Presented at: IS&T/SPIE Annual Symposium on Electronic Imaging, San Francisco, CA, Feb. 9-12, 2015
Published in: Human Vision and Electronic Imaging XX, Proc. SPIE, vol. 9394.

Display Device-Adapted Video Quality-of-Experience Assessment

Abdul Rehman, Kai Zeng, and Zhou Wang

Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada

ABSTRACT

Today's viewers consume video content from a variety of connected devices, including smart phones, tablets, notebooks, TVs, and PCs. This imposes significant challenges for managing video traffic efficiently to ensure an acceptable quality-of-experience (QoE) for the end users as the perceptual quality of video content strongly depends on the properties of the display device and the viewing conditions. State-of-the-art full-reference objective video quality assessment algorithms do not take into account the combined impact of display device properties, viewing conditions, and video resolution while performing video quality assessment. We performed a subjective study in order to understand the impact of aforementioned factors on perceptual video QoE. We also propose a full-reference video QoE measure, named SSIMplus (https://ece.uwaterloo.ca/~z70wang/research/ssimplus/), that provides real-time prediction of the perceptual quality of a video based on human visual system behaviors, video content characteristics (such as spatial and temporal complexity, and video resolution), display device properties (such as screen size, resolution, and brightness), and viewing conditions (such as viewing distance and angle). Experimental results have shown that the proposed algorithm outperforms state-of-the-art video quality measures in terms of accuracy and speed.

Keywords: video quality of experience, subjective study, display device, viewing coniditions, video quality assessment, video monetization

1. INTRODUCTION

Over the past years, we have observed an exponential increase in the demand for video services. Video data dominates Internet video traffic and is predicted to increase much faster than other media types in the years to come. Cisco predicts that video data will account for 79% of Internet traffic by 2018 and mobile video will represent two-thirds of all mobile data traffic by 2018.[1] Well accustomed to a variety of multimedia devices, consumers want a flexible digital lifestyle in which high-quality multimedia content follows them wherever they go and on whatever device they use. This imposes significant challenges for managing video traffic efficiently to ensure an acceptable quality-of-experience (QoE) for the end user, as the perceptual quality of video content strongly depends on the properties of the display device and the viewing conditions. Network throughput based video adaptation, without considering a user's QoE, could result in poor video QoE or wastage of bandwidth. Consequently, QoE management under cost constraints is the key to satisfying consumers and video monetization services.

Digital videos are subject to a wide variety of distortions during acquisition, processing, compression, storage, transmission, reproduction, and display, any of which may result in degradation of visual quality. For applications in which videos are ultimately to be viewed by human beings, the only "correct" method of quantifying visual image quality is through subjective evaluation. In practice, however, subjective evaluation is usually too inconvenient, time-consuming and expensive. Objective video quality assessment (VQA) methods may automatically predict the quality assessment behaviors of humans viewing the video signals. VQA methods have broad applications 1) in the evaluations and comparisons of the quality of videos and the performance of different video acquisition, processing, compression, storage, transmission, reproduction, and display methods and systems; 2)

Further author information: (Send correspondence to Abdul Rehman)
Abdul Rehman: E-mail: abdul.rehman@uwaterloo.ca
Kai Zeng: E-mail: kzeng@uwaterloo.ca
Zhou Wang: E-mail: zhou.wang@uwaterloo.ca

in the control, maintenance, streaming, and resource allocation of visual communication systems; and 3) in the design and optimization of video acquisition, processing, compression, storage, transmission, reproduction, and display methods and systems.

The simplest and most widely used VQA measure is the mean squared error (MSE), computed by averaging the squared intensity differences of distorted and reference image pixels, along with the related quantity of peak signal-to-noise ratio (PSNR). The MSE and PSNR are simple to calculate and are mathematically convenient in the context of optimization, but they are not very well matched to perceived visual quality.[2] State-of-the-art VQA methods include the structural similarity index (SSIM),[3,4] the multi-scale structural similarity index (MS-SSIM),[5] the video quality metric (VQM),[6] and the motion-based video integrity evaluation index (MOVIE).[7] All of them have achieved better quality prediction performance than MSE/PSNR. Among them, the best trade-off of quality prediction performance and computational cost is obtained by SSIM and MS-SSIM.[8] However, none of the aforementioned VQA methods consider the differences between the viewing devices of the end-users, which are an important factor of the visual QoE of the end users. For example, the human quality assessment of the same video can be significantly different when it is displayed on different viewing devices, such as HDTV, digital TV, projectors, desktop PCs, laptop PCs, tablets, and smart phones, and many more. Existing VQA methods ignore such differences and do not contain adaptive frameworks and mechanisms that can adjust themselves to the changes of viewing device parameters. Moreover, the quality analysis information provided by existing methods is limited. For example, VQM and MOVIE do not supply spatially and temporally localized quality maps, SSIM does not produce quality maps at different scales, and SSIM and MS-SSIM do not take into account temporal distortions.

