https://doi.org/10.2352/ISSN.2470-1173.2021.9.IQSP-264

Enhancement of Pixel-based Video Quality Models using Meta-data

Rakesh Rao Ramachandra Rao, Steve Göring, Alexander Raake
Audiovisual Technology Group, Technische Universität Ilmenau, Germany
Email: [rakesh-rao.ramachandra-rao, steve.goering, alexander.raake]@tu-ilmenau.de

Abstract

Current state-of-the-art pixel-based video quality models for 4K resolution do not have access to explicit meta information such as resolution and framerate and may not include implicit or explicit features that model the related effects on perceived video quality. In this paper, we propose a meta concept to extend state-of-the-art pixel-based models and develop hybrid models incorporating meta-data such as framerate and resolution. Our general approach uses machine learning to incooperate the meta-data to the overall video quality prediction. To this aim, in our study, we evaluate various machine learning approaches such as SVR, random forest, and extreme gradient boosting trees in terms of their suitability for hybrid model development. We use VMAF to demonstrate the validity of the meta-information concept. Our approach was tested on the publicly available AVT-VQDB-UHD-1 dataset. We are able to show an increase in the prediction accuracy for the hybrid models in comparison with the prediction accuracy of the underlying pixel-based model. While the proof-of-concept is applied to VMAF, it can also be used with other pixel-based models.

Introduction

With the advent of a variety of "over-the top" (OTT) video streaming providers, such as YouTube, Amazon Prime Video or Netflix, video is fast amounting to a major chunk of consumer internet traffic. Videos accounted for 76% of the consumer internet traffic in 2016, predicted to increase to 82% by 2021 [2]. In this context, HTTP-based adaptive streaming (HAS) such as the standardized Dynamic adaptive streaming over HTTP (DASH) has become the preferred technology for video streaming of different type contents such as traditional 2D videos, 360° videos etc. This necessitates the development of video quality models to take into account the specific properties of HAS-type streaming and also the higher video resolutions such as 4K. In addition to monitoring video quality, these models can also form a basis for optimizing adaptation algorithms applied in the player or the encoding scheme used in the streaming backend. One such example of a video quality model used for optimizing encoding settings and adaptation algorithms is VMAF [14, 18] which is used by Netflix in the dynamic optimizer [12].

Providing end-users with the best Quality of Experience (QoE) is one of the main goals of all OTT video streaming providers. Video quality forms a major part of the user's Quality of Experience (QoE), beside e.g. likability or enjoyment of the content. Here, video quality is one factor that streaming and internet service providers can steer via technology settings such as encoding or network properties. Hence, measuring video quality with high accuracy becomes very important in today's streaming scenario.

In general, video quality models can be classified into the following three main categories: pixel-based, bitstream-based and hybrid models, where the model type depends on the input data that is used for quality assessment [22, 28]. Firstly, pixel-based models use the decoded frames to estimate video quality scores. They can be further categorized into three types, namely, full-reference (FR), reduced-reference (RR) and no-reference (NR), depending on whether a undistorted reference video, partial information about the reference video, or no reference video is being used [9] for quality estimation. The second type of models are bitstream-based models. Similar to the pixel-based models, depending on the level of access to the bitstream, four main types have recently been distinguished in the context of ITU-T Rec. P.1203 [11, 22], namely Mode 0, Mode 1, Mode 2 and Mode 3 for short and long term video quality estimation. These four model types form a hierarchy, thus, a Mode 3 model has a superset of accessible features including those available to a Mode 0 type model. For the short-term video quality prediction, a Mode 0 type model only has access to meta-data such as the codec used, resolution, bitrate and framerate. On the other side, for longer-term quality prediction for more real-life, several minute long viewing sessions, additional side information is required, about stalling events and initial loading delay. In addition to the Mode 0 input information, a Mode 1 model also has access to the video frame types, I, P, B, etc., and frame sizes. Finally, a Mode 3 model can access the full encoded bitstream. Note that the intermediate Mode 2 is a specific approach that principally uses the same information as Mode 3, but is allowed to only parse up to 2% of the Mode-3-type information.

The last general category of video quality models are hybrid models, they are a combination of any of the bitstream model types with any of the pixel-based model types and theoretically can incorporate the best of both the pixel and bitstream models and be better than both these model types based on prediction accuracy.

