https://doi.org/10.2352/ISSN.2470-1173.2019.10.IQSP-314

Adaptive video streaming with current codecs and formats: Extensions to parametric video quality model ITU-T P.1203

***Rakesh Rao Ramachandra Rao*[1]*, Steve Göring*[1]*, Patrick Vogel*[1]*, Nicolas Pachatz*[1]*, Juan Jose Villamar Villarreal*[1]*, Werner Robitza*[1]*, Peter List*[2]*, Bernhard Feiten*[2]*, Alexander Raake*[1]**

[1] ***Dept. of Audio Visual Technology, Technische Universität Ilmenau, Germany;***

[2] ***Technology & Innovation, Deutsche Telekom AG, Germany;***

Email: [rakesh-rao.ramachandra-rao, steve.goering, patrick.vogel, nicolas.pachatz, juan-jose.villamar-villarreal, werner.robitza, alexander.raake]@tu-ilmenau.de, [peter.list, bernhard.feiten]@telekom.de

Abstract

Adaptive streaming is fast becoming the most widely used method for video delivery to the end users over the internet. The ITU-T P.1203 standard is the first standardized quality of experience model for audiovisual HTTP-based adaptive streaming. This recommendation has been trained and validated for H.264 and resolutions up to and including full-HD. The paper provides an extension for the existing standardized short-term video quality model mode 0 for new codecs i.e., H.265, VP9 and AV1 and resolutions larger than full-HD (e.g. UHD-1). The extension is based on two subjective video quality tests. In the tests, in total 13 different source contents of 10 seconds each were used. These sources were encoded with resolutions ranging from 360p to 2160p and various quality levels using the H.265, VP9 and AV1 codecs. The subjective results from the two tests were then used to derive a mapping/correction function for P.1203.1 to handle new codecs and resolutions. It should be noted that the standardized model was not re-trained with the new subjective data, instead only a mapping/correction function was derived from the two subjective test results so as to extend the existing standard to the new codecs and resolutions.

Introduction

Dynamic adaptive streaming over HTTP (DASH) is currently a widely used technology for several video streaming providers, e.g. YouTube [1] or Netflix[2]. Videos accounted for 76% of the consumer Internet traffic in 2016 and that share is predicted to increase to 82% by 2021 [6]. This clearly indicates the importance for better quality monitoring and adaptation algorithms. The main criteria for the development of better quality monitoring and adaptation algorithms is to reduce the overall required bandwidth by not having any perceivable impact on the video and audio.

Recent studies of the YouTube DASH algorithms [2] have shown that there are still improvements possible in the existing algorithms. The adaptation algorithms mainly try to optimize the size of encoded video segments to reduce the usage of transmitted bandwidth. One notable example for this kind of optimization of the storage of encoded video segments is the dynamic optimizer [17] of scenes developed by Netflix based on their VMAF [22, 21] video quality metric.

[1]`https://developers.google.com/youtube/v3/live/guides/encoding-with-dash`

[2]`https://bit.ly/2QNzpFB`

In essence, video quality metrics form a basis for optimization of adaptation algorithms. In general, pixel- and bitstream-based models/metrics can be used to estimate video quality scores [27]. The pixel-based models, namely full-reference, reduced-reference and no-reference models need access to the decoded frames to estimate video quality scores. In contrast, bitstream-based models just need access to the encoded bitstreams to determine video quality scores.

The P.1203 series of standards [14] proposed by ITU-T is one such example for bitstream-based models. P.1203 is a quality model for HTTP based adaptive audiovisual streams [25, 14, 26]. This series consists of three main parts: *Pv*: short term video quality prediction, *Pa*: audio short term quality, *Pq*: overall integration of quality including, e.g. perceived stalling effects. The *Pv* module in P.1203 has four different modes of operation, starting from mode 0 to mode 3. The modes are distinguished based on the amount of bitstream information available, ranging from just the metadata (codec, resolution, bitrate, framerate, segment duration) as in case of mode 0 to full bitstream access as in case of mode 3.