One of the most challenging problems that needs to be addressed to enable video QoE management is the lack of objective VQA measures that predict perceptual video QoE based on viewing conditions across multiple devices.[9,10] There is a lack of publicly available subject-rated video quality assessment databases that investigate the impact on perceptual video quality under the interaction of display device properties, viewing conditions, and video resolution. In this work, we performed a subjective study in order to collect subject-rated data representing the perceptual quality of selected video content in different resolutions viewed on various display devices under varying viewing conditions. A set of raw video sequences, consisting of 1920×1080 and 1136×640 resolutions, was compressed at five distortion levels to obtain bitstreams compliant to H.264 video compression standard. The distorted video sequences were scored by subjects under different viewing conditions on four popular display devices.

In this paper, we propose a full-reference video QoE measure, SSIMplus, that provides real-time prediction of the perceptual quality of a video based on human visual system behaviors, video content characteristics (such as spatial and temporal complexity, and video resolution), display device properties (such as screen size, resolution, and brightness), and viewing conditions (such as viewing distance and angle). We compared the performance of the proposed algorithm to the most popular and widely used FR-VQA measures that include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity[3] (SSIM), Multi-scale Structural Similarity[5] (MS-SSIM), MOtion-based Video Integrity Evaluation[7] (MOVIE), Video Quality Metric[6] (VQM), Picture Quality Ratings (PQR),[11] and Just Noticeable Difference (JND).[12] Experimental results have shown that the proposed algorithm adapts to the properties of the display devices and changes in the viewing conditions significantly better than the state-of-the-art video quality measures under comparison. Additionally, the proposed video QoE algorithm is considerably faster than the aforementioned perceptual VQA measures and fulfills the need for real-time computation of an accurate perceptual video QoE index and a detailed quality map.

2. SUBJECTIVE STUDY

2.1 Video Database

We used four uncompressed videos of natural scenes as source video sequences that contain indoor and outdoor scenes, flat areas and complex patterns, camera zooming/panning and object motion towards different directions. The digital videos are provided in uncompressed YUV 4:2:0 format (which guarantees that the reference videos are distortion free) and do not contain audio. We only used progressively scanned videos to avoid problems associated with video de-interlacing. The reference video sequences were captured in Full HD 1920×1080

resolution. In order to study the impact of varying video conditions on popular display devices, we also included the reference video sequences in the resolution 1136×640, which were created by down-sampling Full HD video sequences frame by frame. All the video sequences are ten seconds long when played at a frame rate of 24 frames per second. We compressed every reference video sequence at five quality levels using x264,[13] a popular software library for encoding video streams into the H.264/MPEG-4 AVC compression format.

2.2 Subjective Test

Our subjective test generally follows the Absolute Category Rating (ACR) methodology, as suggested by ITU-T recommendation P.910.[14] Although SSCQE[14] is designed for continuously tracking instantaneous video quality over time, it is not adopted in our experiment for the following reasons. First, in practice human subjects often opt to judge video quality on per segment basis, discounting the instantaneous quality variations between frames within a scene. Second, in our database, the same coding configuration and parameters are applied to the full duration of each scene, which is also roughly constant in terms of content and complexity. As a result, a single quality score is sufficient to summarize its quality. Third, in SSCQE, there is time delay between the recorded instantaneous quality and the video content, and such delay varies between subjects and is also a function of slider "stiffness". This is an unresolved issue of the general SSCQE methodology, but is avoided when only a single score is acquired. We believe that ACR is much simpler compared to SSCQE and provides more reliable and more realistic quality evaluations under the conditions employed in the subjective study.

Thirty naïve subjects - all university undergraduate and graduate students - 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 were no bias or significant difference between the scores obtained, for the same set of video sequences, in the beginning and at the end of the test. The test video sequences were scored by subjects under the viewing conditions provided in Table 1. Instructions were given to the subjects in both written and oral forms. A training session preceded the test where the subject was shown examples of distorted video sequences expected in the test.

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

In our study, six out of the thirty subjects were found to be outliers and their scores were discarded.[15] After each test session, we talked to subjects about their perceptual quality rating experience on different devices. This step did not affect the data that had been collected, but helped us understand the data better, and also provided us with intuitive ideas that could be employed in the development of computational models that mimic human behaviors.

3. OBJECTIVE MODEL

We propose a novel VQA measure called SSIMplus that has been designed to take into account properties of the human visual system, video content, and viewing conditions. The VQA measure provides straightforward predictions on what an average consumer says about the quality of the video content being delivered on a scale of 0-100 and also categories the quality as either bad, poor, fair, good or excellent. The underlying algorithm uses an advanced perceptual model that allows the VQA measure to adapt video QoE analysis to any display device and viewing conditions.