In this paper, we propose a concept to develop pixel-based hybrid models that take resolution and framerate as additional meta-data input to improve the prediction accuracy as compared to the underlying pixel-based model. We decided to include resolution and framerate as the only meta-data input, since these are the main parameters that are required to properly play out a given video on the corresponding end-device. Moreover, the video codec or bitrate could be used, however, inclusion of these may create a specific codec-dependency. Since the considered pixel-based model was shown to work well for codec- and bitrate-

related distortions [21, 23, 24], specific inclusion of this information may reduce the otherwise rather wide applicability of the underlying pixel-based approach. The general idea was inspired by the observation in [23] that the performance of state-of-the-art full-reference metrics drops in case of framerates different from the typical 30 or 60 fps. To demonstrate the validity of our introduced meta concept, we develop a hybrid full-reference model that uses VMAF [14, 18] as the underlying full-reference component, by including a compensation for the observed limitations of the model. The approach is not restricted to just full-reference pixel-based models and can be extended to reduced- and no-reference models as well.

The paper is organized as follows. In the next section, an overview of the existing full-reference and bitstream video quality models is provided. Following this, the proposed approach and machine learning pipeline for feature selection and model development are described. Subsequently, the dataset that is used for validating the approach demonstrated using VMAF as the underlying pixel-based model is described and the resulting evaluation associated with this is also presented. Finally, we conclude with a discussion and provide an outlook for future work.

Related Work

In the quest for higher prediction accuracy across different content types and also to take into account the effect of perception in video quality prediction, a number of pixel-based and bitstream-based models have been proposed in literature [1, 11, 21, 22, 24, 26, 31].

There are several image quality models that are used for predicting video quality. One of the popular perception-based image quality models is the Structural Similarity (SSIM) index [30]. SSIM is based on the assumption that the Human Visual System (HVS) is highly adapted for extracting structural information from the scene, and thus measuring the structural similarity provides a good approximation of image quality [29]. There exists a multi-scale variant of SSIM known as MS-SSIM [29], which can incorporate image details at different resolutions. These variants are also used for video quality prediction by using them to predicting the quality of each frame and then pooling the per-frame quality to get the overall video quality, e.g. using the average value of all frame scores.

Other popular image quality models that are used either as video quality models or as features in a compound video quality model such as VMAF [19] are the "visual information fidelity" (VIF) in the pixel domain [27] and "detail loss" metric (DLM) [13]. VIF is based on natural scene statistics (NSS). Furthermore, DLM proposes two simple quality measures to correlate detail losses and additive impairments with visual quality, respectively.

Beside full-reference image and video quality metrics, there have been also no-reference models developed. BRISQUE [16], NIQE [17] and deimeq [5] are some of the no-reference equivalents of perception-based image models, where the final model prediction is performed using machine learning, e.g. with a support vector regression or tree-based algorithms.

All video quality models that only use image-based models suffer from the same drawback of lack of motion-related features and provide reduced performance at higher resolutions, because they are usually trained on lower resolutions. For this reason modern video quality models incorporate video-specific features.

On the most popular and widely used full-reference video quality models is Netflix's VMAF [18]. The latest version of VMAF is trained on 4K videos, so it overcomes the resolution limitation of VQM [19]. Moreover, VMAF is a compound video quality model based on two image metrics namely, VIF (at 4 different scales) and DLM. In addition to the image models, a temporal frame-difference feature is included in VMAF. In general, VMAF shows good performance for 4K video quality prediction [7, 21, 23, 24].

In addition to this, the recently standardized P.1204.4 model, a reduced-reference model, has been reported to show good performance on a wide number of different datasets [21]. Most of the widely used pixel-based models such as VMAF suffer from the unavailability of time-related information (e.g. a feature that can estimate the impact on video quality cause by framerate), and also the calculation of the temporal frame difference feature suffers in case of a framerate different from the initial training set, as it is shown e.g. in [23]. In addition to this, Zinner et al. [32] study the impact of resolution and framerate on QoE metrics and proposed a framework for QoE management for content distribution systems based on H.264 SVC, thus showing the need to incorporate features that quantify the impact of resolution and framerate in video quality metrics. Moreover, Madhusudana et al. [15] in their study on the subjective and objective quality assessment of high frame rate videos show how the prediction accuracy varies widely across different framerates starting from 24fps to 120fps for a number of pixel-based metrics such as VMAF, SSIM, MS-SSIM, PSNR etc. In essence, it can be concluded that having framerate and resolution related information as additional features would enhance the accuracy and applicability of pixel-based models.

