Parametric Quality Models for Multiscreen Video Systems

Nabajeet Barman∗, Rahul Vanam†, and Yuriy Reznik‡
∗Brightcove UK Ltd, London, United Kingdom, nbarman@brightcove.com
†Amazon Prime Video, Seattle, USA, rahulv@u.washington.edu
‡Brightcove Inc, Seattle, USA, yreznik@brightcove.com

***Abstract*—We propose simple parametric models for predicting visual quality scores on different devices in multiscreen systems. As input parameters, the proposed models take the distortion measure for the encoded video and parameters of viewing setup: the resolution of projected video, size of the display, and viewing distance. We derive models for the following distortion measures: PSNR, SSIM, VIF, and VMAF. We validate the proposed models using datasets corresponding to three different reproduction environments: standard TV sets, UltraHD TV sets, and mobiles. The obtained results confirm the improved accuracy of the prediction of MOS scores by the proposed techniques. The paper also includes introductory material explaining the usefulness of parametric quality for the analysis and optimizations of multiscreen video systems.**

Keywords—quality metrics, streaming, multiscreen systems

I. Introduction

A. Quality-related problems in multiscreen systems

Most modern-day streaming services deliver videos on many devices: TVs, PCs, tablets, mobiles, and others. Such devices have different screen characteristics and viewing setup parameters: typical viewing distance, viewing angle, ambient illuminance, etc. Consequently, the same videos encoded at the same resolutions and bitrates may appear differently on different devices, resulting in divergent MOS scores in the subjective tests [1]. Such differences are critical for understanding multiscreen systems' performance and posing related optimization problems.

To explain this more specifically, let us consider an HLS [2] or DASH [3]-based adaptive bitrate (ABR) streaming system, presented in Figure 1. In this system, an input video (mezzanine) encoded into n streams, with resolutions $W_1 \times H_1, \dots, W_n \times H_n$ and bitrates $R_1, \dots, R_n$, respectively. Such streams are stored on an origin server connected to a content delivery network (CDN). The origin server also receives an HLS or DASH manifest file describing the properties of such streams. On the receiving end, we may have k possible devices with different form factors and viewing setup parameters $\xi_1, \dots, \xi_k$. Such devices may then pull any combinations of segments/chunks from the encoded streams of the CDN as needed to facilitate continuous playback. In practice, the stream selection logic in players is typically driven by several considerations, such as device capabilities, network conditions, video player sizes, and others. Additional details about implementations of ABR systems can be found in [4-10].

In order to describe *codec-introduced distortions,* it is generally sufficient to measure n quantities: $D_1, \dots, D_n$ - distortion values (e.g., PSNR or SSIM [11]) for each stream. However, since in this system, the receiving devices are different, and we are ultimately interested in Quality of

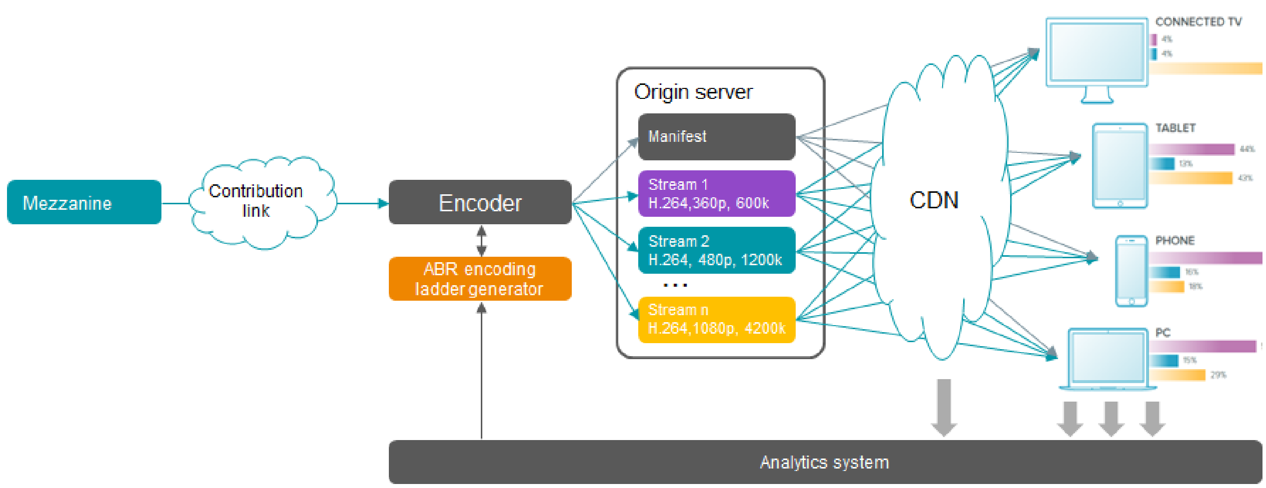

Fig. 1. Example of a multi-screen video delivery system using HTTP-based adaptrive bitrate (ABR) streaming framework.