In this paper, we focus only on the mode 0 *Pv* model. The advantage of mode 0 is that it requires just meta data and is the fastest of all the modes. The speed and simplicity of mode 0 comes at a cost of reduced accuracy as compared to mode 3. In contrast, the increased accuracy of mode 3 comes at an effort of implementing a patched client decoder that extracts bitstream features, e.g. QP values. The existing P.1203 standard doesn't consider newer codecs, i.e., H.265, VP9 and AV1[3] that are currently used in DASH streaming. Also, the standard is restricted in terms of resolution up to 1080p and framerate up to 24fps. In this paper, we focus on extending the existing mode 0 model to support the aforementioned newer codecs and higher resolutions and framerate.

The method that is proposed in this paper is a correction mapping for new codecs, resolutions and framerates and not retrain the existing model. As a result, we use the unmodified mode 0 predictions from the existing model and then do a correction on this prediction for the newer use cases. This approach of just using a correction mapping and not re-training ensures that we can rely on the well-developed P.1203 models. To ensure that the proposed correction reflects quality ratings by humans, two subjective tests were conducted.

[3]`https://bit.ly/2RW5fzY`

This paper is organized as follows. In the next Section, an overview of the state-of-the-art of bitstream-based and pixel-based no-reference video quality models are described. Following this, all the aspects related to the subjective tests conducted for this study are explained in detail. In the subsequent two sections, the development of the proposed extension and the evaluation of the proposed extension are described. Finally, we conclude with a discussion on the proposed extension and provide a few insights for future work.

Video Quality Models

Many pixel-based full-, reduced- and no-reference based models have been proposed in the literature [5, 27]. Also, several bitstream based models which take into account coding artifacts, packet losses etc. are reported in the literature. We will mostly focus on no-reference metrics in the following section.

Torres Vega et al. [28] studied the performance of different NR metrics in the presence of network impairments, e.g. packet-loss, however they don't consider higher resolutions, framerates and newer codecs. In addition, in our paper, we focus on video quality prediction for DASH segments, where no packet-loss is included. In DASH, different distortions occur for long term video quality, e.g. stalling, compare *Pq* part of ITU-T P.1203 [14, 25, 26].

In general, modern no-reference models are able to outperform full-reference models. E.g., Göring, Skowronek, and Raake [9, 8] proposed a no-reference video/image quality model using pre-trained DNN's for feature extraction. The presented model includes higher framerates, newer codecs and 4K resolution, however it is a pixel based no-reference model. A comparison to state-of-the-art full-reference models is done, and the proposed model is able to outperform some full-reference models.

There are also works in literature dealing with no-reference stereoscopic video quality assessment. One such example is the NR-stereoscopic video quality assessment proposed by Appina et al. [3] which uses joint motion and depth statistics.

Similar to the pixel-based no-reference models, a wide range of bitstream-based no-reference models have been proposed in the literature. Usually bitstream-based methods are faster regarding computation time compared to pixel based models, because only the stored bitstream without full decoding of video frames is analyzed. For example, Joskowicz, Bovino, and Arado [16] presents a review of parametric models for video quality estimation. This work concludes that the parametric models show good results when compared with subjective quality ratings.

Furthermore, Keimel et al. [18] present a H.264/AVC bitstream no-reference video quality metric with multiway partial least squares regression. The bitstream features are based on extracted motion values, qp-values, frame-types and more. The presented model is therefore a mode 3 model. They show, that such a model has a good performance compared to full-reference models.

In addition, Raake et al. [24] presented the "T-V Model", a parametric model for video quality estimation for SD and HD TV. This model also considers packet loss. Gustafsson, Heikkila, and Pettersson have proposed in [10] a model that takes into account the combined effects of packet loss and buffering. Garcia et al. [7] have presented a parametric video quality model for IPTV applications which accounts for distortion both due to compression and erroneous transmission. It is divided into three modules, for audio, video and audiovisual quality. The model is applicable to the quality monitoring of encrypted and non-encrypted audiovisual streams. This model is standardized as ITU-T Recommendation P.1201.2 [13], the higher resolution (IPTV and Video on Demand (VoD)) algorithm of Recommendation P.1201.