As the first step towards video QoE assessment, a multi-scale transform on the reference and distorted video frames is performed that decomposes a video frame into multiple scales, each associated with a different frequency range. Subsequently, the quality maps of each scale are computed based on a structure comparison between subsequent reference and distorted scales. Afterwards, the quality of all the scales is determined by performing spatial pooling of the quality maps based on the local information content and distortion. The perceptual quality of the distorted frame is calculated using a weighted combination of the scale-wise quality values. The weights are determined using a method that takes into account the properties of the display device and viewing conditions. The perceptual quality of video content depends on the sampling density of the signal, the viewing conditions, the display device, and the perceptual capability of the observer's visual system. In practice, the subjective evaluation of a given video varies when these factors vary. The contrast perception capability of the human visual system depends strongly on the spatial or spatio-temporal frequency of a visual signal, which is modeled using a function called the contrast sensitivity function (CSF). One may use one or a combination of the following device and viewing parameters to determine the contrast sensitivity of the human visual system: 1) average or range of user viewing distance, 2) sizes of viewing window and screen; 3) screen resolution; 4) video scaling; 5) screen contrast; 6) replay temporal resolution; 7) illumination condition of the viewing environment; 8) viewing angle; 9) viewing window resolution; 10) post-filtering and image resizing methods; 11) device model; 12) screen gamma correction parameter; 13) video scan type (interlaced or progressive). These parameters are used to determine the sensitivity of the human visual system to the individual scales of the input video signals. Subsequently, the sensitivity values are normalized to determine the weight/importance of the scales.

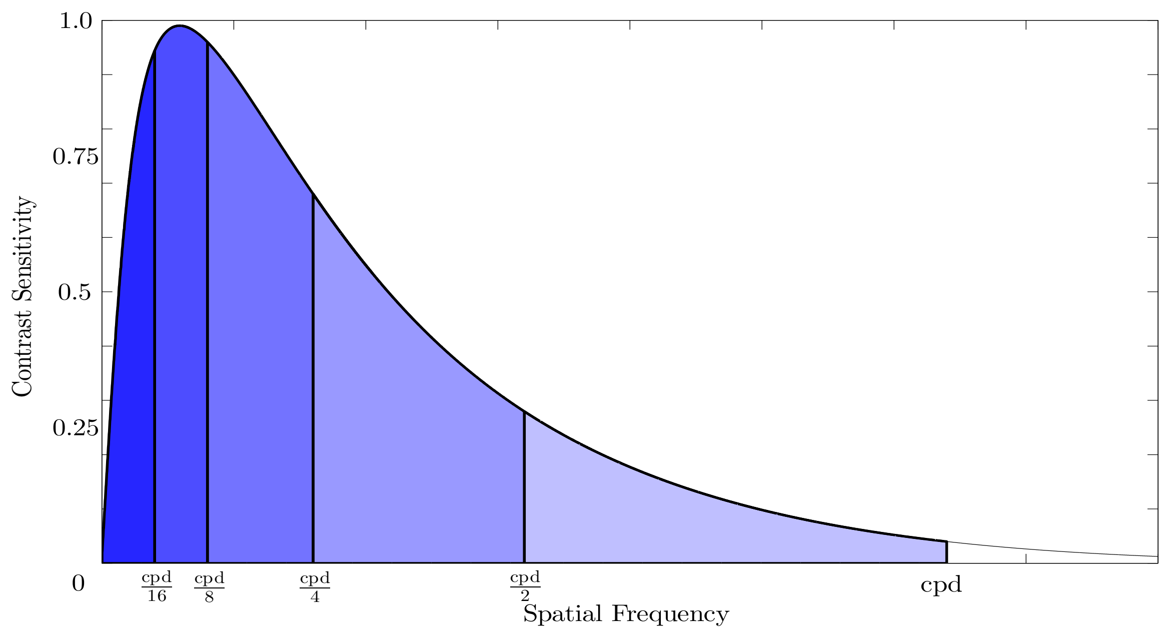

Figure 1: Device and viewing condition-dependent parameters based multi-scale weights calculation scheme

The parameters or a subset of the parameters of viewing window/screen size, device screen resolution, replay temporal resolution, viewing distance, device screen contrast, viewing angle, and viewing window resolution, are converted into a viewing resolution factor in the unit of the number of pixels per degree of visual angle. These parameters are also used to compute the CSF of the human visual system. The viewing resolution factor is subsequently used to determine the frequency covering range of each scale in the multi-resolution transform. The frequency covering ranges of all scales in the multi-resolution transform divide the full CSF into multiple regions, each corresponds to one scale. A weighting factor of each scale is then determined by calculating the area under the CSF function within the frequency covering range of that scale. Since the viewing resolution factor and the CSF computation depend on device parameters and viewing conditions, the frequency covering ranges and subsequently the weighting factor of each scale are also device and viewing condition dependent, which is an important factor that differentiates the proposed method from existing approaches. These device and viewing condition-dependent parameters are used to determine the importance of each scale in the overall quality evaluation of the image or video signal. Figure 1 shows an example of the details of device and viewing condition-dependent parameters based multi-scale weights calculation scheme. In Figure 1 cpd represent cycles per degree of visual angle which is determined by the viewing resolution factor. The frequency covering ranges of the scales in the multi-resolution transform, starting from the finest scale, are between cpd/2 and cpd, cpd/4 and cpd/2, cpd/8 and cpd/4, . . . , respectively. The integrals of the CSF curve under the respective frequency covering range are computed dynamically and used to determine the weighting factor and thus the visual importance of the corresponding scale. Subsequently, these weights are used to pool the scale-wise scores to determine frame-level and sequence-level QoE scores.