[1]The work of on this paper was conducted while Rahul Vanam was affiliated with Brightcove, Inc.

Experience (QoE) and not just distortions, we have to consider *a matrix of* $k \times n$ *parameters*:

$$Q = \begin{bmatrix} Q_{11} & \cdots & Q_{1n} \\ \vdots & \ddots & \vdots \\ Q_{k1} & \cdots & Q_{kn} \end{bmatrix}$$

corresponding to MOS scores, measured for all streams and on all receiving devices. Having such a matrix, we can subsequently express the *average quality* achievable *for each device*:

$$\bar{Q}_i = \sum_{j=1}^{n} p_{ij} Q_{ij}, \;\; i = 1, \ldots, k$$

as well as the *average quality across all devices*:

$$\bar{Q} = \sum_{i=1}^{k} v_i \, \bar{Q}_i.$$

The weight factors p_{ij} and v_i in the above expressions correspond to *load probabilities* of each stream on each device, and *relative volume of content* pulled by each device, respectively.

By further noting that quality scores Q_{ij} are influenced by stream bitrate and resolution parameters as well as parameters of reproduction devices:

$$Q_{ij} = Q_{ij}\left(R_j, W_j, H_j, \xi_i\right),$$

we can pose several related optimization problems.

For example, we can pose a problem of the *design of encoding profiles* (choices of bitrates and resolutions for n streams [12,13]), such that the average quality delivered by the system (for given limits on average bitrates $\bar{R}_{\max,\mathrm{i}}$) is maximal:

$$\bar{Q}^* = \max_{\substack{R_1,\ldots,R_n,\, W_1,\ldots,W_n,\, H_1,\ldots,H_n \\ \sum_{j=1}^{n} p_{ij} R_i \leq \bar{R}_{\max,\mathrm{i}}, i=1,\ldots,k}} \sum_{i=1}^{k} v_i \sum_{j=1}^{n} p_{ij} Q_{ij}\left(R_j, W_j, H_j, \xi_i\right),$$

An alternative problem can also be posed by considering *relative quality gaps*:

$$\delta_{ij} = \frac{Q_i^* - Q_{ij}}{Q_i^*},$$

where Q_i^* denotes maximum quality achievable on i^{th} device:

$$Q_i^* = Q_{ij}\left(R_j \to \infty, W_j \to \infty, H_j \to \infty, \xi_i\right)$$

Using such relative scores, the problem can be posed as follows:

$$\bar{\delta}^* = \min_{\substack{R_1,\ldots,R_n,\, W_1,\ldots,W_n,\, H_1,\ldots,H_n \\ \sum_{j=1}^{n} p_{ij} R_i \leq \bar{R}_{\max,\mathrm{i}}, i=1,\ldots,k}} \max_i \sum_{j=1}^{n} p_{ij} \delta_{ij}\left(R_j, W_j, H_j, \xi_i\right).$$

In other words, we minimize the worst-case average quality gap across all devices. This setting allows the problem to be posed without exact knowledge of content usage distribution across devices, and without the risk of biasing the solution to benefit the most intensively used device while delivering suboptimal performance on the others.

However, as we can see, in both problem settings, the availability of quality estimates Q_{ij} specific for each device is essential. Examples of additional optimization problems utilizing quality estimates on different devices can be found in [14,15].

B. The problem addressed by this paper

In this paper, we propose a set of simple parametric quality models, allowing the matrix of per-device quality values $[Q_{ij}]$ to be easily computed by using a set of n distortion values $D_1, \ldots, D_n$ computed once for each stream, and k viewing setup parameters $\xi_1, \ldots, \xi_k$, corresponding to each category of receiving devices. In other words, what we propose are model functions:

$$Q_{ij} = Q\left(D_j, \xi_i\right), \qquad i = 1, \ldots, k, j = 1, \ldots, n.$$

These models effectively reduce the problem of assessment of quality on each device to a simple computation of distortions (e.g., using basic metrics such as PSNR or SSIM) for each stream and then applying a formula combining such distortion measures with other quality-influencing factors for each device to arrive at final predicted quality scores.