Considering, that mostly all popular video streaming providers use DASH for video transmission, other effects during play-out are introduced and need to be considered in such models. The ITU-T Recommendation P.1203 [14, 26, 25] is the first standardized quality of experience model for audiovisual HTTP-based adaptive streaming. Similar to the ITU-T P.1201 recommendation, this recommendation also is divided into three modules, for audio, video and integrated audiovisual quality. This model is applicable for H.264 encoded videos up to 1080p resolution and a framerate of 24fps. This standard takes into account the stalling information in predicting video quality.

To summarize, we checked several state-of-the-art no-reference models. Our main focus are bitstream based models, as none of the bitstream models are able to handle newer codecs, resolutions and framerates. The proposed solution is a direct extension of the existing standard ITU-T P.1203. More substantial updates will follow with the finalization of the ongoing Phase 2 of the so-called P.NATS standardization work - P.1203 was previously termed "P.NATS". The development of the P.NATS Phase 2 bitstream models is conducted as part of the AVHD / P.NATS Phase 2 development jointly performed by the Video Quality Experts Group (VQEG) and ITU-T Study Group 12, Q14. The present paper here adds another codec, AV1, which is not part of the P.NATS Phase 2 work.

Subjective Tests

This section describes the two subjective tests conducted as part of this study. The details about the test environment, equipment, methodology, sources and test design of both the tests are described in the following sections.

Test Environment, equipment and methodology

The test environment is based on [12] in order to guarantee repeatability. To display the 4K/UHD contents a 55" OLED screen was used. In order to ensure seamless playback, we use an interface to a DeckLink 4K Extreme 12G card.

Also, AVRateNG[4], the new version of AVRate [20], was used as rating software in both of the tests. In this study, we use the absolute category rating (ACR) method for quality assessment according to ITU standard BT.500-11 [12]. Prior to the ACR-based, a simple eye-test was conducted on the subjects using the Snellen chart to ascertain the visual acuity of the subjects. The test lasted 60 minutes with two optional 5-minute breaks in between.

Source Contents

The source contents were selected based on different spatial and temporal complexities. 10 s long video stimuli were used in both tests. This duration is comparable with the segment length in various DASH applications [23].

To quantify the spatial and temporal complexities, the spatial information (SI) and temporal information (TI) metrics based

[4] `https://bit.ly/2QlCGft`

on [15] were used. These metrics were calculated using our publicly available implementation[5].

The SI and TI values of all the source contents that were used in both of the tests are shown in figure 1 and 2. 6 different sources were used in the first test and 7 in the second test. The details of the source contents are mentioned in Tables 1 and 2. As it can be seen in Table 2, the "Space" video was originally in 30 fps. This source was sped up to 60 fps with no negative impact on the content.

