One may also multiplex the predicted QoE scores with the video streams in the packaging module (e.g., being embedded into the headers of the video files, or included as part of the metadata transmitted to the client in an XML file prior to the transmission of the video streams) and transmit them through the network to the receiver device along with the video streams.

At the receiver side, once the SSIMplus scores of all time segments and all available video streams for a number of viewing devices/conditions are received, a dynamic QoE estimation is performed by selecting the

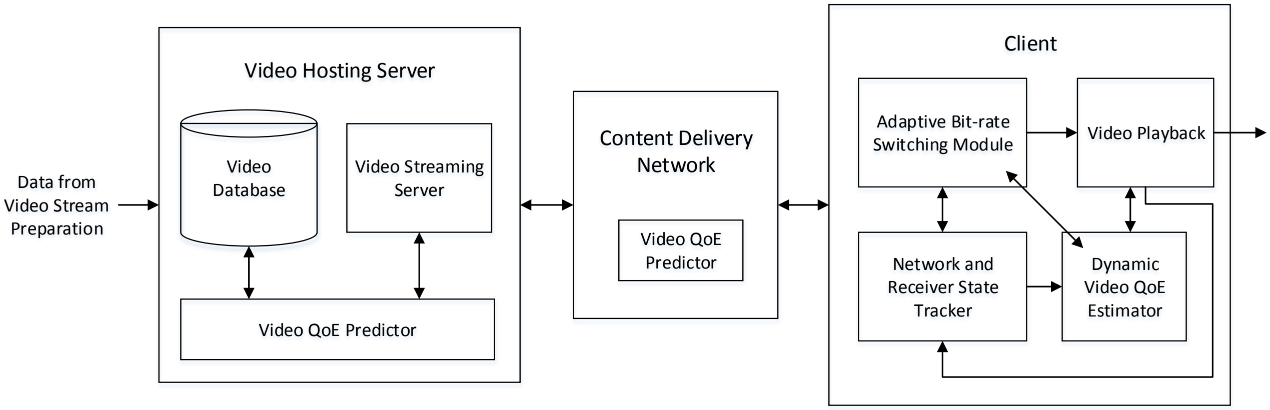


Figure 9: Flow diagram of QoE-driven smart streaming between the video hosting server, content delivery network, and client.

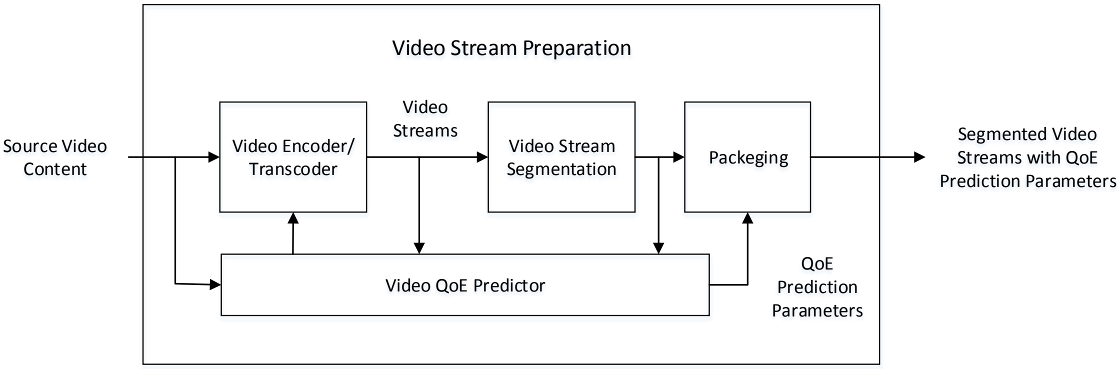


Figure 10: Flow diagram of the video stream preparation stage at the hosting server side that involves video QoE prediction.