C. Benefits

The main benefit of the proposed models is the reduced complexity. Instead of computing quality scores many times and customarily for each stream and device parameters and using sophisticated QoE tools (such as e.g., Tektronix PQA [15,16]), the proposed approach allows the distortion measures to be computed only once per each stream and then reused for arriving at quality estimates for all devices.

The added benefit of the proposed approach is a simple mathematical form of coupling the distortion values with device-specific parameters. Such a connection could enable simplifications in related optimization problems.

The separation of distortion also allows simple modeling of the encoder's performance. Effectively, the design of models for quality-rate functions $Q_{ij}\left(R_j, W_j, H_j, \xi_i\right)$ is reduced to producing models of operational distortion-rate functions $D_j(R_j)$, which is a much simpler and better understood problem.

And finally, as we will show, based on our experimental results, the proposed approach allows high accuracy QOE prediction even when using very simple distortion metrics, such as PSNR or SSIM. Considering that such metrics are much simpler to compute than more sophisticated models and that they only need to be calculated at the encoded video resolutions and not in the upscaled (as displayed) domain, this makes the proposed approach even more appealing from the complexity and ease of use standpoints.

D. Related prior art

Many studies on subjective image quality assessment reveal the importance of physical parameters like image resolution, image/display size, and viewing distance [1, 18-24]. For example, Westerink and Roufs [18] have found that at a constant viewing distance, the subjective quality of still pictures was influenced independently by both the angular resolution and the size of the displayed image. P. Barten [19,20] confirmed the same effects using his SQRI method [20]. This model was also revisited recently by Barman et al. in [25]. They have found that it performs well on modern datasets and suggested a generalized form, extending the ranges of its parameters.

Much more sophisticated models, attempting to model different perceptual effects at optical, retinal, and visual cortex levels have also been proposed. They all naturally require parameters of viewing setups for calibrations. Good examples of such models include VDP [26], Sarnoff model [27], the model of Teo and Heeger [28], and others [29]. However, most of these models are complex and have found limited use in practical applications. Perhaps the best known nowadays is the Sarnoff model [27] due to its inclusion in Tektronix PQA analyzers [16,17].

Fast forwarding to modern practice, we notice that most modern objective quality metrics are purely distortion-based. Good examples are PSNR and SSIM [11]. They both compute summary statistics of differences between pixels in the original and decoded video. VIF [30] is a bit more sophisticated. It looks at logarithmic (information) differences, considering particular statistical models of visual information. But fundamentally, it still measures the differences between the original and reconstructed images. VMAF [31,32] is another popular modern-era metric, using SVM for turning VIF and several additional metrics into final quality scores. However, as explained in [31], it is trained by using a database [33] with DMOS scores and measured in a single viewing environment: ITU-R BT-500 [34] setup. In other words, by design, VMAF is meant to predict DMOS in one specific environment and not MOS scores across different platforms.

As it becomes evident from this review, most popular metrics used in today's practice are simply not suitable for analyzing relative quality effects in multiscreen systems. Therefore, this paper aims to bridge the gap between the broadly available metrics and applications that need device-specific quality assessments.

E. Outline

In Section II, we describe our proposed models and explain the reasoning behind them. In Section III, we conduct the experimental study, assessing the performance of the proposed models on datasets with MOS scores measured on three different categories of receiving devices. In Section IV, we offer conclusions.

II. Parametric Quality Models

A. Distortion metrics used as input

In this work, we will assume that the amount of codec-introduced noise (distortion) in each stream is measured by either PSNR, SSIM, VIF, or VMAF metrics. Average across frames values reported on the sequence level. We will further assume, that all metrics are computed at the resolutions of encoded videos, thus only characterizing the codec-added noise (distortion), and not scaling effects.

B. Viewing setup parameters

We will further assume that for each reproduction device, we may know its *resolution* $W_d \times H_d$ [in pixels], and *relative viewing distance* η, expressed in display heights. If instead of relative distance η, we know the absolute *viewing distance* d [in inches], then *display pixel density* ρ [in pixels per inch], or physical dimensions of the screen [in inches] will also be needed to construct the model.