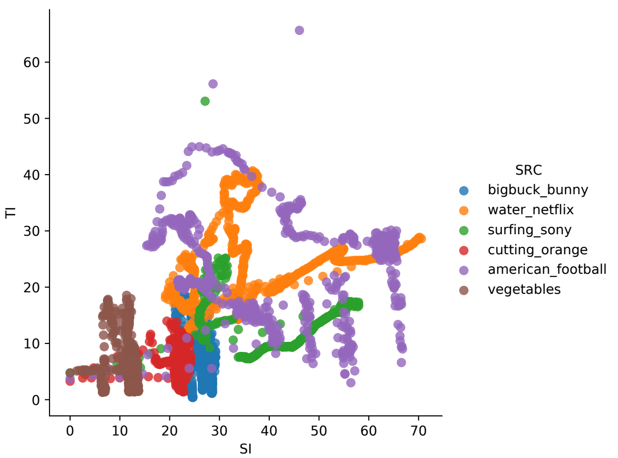

Figure 1. SITI values for sources used in test 1

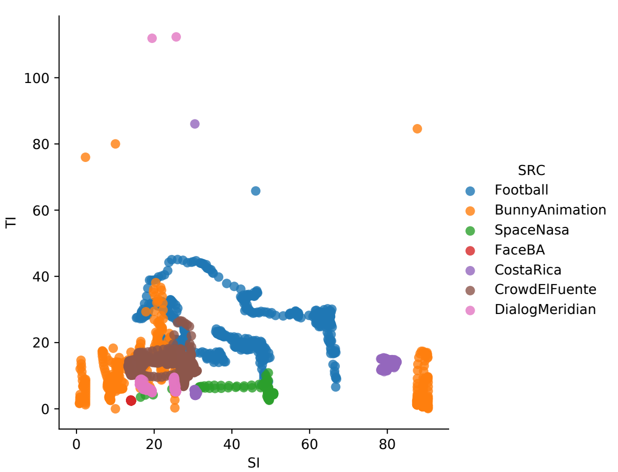

Figure 2. SITI values for sources used in test 2

Test Design

The first subjective test is a codec comparison test using 3 codecs, namely H.264, H.265 and VP9. H.264 was used as one of the codecs to evaluate the performance of the standard for videos above 1080p resolution and 24 fps as the existing standard dealt only up to a resolution of 1080p and framerate of 24 fps for H.264 and possibly retrain the coefficients for H.264 to handle higher resolution and framerate.

[5] https://bit.ly/2oXxQIN

Table 1: Source details for the first subjective test

Content Name	Duration	Details	Source of the content
American Football	10s	3840x2160@60fps	Harmonic Inc. [11]
Bigbuck Bunny	10s	3840x2160@60fps	Blender Foundation [4]
Cutting Orange	10s	3840x2160@60fps	TU Ilmenau
Surfing Sony	10s	3840x2160@60fps	Sony
Vegetables	10s	3840x2160@60fps	TU Ilmenau
Water	10s	3840x2160@60fps	Netflix Inc.

Table 2: Source details for the second subjective test

Content Name	Duration	Details	Source of the content
Animation	10s	3840x2160@60fps	Blender Foundation [4]
Landscape	10s	3840x2160@60fps	Harmonic Inc. [11]
Crowd	10s	3840x2160@60fps	Netflix Inc.
Dialog	10s	3840x2160@60fps	Netflix Inc.
Face	10s	3840x2160@60fps	TU Ilmenau
Sport	10s	3840x2160@60fps	Harmonic Inc. [11]
Space	10s	3840x2160@30fps*	NASA

Table 3: Test Design - Subjective Test 1

Resolution	Bitrate [kbit/s]					
360p	200	750				
720p		750	2000			
1080p			2000	7500	15000	
2160p				7500	15000	40000

In this test, in total 6 different source contents were used. We encoded the source videos with H.264, H.265 and VP9 codecs using ffmpeg. To simulate a realistic encoding setting, we selected a 2-pass encoding scheme with a 8-bit color depth and chroma sub-sampling of 4:2:0. For H.264 and H.265 the default preset of **medium** was used and the corresponding **speed** parameter for VP9 was set to 0 which is the default value. The resolution of the encoded videos ranged between 360p and 2160p and the bitrate between 200kbps and 40000kbps. The detailed test design in terms of bitrate and resolution is described in Table 3. In total 10 (bitrate, resolution) combinations were used for each source and codec, thus resulting in 60 processed video sequences (PVS) for one codec and 180 PVS's in total. In the test, a total of 29 subjects participated.

In contrast to the first test, the second test uses only 2 codecs,

Table 4: Test Design - Subjective Test 2

Resolution	Bitrate [kbit/s]					
360p	512	1024	2048			
720p		1024	2048	4096		
1080p			2048	4096	8192	
2160p				4096	8192	16384

i.e., H.265 and AV1. A total of 7 different source contents were used in this test which were encoded with H.265 and AV1 (version released in April 2018) using ffmpeg 4.0. The encoding followed a 2-pass scheme as in the first test but with a 10-bit color depth and a color sub-sampling of 4:2:2 unlike the first test which used 8-bit color depth and 4:2:0 chroma sub-sampling. The preset used for H.265 was the default **medium** and the corresponding **cpu-used** parameter for AV1 was 4. We decided to use the cpu-used=4 parameter due to performance reasons in the encoding process and to have a comparable preset as for H.265. Each source was encoded in 4 resolutions and 3 bitrates per resolution resulting in a total of 168 PVS's. The detailed test design in terms of bitrate and resolution is described in Table 4. In total, 27 subjects participated in the test.