Similar to the pixel-based full-reference models, a wide range of bitstream-based models have been proposed in the literature [20, 24, 25]. Also, bitstream models have been proposed for other applications such as 360° videos [3]. The main advantage of bitstream- over pixel-based models is their low computation time since these models only use the bitstream as input and no decoding of the bitstream to pixels is required. While generally being of lower computational complexity than pixel-based models, bitstream models also vary in complexity, depending on the input data. The type of models can vary from very simple models which only use meta-data such as bitrate, framerate and resolution, to complex ones which make use of the entire bitstream information. The P.1203 series of recommendations [20] and the P.1204.3 model [24] are some of the examples of very good performing bitstream-based models with P.1204.3 showing a performance of either on par or better in comparison to the existing SoA FR metrics.

To sum up, it can be stated that different model types have different limitations. While showing competitive performance for condition types that they were trained for (codecs and bitrates used), bitstream-based models cannot easily be generalized to other codecs than those they were trained on. In turn, pixel-based models have to estimate the quality-impact of straightforward degradations such as resolution or framerate reduction using pixel-information, with reduced prediction performance, although this information may easily be available from meta-data provided with a given stream. In addition, even though pixel-based models include upscaling degradations, recently developed

DNN-based models are able to outperform SoA upscaling algorithms [4], whereas on the contrary such new upscaling algorithms introduce unknown distortions that are not covered in the training/development process of SoA FR models.

To overcome the respective model limitations in a mutual way, the best features from both types of models can be combined to develop a hybrid model that enhances the overall prediction accuracy of each individual model type. The paper presents a simple but effective approach to extend existing pixel-based models with meta-data input about the two key issues that can easily be captured with the available meta-data.

Proposed Approach

The general model structure is shown in Figure 1. Our approach starts with extracting pixel-based features from the video input. In case of an underlying full-reference model, the video input are the source video and distorted video, or only the distorted video or different variants, depending on the model type. To prove the validity of the concept, we use the popular full-reference model, Netflix's VMAF, as feature extractor. Similar extensions can also be used for no- or reduced-reference models.

After feature extraction, the per frame feature values are then temporally pooled to obtain the final per-video features. Temporal feature pooling is a well-known approach to remove the time dependency of short video sequences and therefore to provide a constant number of features to the underlying machine learning model. Similar methods have been used in other models, e.g. nofu [6]. Such a temporal feature pooling can range from a simple arithmetic mean to more complex methods such as harmonic mean, Minkowski summation, percentile etc. For the present paper, the focus is on the widely used and simple arithmetic mean.

In addition to the pooled features, we include resolution and framerate as meta-data features, resulting in a hybrid extension of the considered FR model. The rationale behind including the meta-data as additional features is that in an HTTP-based adaptive streaming (HAS) scenario as with DASH, such information is accessible to the DASH client via the manifest file required for selecting and playing the current video segments. Moreover, in any kind of play-out scenario the resolution and framerate meta-data is required to ensure a correct playing of the video. In this paper, we use ffprobe which is part of ffmpeg to estimate the required meta-data.

In the hyperparameter and feature optimization step, we consider a large combination of **(hyperparameter, feature set)** values while training the machine learning algorithm for enhancing the pixel-based model using meta-data and choose the combination that performs the best in terms of Pearson correlation coefficient.

Model Instances and Evaluation

This section presents the dataset that was used to train and validate the described meta-data-based hybrid models. We further describe several instances of hybrid models that rely on different machine learning based models. All models use VMAF as the underlying FR model. We also show how our introduced models perform in various evaluation experiments.