scores of the most suitable device (when the actual receiver device is or close to one of the presumed devices in the QoE prediction process at the hosting server) or by combining the scores of multiple received scores. Such a QoE estimation is further refined by the instant network and device/viewing conditions. This results in a matrix of QoE estimation,[21] where each entry is a SSIMplus score of a particular time segment and a particular representation of the video content. The switching decision making process is then carried out by making use of the QoE estimation matrix. A number of *smart* strategies are available to make wiser decisions than the case when the QoE information is missing.[21] For example, one may reject to switch to an affordable higher bitrate and/or higher resolution stream, when without such switching, the QoE maintains at or above a pre-determined target threshold level. This is different from existing *best-effort* approaches which always make the attempt at the client side to request the stream of the highest affordable bitrate, regardless of the actual QoE of that stream. Such difference allows one to save bitrate and reduce the probability of rebuffering while maintaining the QoE level because more video content can be buffered for the same network bandwidth. For another example, one may decide to switch to a lower bitrate and/or lower resolution stream even without seeing a drop in network bandwidth or buffer size, when such switching results in QoE drops lower than a threshold value, and/or when with such switching, the QoE maintains at or above a pre-determined target threshold QoE level. This is again different from existing adaptive streaming approaches which make the best effort at the client side to request the stream of the highest affordable bitrate. Thus when there is no drop in network bandwidth or buffer size, existing adaptive streaming approaches keep requesting the streams equaling or higher than the bitrate of the

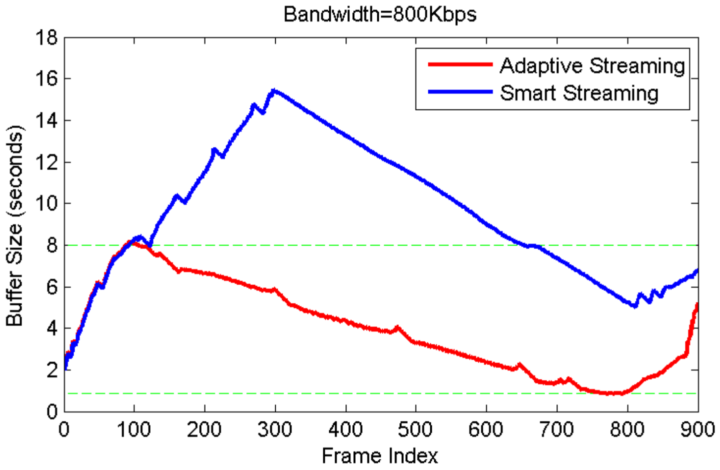


Figure 11: Adaptive streaming vs. smart streaming: receiving device buffer size as a function of frame index.

current stream, but will not switch to a lower bitrate stream. The capability of switching to a lower bitrate in the smart streaming scheme allows one to save bitrate at the current moment of low complexity content and reserve bandwidth and buffer capacities for future video segments that are more complex and desire more bitrates to maintain the QoE. As such, one can improve the smoothness of QoE, reduce the probability of rebuffering and stalling event at future complex segments, and increase the overall QoE. In summary, the availability of reliable QoE estimations allows one to use a number of smart decision making strategies to 1) save the overall bandwidth for the delivery of the video content without sacrificing the client users' QoE; 2) create better overall visual QoE of the client users; 3) create smoother visual QoE of the client users; and 4) reduce the probability of rebuffering or stalling events at the client user device.[21]