Taking both tests into account, in total 348 stimuli were presented to, and rated by 48 participants, 24 in each test after outlier removing. These ratings formed the ground truth for the development of the proposed extension of the model.

Test Results

For checking the reliability of the users, outlier detection was performed during the analysis. The criteria for outlier detection was based on Pearson correlation coefficient (PCC). PCC was computed between the raw scores of each user and the mean opinion score (MOS). A threshold of 0.8 PCC was used as a criteria to detect outliers. Based on this threshold, there were two outliers in each of the two tests. We then computed the MOS and the associated confidence interval (95% CI).

Figure 3 shows the results of the first subjective test. It can be observed that in the lower bitrate ranges, the performance of all the three codecs are similar. Whereas, in the higher bitrate ranges, H.265 and VP9 outperform H.264

Figure 4 shows the results of the second subjective test. Here, we can see that AV1 outperforms H.265 over the entire spectrum of the bitrates. We compare the results of our subjective test with those available in literature [1, 19] w.r.t the comparison between the performance of AV1 and H.265. The results in Akyazi and Ebrahimi [1] clearly highlight that on average AV1 performs better than H.265 both in terms of MOS and bitrate savings. This is in synchronization with the results of our test w.r.t MOS where we see that AV1 outperforms H.265. Laude et al. [19] compares JEM and AV1 with H.265. This study does an objective evaluation based on coding efficiency and computational complexity of the three codecs under well-defined and balanced test conditions. This study concludes that AV1 has increased complexity and it cannot transform this increased complexity into competitiveness in terms of coding efficiency with either H.265 or JEM except for the all-intra configuration. This study does not report any subjective comparison between the codecs and hence we cannot make a substantiative comparison between our test and this particular study, furthermore there is also an ongoing work to speedup AV1 encoding and decoding.

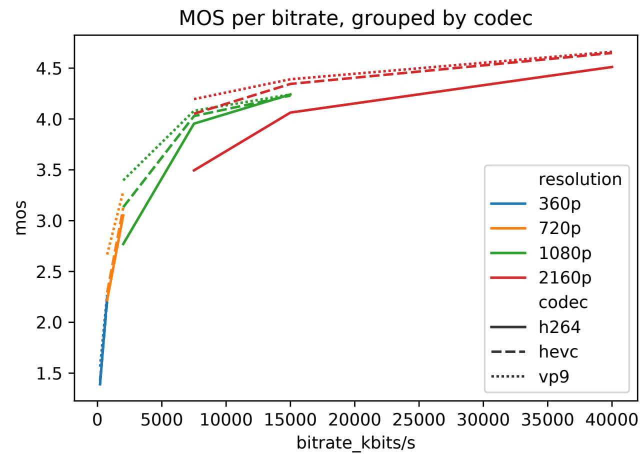

Figure 3. *Results of first subjective test: H.264 vs. H.265 vs VP9.*

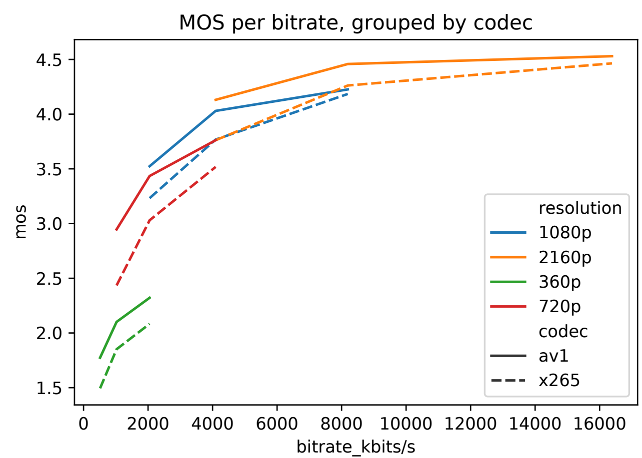

Figure 4. *Results of second subjective test: AV1 vs. H.265.*

Proposed P.1203.1 Mode 0 Extension

The motivation for proposing the extension to the existing P.1203.1 mode 0 model was the wide-ranging usage of codecs other than H.264 such as H.265 and VP9 in DASH and also the advent of new codecs such as AV1. In addition to the new codecs, also higher resolutions and framerates are considered as they form a considerable part of the video contents in various streaming services.