4. PERFORMANCE COMPARISON

This section compares the performance of the SSIMplus algorithm to the following most popular and widely used video quality assessment measures in academia & industry:

- **Peak Signal-to-Noise Ratio (PSNR)** is a simple function of the Mean Squared Error (MSE) between the reference and test video sequences;
- **Structural Similarity (SSIM) Index**[3] is a popular method for quality assessment of still images. The SSIM index was applied frame-by-frame on the luminance component of the video and the overall SSIM index for the video was computed as the average of the frame level quality scores;
- **Multi-Scale Structural Similarity (MS-SSIM) Index**[5] extends the single-scale SSIM Index towards incorporating the variations across scales;
- **Video Quality Metric (VQM)**[6] is a VQA algorithm developed at the National Telecommunications and Information Administration (NTIA). Due to its excellent performance in the VQEG Phase 2 validation tests, the VQM methods were adopted by the American National Standards Institute (ANSI) as a national standard, and as International Telecommunications Union Recommendations (ITU-T J.144 and ITU-R BT.1683, both adopted in 2004);
- **MOtion-based Video Integrity Evaluation (MOVIE) Index**[7] is a VQA index that was recently developed at the Laboratory for Image and Video Engineering (LIVE), University of Texas at Austin;
- **Picture Quality Rating (PQR-Tek)**[11] , a Just Noticeable Difference (JND) based VQA measure, was introduced on the Tektronix PQA200 Picture Quality Analyzer and was offered on its successor, the PQA300;
- **DMOS measure by Tektronix (DMOS-Tek)**[11] is a non-linear mapping of PQR-Tek provided to predict Difference Mean Opinion Score (DMOS) values for test videos;
- **JND measure by Video Clarity (JND-VC)**[12] is an implementation of the JND index by Sarnoff Corporation; and
- **DMOS measure by Video Clarity (DMOS-VC)**[12] is a non-linear mapping of MS-SSIM provided to predict Difference Mean Opinion Score (DMOS) values for test videos;.

4.1 Perceptual quality prediction accuracy

The ultimate goal of VQA algorithms is to predict subjective quality evaluation of a video. Therefore, the most important test is to evaluate how well they predict subjective scores. Recently, a subjective study was conducted by JCT-VC members to quantify the rate-distortion gain of the HEVC codec against a similarly configured H.264/AVC codec.[16] The database is very relevant for evaluation of video quality assessment algorithms developed for media & entertainment industry because it contains videos distorted by most commonly used video compression standard along with the recently developed H.265 codec.[17] We use this independent and challenging subjective database to compare the performance of the VQA algorithms in predicting the perceptual quality. The performance comparison results are provide in Table 2. For this purpose, we employ five evaluation metrics to assess the performance of VQA measures:

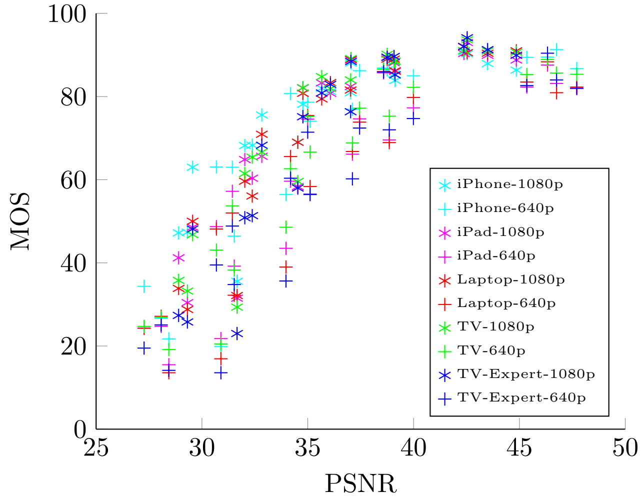

Figure 2: MOS versus PSNR scatter plot

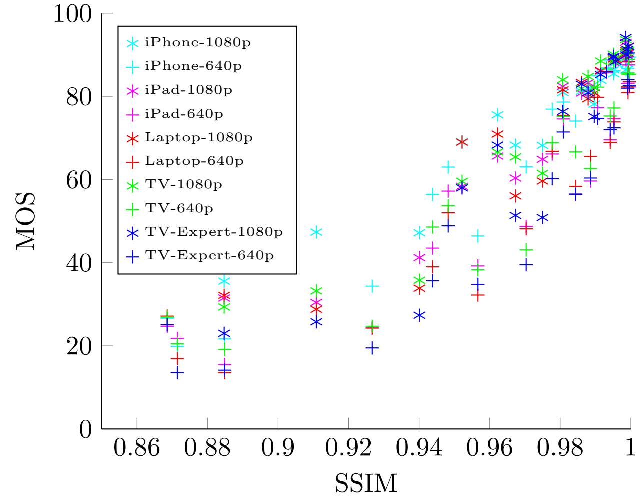

Figure 3: MOS versus SSIM scatter plot

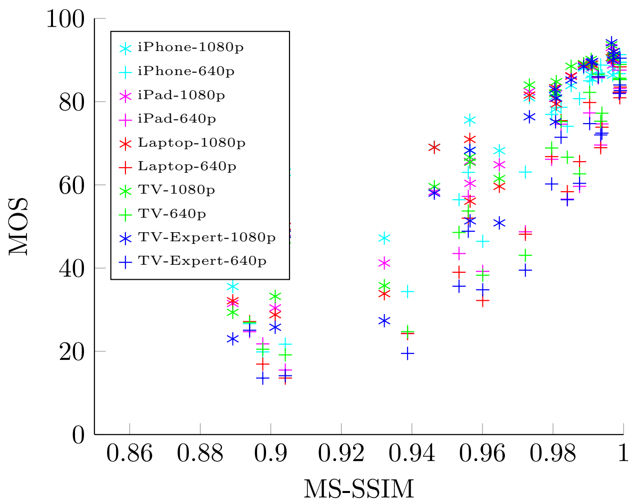

Figure 4: MOS versus MS-SSIM scatter plot

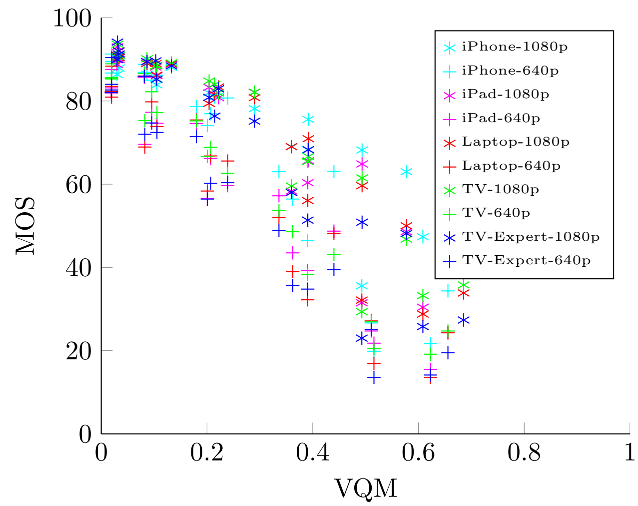

Figure 5: MOS versus VQM scatter plot

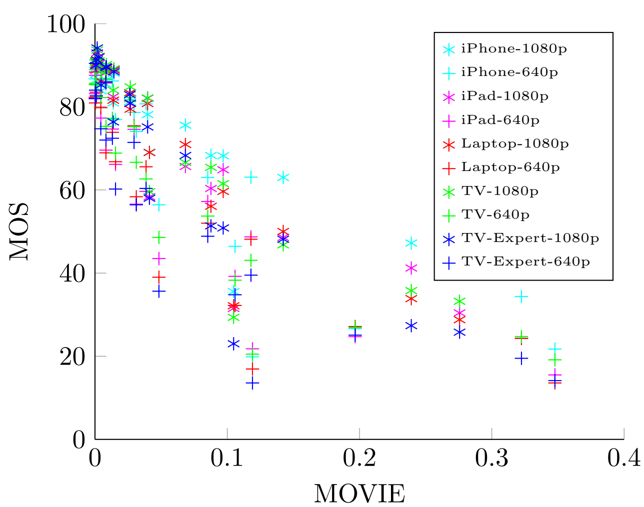

Figure 6: MOS versus MOVIE scatter plot

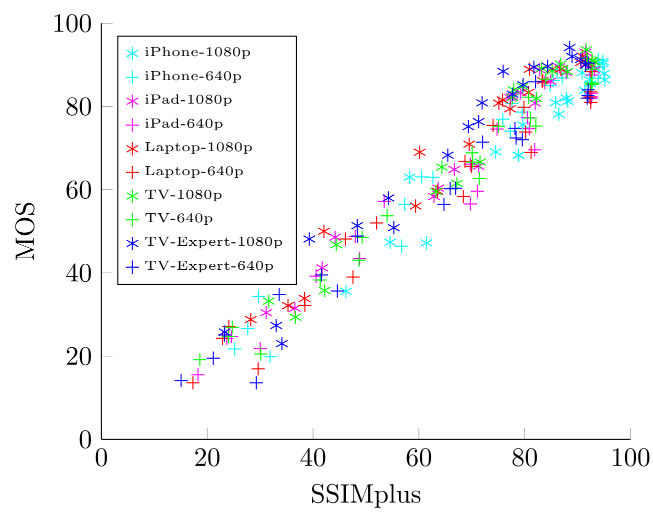

Figure 7: MOS versus SSIMplus scatter plot

- Pearson linear correlation coefficient (PLCC) after a nonlinear mapping between the subjective and objective scores. For the i-th image in an image database of size N, given its subjective score o_i (mean opinion score (MOS) or difference of MOS (DMOS) between reference and distorted images) and its raw objective score r_i, we first apply a nonlinear function to r_i given by[18]

$$q(r) = a_1 \left\{ \frac{1}{2} - \frac{1}{1 + \exp\left[a_2(r - a_3)\right]} \right\} + a_4 r + a_5 \tag{1}$$

  where a_1 to a_5 are model parameters found numerically using a nonlinear regression process to maximize the correlations between subjective and objective scores. The PLCC value can then be computed as

$$\text{PLCC} = \frac{\sum_i (q_i - \bar{q}) * (o_i - \bar{o})}{\sqrt{\sum_i (q_i - \bar{q})^2 * \sum_i (o_i - \bar{o})^2}}. \tag{2}$$

- Mean absolute error (MAE) is calculated using the converted objective scores after the nonlinear mapping described above:

$$\text{MAE} = \frac{1}{N} \sum |q_i - o_i|. \tag{3}$$

- Root mean-squared (RMS) error is computed similarly as

$$\text{RMS} = \sqrt{\frac{1}{N} \sum (q_i - o_i)^2}. \tag{4}$$

- Spearman's rank correlation coefficient (SRCC) is defined as:

$$\text{SRCC} = 1 - \frac{6 \sum_{i=1}^{N} d_i^2}{N(N^2 - 1)}, \tag{5}$$

  where d_i is the difference between the i-th image's ranks in subjective and objective evaluations. SRCC is a nonparametric rank-based correlation metric, independent of any monotonic nonlinear mapping between subjective and objective scores.