To demonstrate how smart streaming differentiates from adaptive streaming approaches, an illustrative example is given in Fig. 11 through Fig. 15. Assume that there are three layers of video streams from the same source content at the hosting server that have bitrates of 500kbps, 1000kbps and 2000kbps, respectively. (The actual bitrate of each video frame fluctuates). Also assume that the network bandwidth is a constant at 800kbps, and the player at the client side initially buffered 2 seconds of video before starting the video playback. Figure 11 compares the buffer size as a function of frame number index in the adaptive streaming and smart streaming approaches, where the adaptive streaming approach uses 8-second buffer and 2-second buffer as two thresholds to trigger the switchings to higher bitrate and lower bitrate, respectively. In this particular example, since the actual network bandwidth of 800kbps is between the bitrates of the first layer video stream of 500kbps and the second layer video stream of 1000kbps, the adaptive streaming approach will alternatively switch between these two layers. The resulting switching decisions can be visualized in the upper plot of Fig. 12, and the resulting actual bitrate as a function of frame index is shown in the upper plot of Fig. 13. Such a performance is normal and indeed ideal in existing adaptive streaming systems that use bitrate as the indicator of video quality, because the curve of bitrate versus frame index is smooth, as can be seen in the upper plot of Fig. 13. However, constant (or similar) bitrate does not mean the same video quality or visual QoE, which largely depends on the complexity of the video content. In this particular case, the last portion of the video content is much more complicated than the earlier portions. As a result, although the last portion of the adaptive streaming video has similar bitrate when compared with the earlier portions, the visual QoE is significantly lower. This can be measured using an effective QoE measure such as SSIMplus, as given in Fig. 14, which successfully indicates that the viewer's QoE changes dramatically from the earlier to the later portions of the video. This could lead to significant drops of the overall visual QoE and largely affect user dissatisfaction and customer engagement. By contrast, the smart streaming approach behaves differently in this scenario. Fig. 11 shows the buffer size as a function of frame

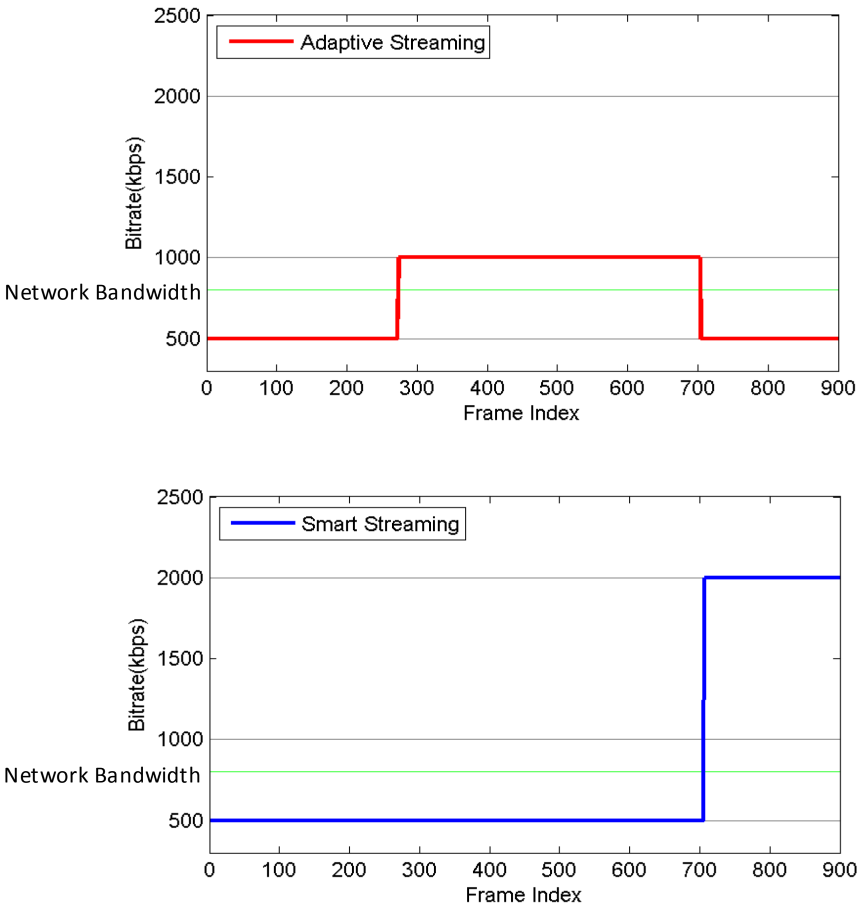


Figure 12: Adaptive streaming vs. smart streaming: switching decision as a function of frame index. Upper: adaptive streaming; Bottom: smart streaming.