Dataset

To evaluate the proposed framework and the resulting hybrid variant of VMAF, we use the publicly available AVT-VQDB-UHD-1 dataset [23]. This dataset consists of four different subjective video quality tests. All four tests follow a full-factorial test design. All degraded videos videos were presented on a 4K/UHD-1 screen. The video resolutions used in these tests ranged from 360p to 2160p, with framerates between 15 fps and 60 fps and bitrates between 200 kbps and 40000 kbps. In total, three codecs, namely, H.264, H.265 and VP9, were used for encoding. These codecs cover the range of practically used video codes.

Test #1 of the AVT-VQDB-UHD-1 is a codec comparison test using all three codecs H.264, H.265 and VP9. In total, 6 different source contents of 10s duration were used in the test. The framerate of the encoded videos was not changed for this test. A total of 180 processed video sequences (PVSs) were used in the test, collecting quality ratings from 29 participants for each of the videos. A Pearson Correlation Coefficient (PCC) of 0.75 between the individual subjects' ratings to the mean ratings across all subjects was used as criterion for outlier detection [11], with no outliers being detected for this test.

Furthermore, Tests #2 and #3 followed a similar design as that of Test #1, using three source contents from Test #1 and three new source contents. Like in Test #1, the framerate of the encoded videos was kept at the source video framerate. Due to increase in the number of Hypothetical Reference Circuits (HRCs), each test was a codec comparison test between 2 codecs with Test #2 using H.264 and H.265, and Test #3 H.265 and VP9. Each of the 2 tests had a total of 192 PVSs. There was a total of 24 participants in Test #2 and 25 participants in Test #3, with no outliers detected using the criterion of 0.75 PCC.

The last included test, Test #4 differed in design to the other three tests, with the framerate of the encoded videos now being varied from 15fps to 60fps using 8 different source contents of originally 60fps. Also, this test was not a codec comparison test but focused mostly on the difference in perceived quality with videos of different bitrates, resolutions and framerates. Hence, only one codec, H.264, was used for encoding. A total of 192 PVSs were rated by each of the 26 participants. Overall 2 outliers were detected using the criterion of 0.75 PCC.

In total, across the 4 tests 756 PVSs were rated by 104 participants.

Evaluation

To validate the proposed approach, we use VMAF as the underlying FR model to develop the hybrid models. In total, three machine learning algorithms, namely, support vector regression (SVR), random forest (RF) and extreme gradient boosting trees (XGBoost), are considered for this purpose. Besides evaluation of different machine learning approaches, we further optimize the hyper-parameters and feature set of the different hybrid-FR models. We selected these machine learning algorithms, because they have been used already in other models, e.g. SVRs in VMAF, RF in P.1203/1204.3 and extreme gradient boosting trees to predict the number of video encoding passes in [8]. Other models are possible, however, due to the still low number of training samples within the used databases, other models like neural networks are out of the scope.

In case of RF and XGBoost-based models, only the **number**

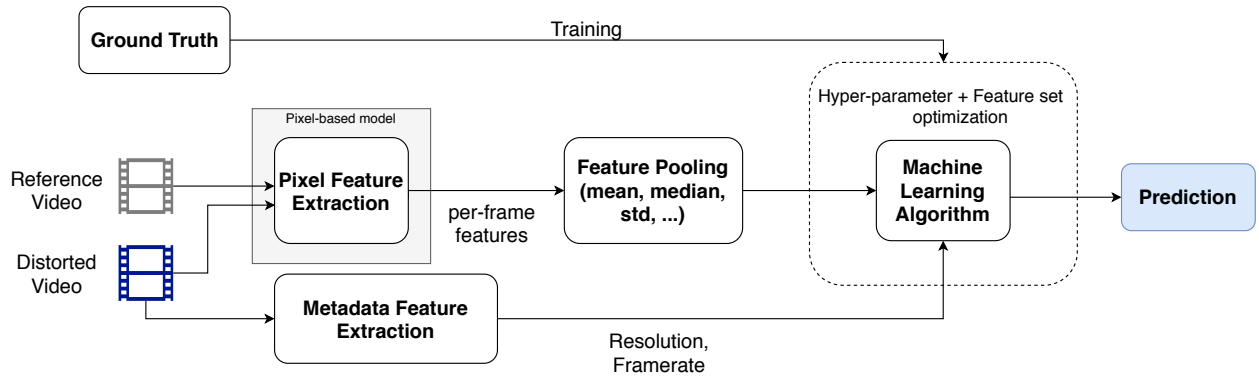


Figure 1: General Machine Learning Pipeline, indicating the involved steps from pixel-based feature extraction based on any kind of pixel-based model, over temporal feature pooling to the final training of the included machine learning algorithm.