- Kendall's rank correlation coefficient (KRCC) is another non-parametric rank correlation metric given by

$$\text{KRCC} = \frac{N_c - N_d}{\frac{1}{2} N(N - 1)}, \tag{6}$$

  where N_c and N_d are the numbers of concordant and discordant pairs in the data set, respectively.

Among the above metrics, PLCC, MAE and RMS are adopted to evaluate prediction accuracy,[19] and SRCC and KRCC are employed to assess prediction monotonicity.[19] A better objective VQA measure should have *higher* PLCC, SRCC and KRCC while *lower* MAE and RMS values. The best results are highlighted in bold font. All of these evaluation metrics are adopted from previous VQA studies.[18,19]

We can observe from Table 2 that SSIMplus not only outperforms the popular VQA quality measures in terms of perceptual quality prediction accuracy but also in terms of computation time.

4.2 Device-adaptation capability

The above test results assume a single fixed viewing device, which is a common assumption made by existing state-of-the-art VQA models. The capability of SSIMplus is beyond the limitation of existing models. In particular, SSIMplus is designed to inherently consider the viewing conditions such as display device and viewing distance. Due to the unavailability of public subject-rated video quality assessment databases that contain subject-rated video sequences watched under varying viewing conditions, we use the subjective study results described in Section 2 to test the device-adaptive capability of the SSIMplus algorithm.

Table 2: Perceptual quality prediction performance comparison

	PLCC	**MAE**	**RMS**	**SRCC**	**KRCC**	**Computation time** (normalized)
PSNR	0.5408	1.1318	1.4768	0.5828	0.3987	1
MOVIE	0.7164	0.9711	1.2249	0.6897	0.4720	3440.27
VQM	0.8302	0.7771	0.9768	0.8360	0.6243	174.53
SSIM	0.8422	0.8102	0.9467	0.8344	0.6279	22.65
MS-SSIM	0.8527	0.7802	0.9174	0.8409	0.6350	48.49
SSIMplus	**0.8678**	**0.7160**	**0.8724**	**0.8745**	**0.6737**	7.83

Table 3: Performance Comparison between PSNR, SSIM, MS-SSIM, VQM, PQR-Tek, DMOS-Tek, JND-VC, DMOS-VC and SSIMplus (device: iPhone 5S, viewing distance: 10 inches)