index, the lower plot of Fig. 12 illustrates the actual switching decisions of the smart streaming case, and the lower part of Fig. 13 gives the resulting actual bitrate as a function of frame index. There are several important differences that have been made when compared with the adaptive streaming approach. First, because the QoE of future frames in each layer of video stream is available, the smart streaming module does not trigger switching to a higher bit rate in the middle part of the video, because such switching does not lead to sufficient improvement of QoE, and also because the smart streaming module is forseeing the highly difficult future segments (last portion of the video). Second, to maintain the smoothness of visual QoE for the last portion of the video, the smart streaming module triggers a switch to the third layer of video stream of 2000kbps, which is a much higher bitrate than the network bandwidth. The resulting switching decision in the lower plot of Fig. 12 and the actual bitrate as a function of frame index in the lower plot of Fig. 13 exhibit very large jumps to higher bitrate at the last portion of the video, which is an effect that is not observed in existing adaptive streaming approaches. Such a new decision making strategy leads to a SSIMplus based QoE curve in Fig. 14, where the smart streaming curve maintains at a high quality level throughout the entire video, with significantly better smoothness and a better overall performance in QoE. Some visual examples are given in Fig. 15, where although the frames extracted from adaptive streaming video better maintains a similar bitrate, their perceptual quality appears to be unstable (in particular, the quality of Frame 800 drops largely from Frames 200 and 500). By contrast, the

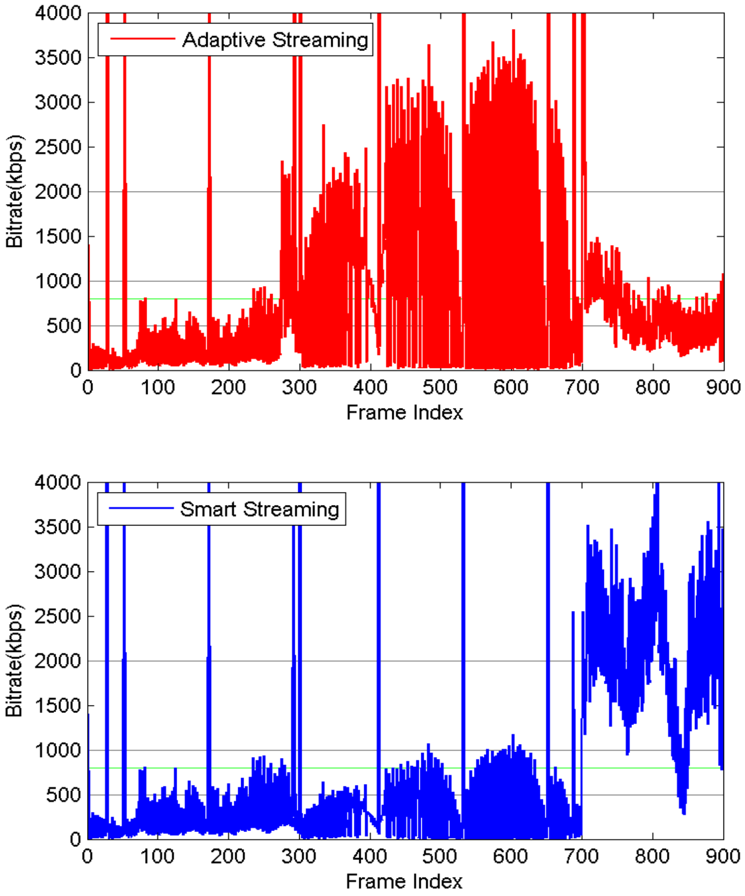


Figure 13: Adaptive streaming vs. smart streaming: frame bit-rate as a function of frame index. Upper: adaptive streaming; Bottom: smart streaming.

smart streaming video exhibits much more consistent quality across the video content. In this example, it is also noted that the total bitrate of the smart streaming case is even lower than that of the adaptive streaming case (which can be determined by the buffer sizes at the end of the curves in Fig. 11). Furthermore, the buffer size curve of the smart streaming case in Fig. 11 is mostly higher than that of adaptive streaming, meaning that smart streaming is better prepared to reduce rebuffering and stalling events. In summary, because of the adoption of the smart streaming approach driven by an effective QoE predictor, one can gain better overall and smoother user QoE than existing adaptive streaming methods, together with the potential benefits of using a lower overall bitrate and maintaining a healthier buffer.