The subjective ratings from the two tests were used to derive a mapping/correction function for the ITU-T P.1203.1 mode 0 model to handle new codecs, resolution and framerate. As a pre-processing step, mean opinion scores (MOS) were obtained by averaging the individual ratings over all individual sources to eliminate content dependency as the mode 0 model is unable to handle this content dependency in a meaningful way. Various extensions were developed using different parameters such as bitrate, resolution and codec as input parameters in several combinations. In addition to these parameters, the output of P.1203.1 mode 0 model, referred as $mode0\_output$, was also used as an additional input to the mapping/correction function. While computing the P.1203.1 mode 0 output for the subjective data, appropriate changes in terms of codec and resolution handling were made to the existing model to take into account newer codecs and resolution. We analyzed different possible input parameters, e.g. only {codec}, only {resolution, codec}, only {bitrate} and {resolution, codec, bitrate}. We found that {resolution, codec, bitrate} parameters are required for good performance of such a correction function, due to the fact that modern codecs are designed to handle lower and higher resolutions with varying bitrates differently than H.264.

We used curve fitting for training. The software is based on Python 3 and uses LmFit[6]. Several candidate functions were checked to determine the best performing function. Out of these candidates, the below mentioned candidate, see Equation 1, was the best performing extension.

The final correction/mapping function is as follows:

$$predicted\_mos = a + b * mode0\_output + c * log(bitrate) + d * log(resolution) \quad (1)$$

For each video codec a different set of coefficients is used. In Table 5 all coefficients are summarized.

Table 5: Correction Mapping - Coefficients per codec

Codec	a	b	c	d
H.264	-0.19	0.04	0.72	-0.18
H.265	0.05	0.47	0.40	-0.09
VP9	-3.55	-0.008	0.43	0.25
AV1	-7.38	-0.18	0.46	0.54

Evaluation of the proposed extension

In this section, a comparison of the correction method with non-adjusted P.1203 is conducted. At first, the non-adjusted P.1203 mode 0 Pv model was used to obtain the predicted MOS for the two subjective tests that were conducted. This was done to ascertain if the existing model works well for all the codec, resolution and framerate extensions or if there is a need for a correction mapping to handle these extensions. The evaluation of the unadjusted model showed that the performance in terms of rmse is as follows for the three new codecs: $rmse_{h264} = 0.66$, $rmse_{h265} = 0.65$, $rmse_{vp9} = 0.69$ and $rmse_{av1} = 0.9$. These values diverge considerably from the rmse of 0.465 for mode 0 as reported in the standard. Even for the case of H.264, the difference between the rmse reported in the standard and the one for our test is also considerably large. This deviation in performance is expected as the existing model was trained for H.264 for only resolutions up to 1080p and framerates up to 24fps. In order for the model to take into account the new extensions, either the model has to be re-trained for these extensions or else a correction mapping can be done.

[6]https://bit.ly/2K8QcQZ

The decision to go for a simple correction mapping over re-training the model was based on the rationale that such a correction mapping would keep the structure and coefficients of the original model intact and ensures that we can rely on the well-developed P.1203 models. Keeping in view this rationale, we proposed a correction mapping which takes the output of the original mode 0 model as an input and performs the correction based on bitrate, resolution and framerate. Figure 5 shows the performance of the correction mapping for all the four codecs. It can be seen that the RMSE is lower than 0.3 for all the codecs for the mapped version, and hence it can be concluded that the developed method is better in terms of RMSE than the original, non-adjusted model for the new application scenarios.