To define our models, we will further use two *angular parameters* [18]:

- *viewing angle* ϕ [in degrees] – capturing the horizontal angular size of the video, as seen by the viewer on the screen, and
- *angular resolution* u [in cycles per degree] – expressing the highest possible spatial frequency in the video projected to the screen. Effectively, the Nyquist frequency of the video translated into cycles per degree.

We provide formulae connecting the above angular parameters to standard parameters of viewing setup in Table I.

TABLE I. Relationships Between Parameters of Viewing Setup

Parameter	Notation / formula	Units
Height of the encoded video	H	Pixels
Width of the encoded video	W	Pixels
Height of the displayed video	H_p	Pixels
Height of the displayed video	W_p	Pixels
Display pixel density	ρ	Pixels per inch
Viewing distance in inches	d	Inches
Viewing distance in display heights	$\eta = \frac{d\rho}{H_p}$	Display heights
Horizontal viewing angle	$\phi = 2\arctan\left(\frac{W_p}{2d\rho}\right) = 2\arctan\left(\frac{W_p}{2\eta H_p}\right)$	Degrees
Maximum horizontal spatial frequency (Nyquist) reproducible by the display	$u_d = \left(2\arctan\left(\frac{1}{d\rho}\right)\right)^{-1} = \left(2\arctan\left(\frac{1}{\eta H_p}\right)\right)^{-1}$	Cycles per degree
Maximum horizontal spatial frequency (angular resolution) that can be present in encoded video	$u = \left(2\arctan\left(\frac{W_p/W}{d\rho}\right)\right)^{-1} = \left(2\arctan\left(\frac{W_p/W}{\eta H_p}\right)\right)^{-1}$	Cycles per degree

C. Quality model based on viewing setup parameters

As a basic model of perceived quality based on the parameters of viewing setup, we will use the well-known model of J. Westerink and J. Roufs [18]. The specific formula that we will use in our work comes from [25]:

$$Q_{WR}(\phi, u) = \ln\left(a + b\left(1 + \left(\frac{\phi}{\phi_s}\right)^{-k}\right)^{-\frac{c}{k}}\left(1 + \left(\frac{u}{u_s}\right)^{-l}\right)^{-\frac{d}{l}}\right), \quad (1)$$

where $a = 2.718$, $b = 145.69$, $c = 1.55$, $d = 2.12$, $k = 6.01$, $l = 2.11$, $\phi_s = 35.0$, and $u_s = 16.93$ are the constants. This model predicts how the parameters of viewing setup (and specifically, viewing angle ϕ, and angular resolution u) impact the perceived quality of videos projected on the screen.

D. Quality models based on distortion measures

We are now ready to introduce parametric models proposed in this paper. In the most general form, they can be expressed as:

$$Q(D,\phi,u) = \alpha + \beta\big(1+\gamma Q_{WR}(\phi,u)\big)\, Q_D(D) + \delta Q_{WR}(\phi,u), \tag{2}$$

where $Q_{WR}(\phi,u)$ is a Westerink-Roufs model, $Q_D(D)$ is a fitting function for translating distortion scores to MOS domain, and where α, β, γ, and δ are the calibration constants.

We note that in a particular case when both parameters $\gamma = \delta = 0$, the model (1) turns into a direct mapping of distortion scores to MOS:

$$Q(D) = \alpha + \beta\, Q_D(D). \tag{3}$$

In a case when $\gamma = 0$, but $\delta > 0$, the proposed model (1) treats the impacts of distortion (D) and viewing factors (ϕ, u) as additive terms in the overall quality expression:

$$Q'(D,\phi,u) = \alpha + \beta\, Q_D(D) + \delta\, Q_{WR}(\phi,u).$$

Finally, when $\delta = 0$, and $\gamma \gg 1$, the model uses multiplicative coupling of effects of distortion and viewing-related factors:

$$Q''(D,\phi,u) \sim \alpha + \beta\gamma\, Q_{WR}(\phi,u)\cdot Q_D(D).$$

In other words, by using two fitting parameters γ and δ, these models can exploit both additive and multiplicative effects in combinations of distortion-based and viewing-based factors.

As in earlier studies [35,11.30], for translation of PSNR, SSIM, and VIF scores to MOS, we use the logistics function:

$$Q_D(D) = f(D,\varepsilon,\zeta) = \frac{1}{1+\exp(-\varepsilon(D-\zeta))}, \tag{4}$$

where ε defines the slope, and ζ represents the mid-point of this function. For translation of VMAF scores to MOS, we will use a linear model:

$$Q_D(D) = D, \tag{5}$$

since it is known that by design, VMAF is trained to operate at DMOS scale.