It is worth mentioning that the employment of device-adaptive QoE predictors such as SSIMplus together with the smart streaming strategies exemplified above and described in more detail in[21] allows the adaptive switching module at the user device to make different switching decisions depending on the device and window size/resolution being used to display the video content. This could potentially lead to large differences in the streams being transmitted from the hosting server. For example, a consumer using a smartphone may request a video stream of a much lower bitrate than that requested by another consumer watching the same video content on an HDTV. Based on our experience with SSIMplus, this could often save the smartphone user more than half of his/her data usage without sacrificing the presentation QoE, along with added benefits such as improved

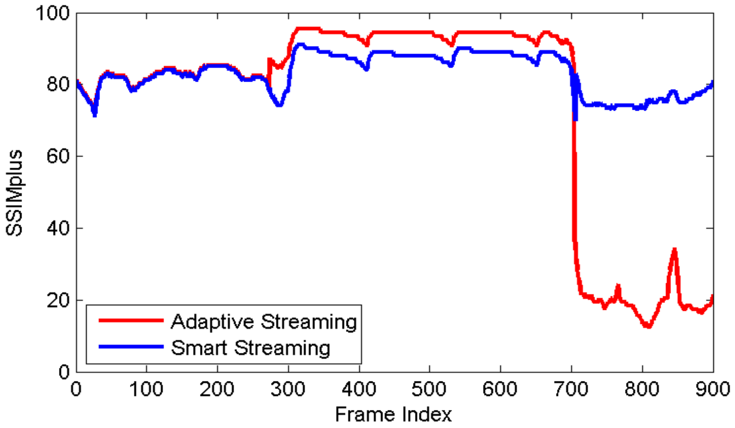


Figure 14: Adaptive streaming vs. smart streaming: SSIMplus quality index as a function of frame index.

battery life and reduced likelihood of video freezing.