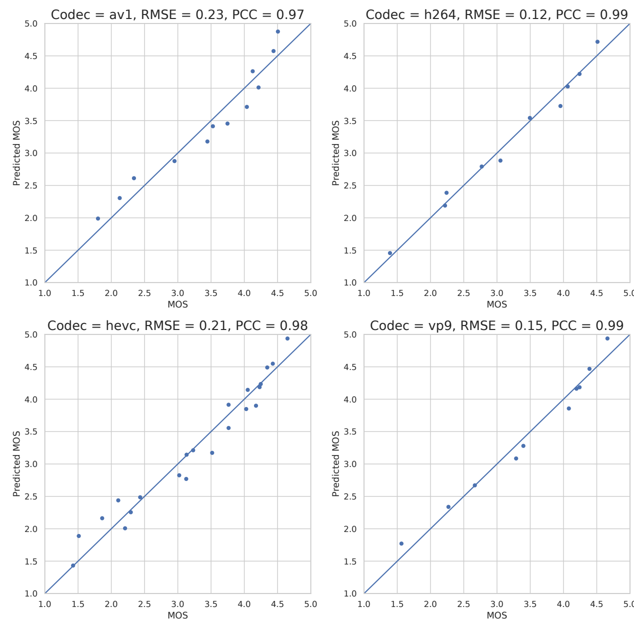

Figure 5. Performance of the correction mapping for all considered video codecs

Conclusion and Future Work

We analyzed modern state-of-the-art no-reference models, with specific focus to bitstream based models. We found out, that current bitstream based no-reference video quality models are not able to handle newer codecs, higher resolutions and framerates. It is even hard to extend such models, because for new codecs there is a complete change in the bitstream. In this paper, we propose and describe a first method to extend P.1203 Mode 0, a well-developed and standardized bitstream model, to handle new codecs such as H.265, VP9 and AV1, higher resolution (up to 4K) and framerates (up to 60fps). We only consider Mode 0 of the P.1203 model, because this mode is based on meta-data, i.e., framerate, bitrate, resolution and codec, that is available in nearly all real world scenarios. To develop such an extension, we conducted two subjective tests. The first test considered, H.264, H.265 and VP9 with resolutions up to 4K, 60 fps as framerate and realistic bitrate settings. In addition, we conducted a second subjective test, where we included H.265 and AV1 as codecs, with settings comparable to the first test. Our general approach for both tests was that we include similar conditions, so that later a combination of both tests can be used. In our paper, we describe a mapping function, that uses the original predicted quality score of P.1203, assuming that e.g. video codec H.264 was used, and meta-data of the video stream. Our approach can also be used for other bitstream based models as it uses the originally predicted score by the model and performs a correction on it. This work is a first step to describe methods that are able to "update" state-of-the-art models for new domains of videos (e.g. codecs, resolutions, framerates or even HDR vs LDR) using a simple correction mapping.

References

[1] P. Akyazi and T. Ebrahimi. "Comparison of Compression Efficiency between HEVC/H.265, VP9 and AV1 based on Subjective Quality Assessments". In: (2018), p. 6.

[2] J. Añorga et al. "YouTube's DASH implementation analysis". In: *19th International Conference on Circuits, Systems, Communications and Computers (CSCC)*. 2015, pp. 61–66.

[3] B. Appina et al. "No Reference Stereoscopic Video Quality Assessment Using Joint Motion and Depth Statistics". In: *CoRR* abs/1711.05480 (2017).

[4] Blender Foundation. *Bick Buck Bunny Distribution*. URL: http://distribution.bbb3d.renderfarming.net/video/png (Accessed: 10/20/2018).

[5] S. Chikkerur et al. "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison". In: *IEEE Transactions on Broadcasting* 57.2 (2011), pp. 165–182.

[6] Cisco. *Cisco Visual Networking Index: Forecast and Methodology, 2016–2021*. 2017.

[7] M. Garcia et al. "Parametric model for audiovisual quality assessment in IPTV: ITU-T Rec. P.1201.2". In: *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*. 2013, pp. 482–487.