The full set of models and their parameters that we will be considering in this paper is presented in Table II.

TABLE II. PROPOSED QUALITY MODELS

Quality models	**Distortion metric D**	**Function $Q_D(D)$**	**Model formula**	**Model Parameters**
WR+PSRN2MOS	PSNR	$f(D,\varepsilon,\zeta)$	(2)	$\alpha,\beta,\gamma,\delta,\varepsilon,\zeta$
WR+SSIM2MOS	SSIM	$f(D,\varepsilon,\zeta)$	(2)	$\alpha,\beta,\gamma,\delta,\varepsilon,\zeta$
WR+VIF2MOS	VIF	$f(D,\varepsilon,\zeta)$	(2)	$\alpha,\beta,\gamma,\delta,\varepsilon,\zeta$
WR+VMAF2MOS	VMAF	D	(2)	$\alpha,\beta,\gamma,\delta$
PSNR2MOS	PSNR	$f(D,\varepsilon,\zeta)$	(3)	$\alpha,\beta,\varepsilon,\zeta$
SSIM2MOS	SSIM	$f(D,\varepsilon,\zeta)$	(3)	$\alpha,\beta,\varepsilon,\zeta$
VIF2MOS	VIF	$f(D,\varepsilon,\zeta)$	(3)	$\alpha,\beta,\varepsilon,\zeta$
VMAF2MOS	VMAF	D	(3)	α,β
xPSNR2MOS	↑ PSNR	$f(D,\varepsilon,\zeta)$	(3)	$\alpha,\beta,\varepsilon,\zeta$
xSSIM2MOS	↑ SSIM	$f(D,\varepsilon,\zeta)$	(3)	$\alpha,\beta,\varepsilon,\zeta$
xVIF2MOS	↑ VIF	$f(D,\varepsilon,\zeta)$	(3)	$\alpha,\beta,\varepsilon,\zeta$
xVMAF2MOS	↑ VMAF	D	(3)	α,β

The top 4 models in Table II are our main parametric models translating distortions and viewing factors to MOS scores. These models use the complete set of parameters present in formula (2). We name them as *"WR+<distortion>2MOS"*, where distortion metric can be either PSNR, SSIM, VIF, or VMAF. In all cases, the distortion metrics are assumed to be computed at encoded video resolutions.

The remaining models define direct mappings between distortion metrics and MOS scores. They are defined by using simplified model formula (3). We will use them for comparison purposes. There are 2 sets of such models. The first set uses names *"<distortion>2MOS"*, with all same distortion metrics and their meanings as in first 4 models. The second set uses names *"x<distortion>2MOS"*, where prefix "x" implies that distortion is computed not at resolution as encoded, but rather in upscaled form, where both the original video and decoded one are upscaled to match the anticipated resolution of the display, and then the distortion is computed by comparing such larger videos.

III. EXPERIMENTAL STUDY OF THE PROPOSED MODELS

A. Devices and datasets

The parameters of target devices and databases that we used in our work are summarized in Table III. These datasets come from several sources [31,36,37], and they all have been used extensively in studies on visual quality topics in the past.

TABLE III. DEVICES AND DATASETS USED IN THIS WORK

Parameter	**UHDTV**	**HDTV**	**Mobile**
Dataset name & reference	AVT-VQDB [36]	Netflix [31]	Mobile VQDB [37]
User devices	55” and 65" UHTVs	Consumer-grade TVs	6.39" smartphones
Viewing distance	1.5H	3H	3.67H
Display resolution	3840x2160	1920x1080	2340x1080
Display area used	Full screen	Full screen	1920x1080
Viewing angle [degrees]	61.3	33	27.2
Display Nyquist [cpd]	28.28	28.28	34.6
Resolutions of encoded videos [pixels→cpd]	480x360→4.71 1280x720→9.42 1920x1080→14.1 3840x2160→28.3	384x288→5.65 512x384→7.54 720x480→10.60 1280x720→18.85 1920x1080→28.3	1920x1080→34.6
Color space / dynamic range	BT.709 / SDR	BT.709 /SDR	BT.709 /SDR
Number of videos	30	70	128

As shown in Table III, our datasets include devices with a broad range of form factors: from 6" mobiles to 65" UHDTVs. The gap in relative viewing distances is also considerable: from 1.5 to 3.67 heights. The smallest encoded resolutions are 288p, the highest 2160p. All videos in datasets are progressive and use 16:9 DAR, BT.609 colors, and SDR transfer functions. The data sets include many files with a broad variety of visual content and artifacts introduced by transcoding them at different bitrates and resolutions.