Model	Resolution	PLCC	MAE	RMS	SRCC	KRCC
PSNR	640p & 1080p	0.8974	6.2667	9.0641	0.9277	0.7633
SSIM	640p & 1080p	0.9498	4.1694	6.4252	0.9604	0.8249
MS-SSIM	640p & 1080p	0.9186	5.2874	8.1157	0.9438	0.7941
VQM	640p & 1080p	0.8939	6.2125	9.2098	0.9324	0.7736
MOVIE	640p & 1080p	0.9030	6.1677	8.8268	0.9318	0.7710
PQR-Tek	640p & 1080p	0.4605	13.853	18.234	0.4694	0.3323
DMOS-Tek	640p & 1080p	0.4645	13.864	18.191	0.4694	0.3323
JND-VC	640p & 1080p	0.9423	4.8685	6.8740	0.9448	0.7941
DMOS-VC	640p & 1080p	0.8729	6.9934	10.022	0.9116	0.7377
SSIMplus	640p & 1080p	**0.9781**	**3.0251**	**4.2715**	**0.9529**	**0.8275**

Table 4: Performance Comparison between PSNR, SSIM, MS-SSIM, VQM, PQR-Tek, DMOS-Tek, JND-VC, DMOS-VC and SSIMplus (device: iPad Air, viewing distance: 16 inches)

Model	Resolution	PLCC	MAE	RMS	SRCC	KRCC
PSNR	640p & 1080p	0.9097	7.7111	9.4030	0.8616	0.6684
SSIM	640p & 1080p	0.9332	6.5561	8.1391	0.8860	0.7146
MS-SSIM	640p & 1080p	0.8986	8.3154	9.9370	0.8364	0.6427
VQM	640p & 1080p	0.8971	8.2887	10.003	0.8457	0.6479
MOVIE	640p & 1080p	0.9114	7.8819	9.3206	0.8709	0.6812
PQR-Tek	640p & 1080p	0.8656	8.2467	11.337	0.8730	0.7146
DMOS-Tek	640p & 1080p	0.8681	8.1190	11.239	0.8730	0.7146
JND-VC	640p & 1080p	0.9395	6.2170	7.7568	0.9193	0.7556
DMOS-VC	640p & 1080p	0.8420	9.4575	12.216	0.7812	0.5863
SSIMplus	640p & 1080p	**0.9701**	**4.5263**	**5.4991**	**0.9131**	**0.7659**

The mean opinion scores (MOS) provided by subjects were used to compare the performance of SSIMplus with state-of-the-art VQA measures. The scatter plots of the VQA algorithms under comparison are shown in Figures 2 - 7. The superior performance of the SSIMplus algorithm compared to the other VQA algorithms is evident from the figures.

Comparisons between the VQA algorithms using PLCC, MAE, RMS, SRCC, and KRCC are provided in Tables 3 - 8. We can observe from the results that SSIMplus algorithm outperforms other state-of-th-art video compression algorithms. The main purpose of the subjective study (refer to Section 2) is to observe the adaptation behavior of state-of-the-art VQA measures when deployed for predicting the perceptual quality of video content viewed under different viewing conditions. Table 7 compares the performance of the VQA measures when the TV viewing distance is reduced to 20 inches (referred to as *expert mode*). SSIMplus adapts to the changes in the viewing conditions better than the VQA algorithms under comparison. SSIMplus algorithm is considerably

Table 5: Performance Comparison between PSNR, SSIM, MS-SSIM, VQM, PQR-Tek, DMOS-Tek, JND-VC, DMOS-VC and SSIMplus (device: Lenovo W530 laptop, viewing distance: 20 inches)

Model	Resolution	PLCC	MAE	RMS	SRCC	KRCC
PSNR	640p & 1080p	0.9122	7.6379	9.6722	0.8751	0.6940
SSIM	640p & 1080p	0.9216	7.4738	9.1659	0.8876	0.7146
MS-SSIM	640p & 1080p	0.8883	8.5300	10.841	0.8388	0.6427
VQM	640p & 1080p	0.8981	8.5620	10.383	0.8560	0.6607
MOVIE	640p & 1080p	0.9175	7.5530	9.3934	0.8852	0.7017
PQR-Tek	640p & 1080p	0.9350	6.3503	8.3737	0.9304	0.7890
DMOS-Tek	640p & 1080p	0.9446	5.7194	7.7513	0.9304	0.7890
JND-VC	640p & 1080p	0.9312	7.0154	8.6077	0.9177	0.7428
DMOS-VC	640p & 1080p	0.8248	10.5753	13.349	0.7772	0.5786
SSIMplus	640p & 1080p	**0.9698**	**4.7388**	**5.7593**	**0.9227**	**0.7813**

Table 6: Performance Comparison between PSNR, SSIM, MS-SSIM, VQM, PQR-Tek, DMOS-Tek, JND-VC, DMOS-VC and SSIMplus (device: Samsung TV 55" viewing distance: 90 inches)