[8] S. Göring and A. Raake. "deimeq – A Deep Neural Network Based Hybrid No-reference Image Quality Model". In: *Visual Information Processing (EUVIP), 2018 7th European Workshop on*. to appear. IEEE. 2018, pp. 1–6.

[9] S. Göring, J. Skowronek, and A. Raake. "DeViQ – A deep no reference video quality model". In: *Electronic Imaging, Human Vision Electronic Imaging* (2018).

[10] J. Gustafsson, G. Heikkila, and M. Pettersson. "Measuring multimedia quality in mobile networks with an objective parametric model". In: *2008 15th IEEE International Conference on Image Processing*. 2008, pp. 405–408.

[11] Harmonic. *Free 4K Demo Footage - Ultra HD Demo Footage*. URL: https://www.harmonicinc.com/4k-demo-footage-download/ (Accessed: 10/20/2018).

[12] ITU-T. *RECOMMENDATION ITU-R BT.500-11 - Methodology for the subjective assessment of the quality of television pictures*. Tech. rep. International Telecommunication Union, 2014.

[13] ITU-T. *Recommendation P.1201 - Parametric non-intrusive assessment of audiovisual media streaming quality*. Tech. rep. International Telecommunication Union, 2014.

[14] ITU-T. *Recommendation P.1203 - Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport*. Tech. rep. International Telecommunication Union, 2016.

[15] ITU-T. *Subjective video quality assessment methods for multimedia applications*. Serie P: Telephone Transmission Quality, Telephone Installations, Local Line Networks. Vol. Recommendation ITU-T P.910. International Telecommunication Union. Geneva, 2008.

[16] J. Joskowicz, R. G. S. Bovino, and J. C. L. Arado. "Comparison of parametric models for video quality estimation: Towards a general model". In: *IEEE international Symposium on Broadband Multimedia Systems and Broadcasting* (2012), pp. 1–7.

[17] I Katsavounidis. "Dynamic optimizer–a perceptual video encoding optimization framework". In: *The Netflix Tech Blog* (2018).

[18] C. Keimel et al. "Design of no-reference video quality metrics with multiway partial least squares regression". In: *2011 Third International Workshop on Quality of Multimedia Experience*. 2011, pp. 49–54.

[19] T. Laude et al. "A Comparison of JEM and AV1 with HEVC: Coding Tools, Coding Efficiency and Complexity". In: *2018 Picture Coding Symposium (PCS)*. 2018, pp. 36–40.

[20] P. Lebreton, M.-N. Garcia, and A. Raake. "AVRate: An Open Source Modular Audio/Visual Subjective Evaluation Test Interface". In: *8th International Workshop on Quality of Multimedia Experience (QoMEX)* (2016).

[21] J. Y. Lin et al. "A fusion-based video quality assessment (fvqa) index". In: *APSIPA, 2014 Asia-Pacific*. Dec. 2014, pp. 1–5.

[22] Netflix. *Netflix VMAF*. URL: https://github.com/Netflix/vmaf (Accessed: 10/20/2018).

[23] R. Pantos. *HTTP Live Streaming*. 2011. URL: https://tools.ietf.org/html/draft-pantos-http-live-streaming-13 (Accessed: 10/20/2018).

[24] A. Raake et al. "T-V-model: Parameter-based prediction of IPTV quality". In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2008, pp. 1149–1152.

[25] A. Raake et al. "Scalable Video Quality Model for ITU-T P.1203 (aka P.NATS) for Bitstream-based Monitoring of HTTP Adaptive Streaming". In: *QoMEX 2017*. IEEE. 2017.

[26] W. Robitza et al. "HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 – Open Databases and Software". In: *9th ACM Multimedia Systems Conference*. Amsterdam, 2018.

[27] M. Shahid et al. "No-reference image and video quality assessment: a classification and review of recent approaches". In: *EURASIP Journal on Image and Video Processing* 2014.1 (2014), p. 40.

[28] M. Torres Vega et al. "An experimental survey of no-reference video quality assessment methods". In: *International Journal of Pervasive Computing and Communications* 12.1 (2016), pp. 66–86.