For our study, we have used absolute MOS scores (not relative or DMOS scores) as reported in the datasets. However, we recomputed all objective distortion metrics for videos in datasets. This was done to ensure the consistency of all scores. To compute all metrics, we used FFmpeg tool [38]. In the computation of scaled metrics, the Lanczos3 filter was used for upscaling. For PSNR and SSIM, we have used Y-PSNR and Y-SSIM metrics as reported by FFmpeg. For VIF, we used the averages of 4 VIF scores, as reported by FFmpeg. For the computation of VMAF we used libvmaf v2.3.0, integrated into the FFmpeg. Both upscaled and non-scaled versions of all 4 metrics have been computed.

B. Fitting of models to dataset MOS scores

For fitting our models, a combined pool of data points was created, including all data points from Mobile data set, plus 2x replicated points from HDTV dataset, and 4x replicated data points from UHDTV dataset. These replication factors have been chosen to ensure that we have a balanced representation of data from each category of receiving devices in the combined set. For finding model parameters $(\alpha, \beta, \ldots, \zeta)$ we have used MAPLE computer algebra tool [39], and its *Minimize* function. The sum of square differences between model-predicted and actual MOS scores in the combined pool was used as an objective function for minimization.

C. The results

The values of optimal parameters found for each of our models are summarized in Table IV.

TABLE IV. MODEL PARAMETERS

Quality models	α	β	γ	δ	ε	ζ
WR+PSRN2MOS	-6.906	6.130	-0.048	1.476	0.228	23.83
WR+SSIM2MOS	-7.181	7.662	-0.089	1.753	7.492	0.777
WR+VIF2MOS	-12.09	12.117	-0.137	2.763	4.846	0.416
WR+VMAF2MOS	-7.682	0.0753	-0.122	2.01		
PSNR2MOS	0	3.86			0.216	23.49
SSIM2MOS	1.106	2.863			11.751	0.789
VIF2MOS	0.831	2.941			8.124	0.408
VMAF2MOS	1.164	0.0286				
xPSNR2MOS	0	4.14			0.212	25.38
xSSIM2MOS	0	6.414			4.963	0.865
xVIF2MOS	0.305	5.461			4.127	0.598
xVMAF2MOS	0.523	0.0428				

The resulting RMSE accuracy values are provided in Table V. These values are provided separately for each dataset. The compound average RMSE values across all sets are also shown. The lowest values in each category are marked in bold.

D. Discussion

Based on results presented in Table V, we observe that the proposed combined distortion + viewing factors metrics achieve much higher accuracy of matching MOS scores than metrics that rely on distortion alone. The differences are particularly dramatic for simple metrics, such as PSNR and SSIM. E.g., we see that WR+PSRN2MOS reports 0.466 overall RMSE, while PSNR2MOS reports 0.908. Similarly, WR+SSIM2MOS reports 0.441, while SSIM2MOS reports 0.893.

TABLE V. RMSE ACCURACY RESULTS ACROSS DATASETS

Quality models	**UHDTV**	**HDTV**	**Mobile**	**Average**
WR+PSRN2MOS	0.384	0.396	0.618	0.466
WR+SSIM2MOS	0.369	0.423	0.532	0.441
WR+VIF2MOS	0.356	0.389	**0.362**	0.369
WR+VMAF2MOS	0.303	**0.375**	0.382	**0.354**
PSNR2MOS	0.886	1.105	0.734	0.908
SSIM2MOS	0.891	1.139	0.649	0.893
VIF2MOS	0.901	1.170	0.583	0.885
VMAF2MOS	0.830	1.103	0.629	0.854
xPSNR2MOS	0.859	0.884	0.710	0.818
xSSIM2MOS	0.839	0.842	0.592	0.757
xVIF2MOS	0.418	0.739	0.470	0.543
xVMAF2MOS	**0.286**	0.422	0.396	0.364

We also note that our proposed metrics *outperform metrics relying on upscaling videos*. As expected, upscaling shows somewhat improved performance, but the combination of non-scaled metrics using our parametric approach seems to work even better. This can be easily followed noting that RMSE for WR+PSRN2MOS ~ 0.466 vs xPSNR2MOS ~ 0.818, or that for WR+SSIM2MOS ~ 0.441 vs xSSIM2MOS ~ 0.757.