4. CONCLUSION

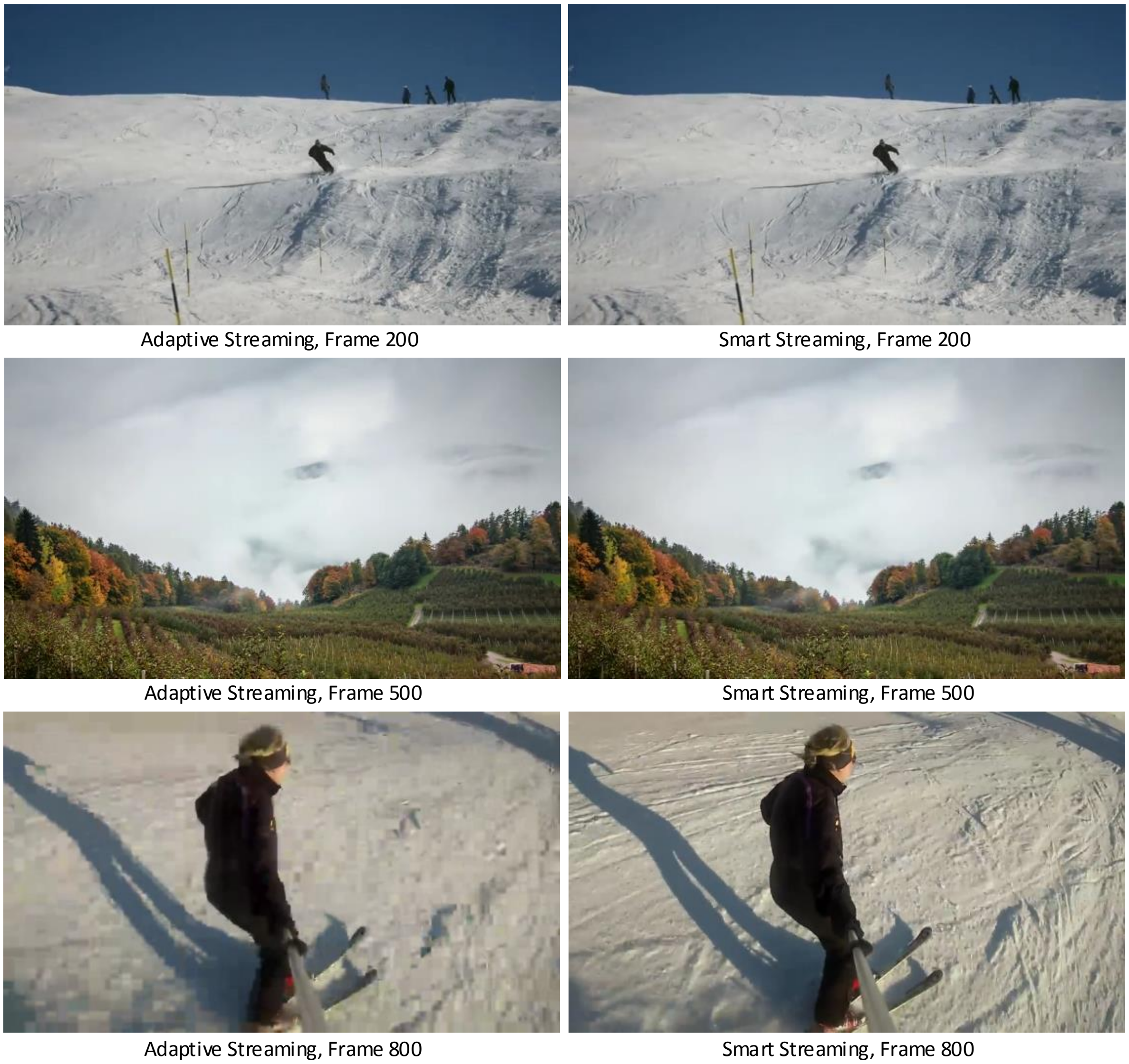
Real-world video streaming systems typically use bitrate as the key indicator of presentation quality. However, bitrate is an extremely poor measure of end consumer's visual QoE, which varies dramatically with the video content, display devices and viewing conditions. As a result, current systems often make subpar decisions that are far from optimal, which occur frequently in the popular adaptive bitrate streaming systems in the current OTT applications. To correct this, we need to redesign video delivery systems by employing a reliable objective presentation QoE predictor that is accurate, fast, easy-to-understand, and provides meaningful QoE predictions across device, resolution and content. In this paper, we recommend the newly proposed SSIMplus index (`https://ece.uwaterloo.ca/~z70wang/research/ssimplus/`) for this role, which overcomes many fundamental limitations in existing and state-of-the-art VQA models. Furthermore, we introduce the concept of QoE-driven smart streaming, where the decision making strategy in the adaptive streaming process at the receiver device takes into account QoE predictions such as SSIMplus scores as critical factors. We show that many *smarter* decisions can be made that may lead to a number of highly desirable benefits including reduced bandwidth, improved user QoE, smoother user QoE, and reduced probability of rebuffering or stalling events at end users' display devices.

The current paper focuses on OTT-based VoD applications. The core technologies, i.e., the SSIMplus QoE predictor and the QoE-driven smart streaming strategies, may find extensive usage in a much wider range of applications. For example, SSIMplus may be used to benchmark and optimize video encoders and transcoders, to guide the definition of video delivery system profiles, to monitor the quality of live and file-based video streams delivered over the networks, and to drive network resource allocations. The smart streaming idea may also be adapted to improve the switching decisions in live-streaming applications.