Model	Resolution	PLCC	MAE	RMS	SRCC	KRCC
PSNR	640p & 1080p	0.9343	6.4934	8.2855	0.9034	0.7248
SSIM	640p & 1080p	0.9438	6.1363	7.6822	0.9140	0.7505
MS-SSIM	640p & 1080p	0.9126	7.3825	9.5003	0.8742	0.6786
VQM	640p & 1080p	0.9242	7.3915	8.8743	0.8914	0.6992
MOVIE	640p & 1080p	0.9345	6.6421	8.2690	0.9108	0.7377
PQR-Tek	640p & 1080p	0.9572	5.3377	6.7291	0.9435	0.8044
DMOS-Tek	640p & 1080p	0.9462	5.9821	7.5147	0.9435	0.8044
JND-VC	640p & 1080p	0.9512	5.8066	7.1664	0.9394	0.7864
DMOS-VC	640p & 1080p	0.8485	9.4566	12.295	0.8150	0.6171
SSIMplus	640p & 1080p	**0.9856**	**3.2147**	**3.9271**	**0.9464**	**0.8172**

Table 7: Performance Comparison between PSNR, SSIM, MS-SSIM, VQM, PQR-Tek, DMOS-Tek, JND-VC, DMOS-VC and SSIMplus (device: Samsung TV 55", viewing distance: 20 inches)

Model	Resolution	PLCC	MAE	RMS	SRCC	KRCC
PSNR	640p & 1080p	0.9204	7.5788	9.7918	0.8891	0.7077
SSIM	640p & 1080p	0.9322	7.5625	9.0674	0.9113	0.7487
MS-SSIM	640p & 1080p	0.9019	8.9489	10.820	0.8709	0.6872
VQM	640p & 1080p	0.9185	8.3203	9.9051	0.8777	0.6821
MOVIE	640p & 1080p	0.9240	7.6532	9.5777	0.9000	0.7205
PQR-Tek	640p & 1080p	0.7472	14.312	16.647	0.7227	0.5205
DMOS-Tek	640p & 1080p	0.7632	13.773	16.185	0.7221	0.5196
JND-VC	640p & 1080p	0.9357	7.1399	8.8356	0.9300	0.7692
DMOS-VC	640p & 1080p	0.8481	10.3025	13.272	0.8174	0.6205
SSIMplus	640p & 1080p	**0.9708**	**5.1424**	**6.0055**	**0.9311**	**0.7897**

faster than the other quality measures proposed to predict perceptual quality of video content and meets the requirements for real-time predictions of both perceptual video QoE and the detailed quality map.

5. CONCLUSION

A subjective study to evaluate the effects of display device and viewing conditions on the perceptual quality of digital video was presented. This study included forty videos derived from eight reference videos compressed at five quality levels and were evaluated by thirty subjects under varying viewing condition on four display devices. We evaluated the performance of several publicly available objective VQA models and SSIMplus video QoE

Table 8: Performance Comparison between PSNR, SSIM, MS-SSIM, VQM, PQR-Tek, DMOS-Tek, JND-VC, DMOS-VC and SSIMplus including all devices

Model	PLCC	MAE	RMS	SRCC	KRCC	Computation time (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**

measure on this database. Experimental results showed that the proposed algorithm outperforms state-of-the-art video quality measures in terms of its device adaptation capability, perceptual video quality prediction accuracy, and speed. More information can be found at https://ece.uwaterloo.ca/~z70wang/research/ssimplus/.

REFERENCES

[1] Cisco Systems, Inc., "Cisco visual networking index: Global mobile data traffic forecast update, 2013-2018," (Feb. 2014).

[2] 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).

[3] 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).

[4] 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).

[5] 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).

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

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

[8] Zeng, K., Rehman, A., Wang, J., and Wang, Z., "From H.264 to HEVC: coding gain predicted by objective video quality assessment models," in [*7th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*], 42–46 (Jan. 2013).

[9] Moorthy, A. K., Choi, L. K., Bovik, A. C., and de Veciana, G., "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *J. Sel. Topics Signal Processing* **6**(6), 652–671 (2012).

[10] Liao, Y., Younkin, A., Foerster, J., and Corriveau, P., "Achieving high qoe across the compute continuum: How compression, content, and devices interact," in [*7th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*], 91–95 (Jan. 2013).

[11] Inc, T., "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].

[12] Clarity, V., "Understanding MOS, JND and PSNR." http://videoclarity.com/wpunderstandingjnddmospsnr/ (2014). [Online; accessed September 07, 2014].

[13] Aimar, L., Merritt, L., Petit, E., Chen, M., Clay, J., Rullgrd, M., Heine, C., and Izvorski, A., "x264 - a free H264/AVC encoder." http://www.videolan.org/developers/x264.html (2015). [Online; accessed January 06, 2015].

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

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

[16] Baroncini, V., Ohm, J. R., and J., S. G., "Report on preliminary subjective testing of hevc compression capability," in [*JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*], (Feb. 2012).

[17] Sullivan, G., Ohm, J., Han, W., and Wiegand, T., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Techn.* **22**(12), 1649–1668 (2012).

[18] 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).

[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).