However, the most surprising is that all the above improvements seem to stay in place not only for simple metrics such as PSNR and SSIM, but also for much more complex ones – VIF and VMAF, which both exploit multi-scale processing, and complex statistical models. We notice that WR+VIF2MOS achieves RMSE of 0.369, easily beating upscaled VIF: xVIF2MOS ~ 0.543, and that WR+VMAF2MOS, computed without upscaling, achieves an overall RMSE of 0.354, which is also better than overall RMSE of upscaled VMAF xVMAF2MOS ~0.364.

The only case in which the regular upscaled VMAF has performed better than our WR+VMAF2MOS, was in fit to UHDTV dataset. It seems that this device's characteristics are closer to the training dataset used in the design phase of VMAF. But then the performance of scaled VMAF on other devices is worse, as well it is slightly worse overall.

These findings confirm the merits of our proposed technique of incorporating parameters of viewing setups for improved predictions of MOS scores on different devices.

IV. CONCLUSIONS

Several parametric models have been proposed for predicting visual quality scores on different devices. These models use distortion metrics (PSNR, SSIM, VIF, or VMAF) and viewing setup characteristics as input parameters. The proposed models have been validated using datasets with MOS scores measured on TV sets, UltraHD TV sets, and mobile devices. The obtained results confirm the improved accuracy of the prediction of MOS scores on different devices. Other practical benefits and applications of the proposed techniques are also discussed.

REFERENCES

[1] A. Catellier, M. Pinson, W. Ingram, and A. Webster, "Impact of mobile devices and usage location on perceived multimedia quality," in Proc. International Workshop on Quality of Multimedia Experience (QoMEX), 2012, pp. 39-44.

[2] “HTTP live streaming, RFC 8216,” https://tools.ietf.org/html/rfc8216, 2019, [Online: accessed 19-March-2022].

[3] “ISO/IEC 23009-1:2019 Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats,” https://www.iso.org/standard/79329.html, 2019, [Online: accessed 17-March-2022].

[4] B. Girod, M. Kalman, Y.J. Liang, and R. Zhang, "Advances in channel-adaptive video streaming," Wireless Comm. and Mobile Comp., vol. 2, no. 6, pp. 573-584, 2002.

[5] D. Wu, Y.T. Hou, W. Zhu, Y-Q. Zhang, and JM Peha, "Streaming video over the internet: approaches and directions," IEEE Trans. CSVT, vol. 11, no. 3, pp. 282-300, 2001.

[6] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, "Video coding for streaming media delivery on the internet," IEEE Trans. CSVT, vol. 11, no. 3, pp. 269-281, 2001.

[7] D. Lee, C. Dovrolis, A. Begen, "Caching in HTTP Adaptive Streaming: Friend or Foe?," in Proc. ACM Network and Operating System Support on Digital Audio and Video Workshop, 2014, pp. 31-36.

[8] S. Hesse, "Design of scheduling and rate-adaptation algorithms for adaptive HTTP streaming," in Proc. SPIE 8856, Applications of Digital Image Processing XXXVI, 88560M, 2013.

[9] C. Zhou, X. Zhang, L. Huo, and Z. Guo, "A control-theoretic approach to rate adaptation for dynamic HTTP streaming," in Proc. Visual Comm. Image Processing, San Diego, CA, 2012, pp. 1-6.

[10] K. Spiteri, R. Urgaonkar, R. K. Sitaraman, BOLA: Near-Optimal Bitrate Adaptation for Online Videos. IEEE/ACM Trans. Netw. 28(4): 1698-1711 (2020)

[11] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: from error visibility to structural similarity". IEEE Transactions on Image Processing 13 (4): 600–612 (2004-04-01).

[12] Y. Reznik, K. O. Lillevold, A. Jagannath, J. Greer, and J. Corley, "Optimal design of encoding profiles for ABR streaming," in Proc. Packet Video Workshop, Amsterdam, NL, June 12, 2018, pp. 43-47.

[13] C. Chen, Y. Lin, S. Benting, and A. Kokaram, "Optimized transcoding for large scale adaptive streaming using playback statistics," in Proc. IEEE Int. Conf. Image Proc., Oct 2018, pp. 3269-3273.