REFERENCES

[1] Stockhammer, T., "Dynamic adaptive streaming over HTTP: Standards and design principles," in [*Proceedings of the Second Annual ACM Conference on Multimedia Systems*], *MMSys '11*, 133–144, ACM, New York, NY, USA (2011).

[2] Wang, Z., Sheikh, H. R., and Bovik, A. C., "Objective video quality assessment," in [*The Handbook of Video Databases: Design and Applications*], Furht, B. and Marques, O., eds., 1041–1078, CRC Press (Sept. 2003).



Adaptive Streaming, Frame 200

Smart Streaming, Frame 200

Adaptive Streaming, Frame 500

Smart Streaming, Frame 500

Adaptive Streaming, Frame 800

Smart Streaming, Frame 800

Figure 15: Adaptive streaming vs. smart streaming: sample frames extracted from the received videos. Left: Frames 200, 500 and 800 from adaptive streaming video; Right: Frames 200, 500 and 800 from smart streaming video. The bitrate-driven adaptive streaming video maintains a similar bit-rate across content but results in unstable video quality (detected by the corresponding SSIMplus curve in Fig. 14). The QoE-driven smart streaming video better maintains perceptual quality across content (confirmed by the corresponding SSIMplus curve in Fig. 14).

[3] Wang, Z. and Bovik, A. C., [*Modern Image Quality Assessment*], Morgan & Claypool Publishers (Mar. 2006).

[4] Wang, Z. and Bovik, A., "Mean squared error: love it or leave it? - a new look at signal fidelity measures," *IEEE Signal Processing Magazine* **26**, 98–117 (Jan. 2009).

[5] Wang, Z. and Bovik, A. C., "A universal image quality index," *IEEE Signal Processing Letters* **9**, 81–84 (Mar. 2002).

[6] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing* **13**, 600–612 (Apr. 2004).

[7] Wang, Z., Lu, L., and Bovik, A. C., "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication,* special issue on objective video quality metrics **19**, 121–132 (Feb. 2004).

[8] Wang, Z., Simoncelli, E. P., and Bovik, A. C., "Multi-scale structural similarity for image quality assessment," in [*Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*], 1398–1402 (Nov. 2003).

[9] Pinson, M. H. and Wolf, S., "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting* **50**(3), 312–322 (2004).

[10] Seshadrinathan, K. and Bovik, A. C., "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Processing* **19**, 335–350 (Feb. 2010).

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

[12] Conviva Inc., *Viewer Experience Report* (2013).

[13] Symmetricom Inc., *Cable Operator Video Quality Study* (2008).

[14] Wang, Z. and Bovik, A. C., "Reduced- and no-reference visual quality assessment - the natural scene statistic model approach," *IEEE Signal Processing Magazine* **28**, 29–40 (Nov. 2011).

[15] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," tech. rep., International Telecommunication Union, Geneva, Switzerland (Apr. 2008).

[16] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," (Mar. 2002).

[17] Tektronix Inc., "Understanding PQR, DMOS and PSNR Measurements." `http://www.tek.com/document/fact-sheet/understanding-pqr-dmos-and-psnr-measurements/` (2014). [Online; accessed September 12, 2014].

[18] VideoClarity Inc., "Understanding MOS, JND and PSNR." `http://videoclarity.com/wpunderstandingjnddmospsnr/` (2014). [Online; accessed September 07, 2014].

[19] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," tech. rep., available at `http://www.vqeg.org/` (Apr 2000).

[20] Sheikh, H. R., Sabir, M., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing* **15**, 3440–3451 (Nov. 2006).

[21] Wang, Z., Zeng, K., and Rehman, A., "Method and system for smart adaptive video streaming driven by perceptual quality-of-experience estimations," in [*US Provisional Patent Application*], (Feb. 2015).