[14] Y. Reznik, K. Lillevold, R. Vanam, "Perceptually optimized ABR ladder generation for Web streaming," *Proc. IS&T Electronic Imaging*, San Francisco, CA, January 18-21, 2021.

[15] Y. A. Reznik, "Average Performance of Adaptive Streaming," in Proc. IEEE Data Compression Conference (DCC), Snowbird, UT, March 21 - 24, 2021, pp. 263–272.

[16] Tektronix, "Understanding PQR, DMOS and PSNR Measurements," https://www.tek.com/en/documents/fact-sheet/understanding-pqr-dmos-and-psnr-measurements

[17] Tektronix PQA600 Picture Quality Analyzer, User Manual, https://download.tek.com/manual/PQA600C-and-PQASW-Picture-Quality-Analyzer-User-Manual-077113700.pdf

[18] J. Westerink and J. Roufs, "Subjective image quality as a function of viewing distance resolution and picture size," SMPTE Journal, vol. 98, 1989, pp. 113-19.

[19] A. Lund, "The influence of video image size and resolution on Viewing-Distance preferences " SMPTE Journal, vol. 102, 1993, pp. 406-15.

[20] P. G. J. Barten, "Effect of picture size and definition on perceived image quality," IEEE Transactions on Electron Devices vol. 36 no. 9 Part 2 pp. 1865-9, 1989.

[21] P. G. J. Barten, "Contrast Sensitivity of the Human Eye and Its Effects on Image Quality," SPIE Press, 1999.

[22] L. Jesty, "The relation between picture size viewing distance and picture quality," Proc. IEE, 1958, pp. 425-39.

[23] R. Fish and T. Judd "A subjective visual comparison of NTSC VHS and Compressed DS1-Compatible video " Proc. of the Society for Information Display (SID) vol. 32 no. 2 pp. 157-164 1991.

[24] H. Knoche J. D. McCarthy and M. A. Sasse "Can small be beautiful? assessing image resolution requirements for mobile TV," in Proc. ACM Multimedia, 2005, pp. 829-38.

[25] N.Barman, R. Vanam, Y. Reznik, "Generalized Westerink-Roufs Model for Predicting Quality of Scaled Video", in Proc. International Conference on Quality of Multimedia Experience (QoMEX), Lippstadt, Germany, Sept 5-7, 2022.

[26] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity", in Digital Images and Human Vision (A. B. Watson, ed.), pp. 179-206, Cambridge, MA: The MIT Press, 1993.

[27] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance", in Digital Images and Human Vision (A. B. Watson, ed.), pp. 163-178, Cambridge, MA: The MIT Press, 1993.

[28] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in Proc. ICIP-94, vol. II, (Austin, TX), pp. 982-986, Nov. 1994.

[29] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in Handbook of Image and Video Proc. (A. Bovik, ed.), Academic Press, 2000.

[30] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," in IEEE Transactions on Image Processing, vol. 15, no. 2, pp. 430-444, Feb. 2006.

[31] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," The Netflix Tech Blog, vol. 6, p. 2, 2016. https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652 [Online: accessed 17-March-2022].

[32] Z. Li, Netflix dataset: https://drive.google.com/drive/folders/0B3YWNICYMBIweGdJbERlUG9zc0k

[33] Netflix, "VMAF - Video Multi-Method Assessment Fusion," https://github.com/Netflix/vmaf, [Online: accessed 12-May-2022].

[34] ITU-R BT-500, "BT.500: Methodologies for the subjective assessment of the quality of television images", October 2019.

[35] U. Engelke, M. Kusuma, H.-J. Zepernick, M. Caldera, “Reduced-reference metric design for objective perceptual quality assessment in wireless imaging,” Signal Processing: Image Communication, Volume 24, Issue 7, 2009, Pages 525-547.

[36] R. R. Ramachandra Rao, S. Göring, W. Robitza, B. Feiten and A. Raake, "AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1," 2019 IEEE International Symposium on Multimedia (ISM), 2019, pp. 17-177

[37] Y. Zhang, Z. Liu, W. Wu, R. Yao, Z. Chen, and S. Liu, "A Subjective Quality Assessment Database for Mobile Video Coding," 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2020, pp. 225-228.

[38] FFmpeg, https://ffmpeg.org/download.html [Online: accessed 17-March-2022].

[39] MAPLE, https://www.maplesoft.com/products/Maple/ [Online: accessed 17-March-2022].