of trees parameter was considered for hyper-parameter optimization. Otherwise, default values were used for all other parameters as included in the scikit-learn and xgboost implementation of RF and XGBoost, respectively. For the proof-of-concept provided with this paper, no hyper-parameter optimization for the SVR-based approach was employed. For SVR, we used the radial basis function (RBF) as the kernel and default values for all the parameters that are included in scikit-learn; this is similar to VMAF, BRISQUE and NIQE.

For optimizing the feature set, we perform a full grid search selecting all possible combinations of features, finally selecting the combination of features that results in the best performance in terms of PCC and root mean square error (RMSE). In our case it results in 255 possible feature combinations, where 6 features are based on VMAF's pixel-based calculations and 2 result from our added meta-data, namely resolution and framerate. The hyper-parameter and feature set optimization is a joint optimization process where for every value of the **number of trees parameter**, all possible feature combinations were tested and finally the **(number of trees; feature set)** that results in the best performance is selected. We considered 20 different values for the **number of trees** parameter ranging between 1 and 101 with a step size of 5.

Other parameters of RF and XGBoost can be optimized in a similar manner. In our case, some of the parameters were checked with several pre-tests, and it was finally decided to use the default values, as the optimization of additional parameters did not show any significant improvement of the overall performance of the models. For each of the aforementioned variations of parameters, a separate model is trained and evaluated, considering the prediction performance of the validation set.

The training-validation ratio for all three machine learning algorithms was chosen to be 50:50, ensuring that none of the common sources are used in training, so that the model variants are validated with completely unknown videos.

Table 1 shows the results using the best combination of **(number of trees; feature set)**. In addition, in Table 2 summarizes the performance using a **(number of trees; feature set)** set with as few values as possible for both parameters, with comparable performance with the best case. For the case of SVR, even with just 4 features, the performance is comparable to the best case.

The best combination is 26 trees with 4 features for the RF case, and 101 trees with 4 features for the XGBoost-based model. Even with only 6 trees and 4 features, the RF model has comparable performance with the best RF case. Similarly, a model with 71 trees and 4 features shows comparable performance with the best case for the XGBoost based model. All these best performing cases included resolution and framerate as features. We further perform a detailed feature relevance analysis by counting the number of occurrence of features which is detailed in Figure 2. It can be observed that all hybrid-VMAF instances outperform the retrained VMAF significantly, in terms of all applied performance criteria, namely Pearson correlation coefficient (PCC), Spearman Rank Order Correlation (SROCC), Kendall rank correlation coefficient and Root Mean Squared Error (RMSE). This demonstrates the validity of our approach and also the suitability of the used machine learning algorithms to develop such models. In addition to the performance metrics, a significance analysis was performed according to ITU-T P.1401 [10].

Besides VMAF, we have compared model performance with a number of further metrics, as shown in Tables 1 and 2. It should be noted that for BRISQUE and NIQE, the performance numbers reported in Tables 1 and 2 result after retraining, as described in the AVT-VQDB-UHD-1 study [23].

In addition to the performance analysis in terms of correlation and RMSE, we also analyzed the number of times of occurrence of all the features for the top 100 performing **(number of trees; feature set)** set. It can be seen from Figure 2 that for all the three cases (SVR, RF, XGBoost), resolution and framerate occurred the highest number of times. Also, in the best performing **(number of trees; feature set)** set in both Tables 1 and 2, resolution and framerate were part of the feature set for all three ML algorithms. This further shows that the additional meta-data-based input features, resolution and framerate, indeed plays an important role in improving the performance of FR models. It further indicates that only a small amount of additional data is required to develop a hybrid model variant with good overall performance. As mentioned before, we also evaluated the inclusion of bitrate and video codec as meta-data features, but no improvement was observed; hence these meta-data features were removed for our proposed hybrid model framework.

Conclusion

We evaluated the performance of SoA video quality models with the focus towards pixel-based models. The existing pixel-

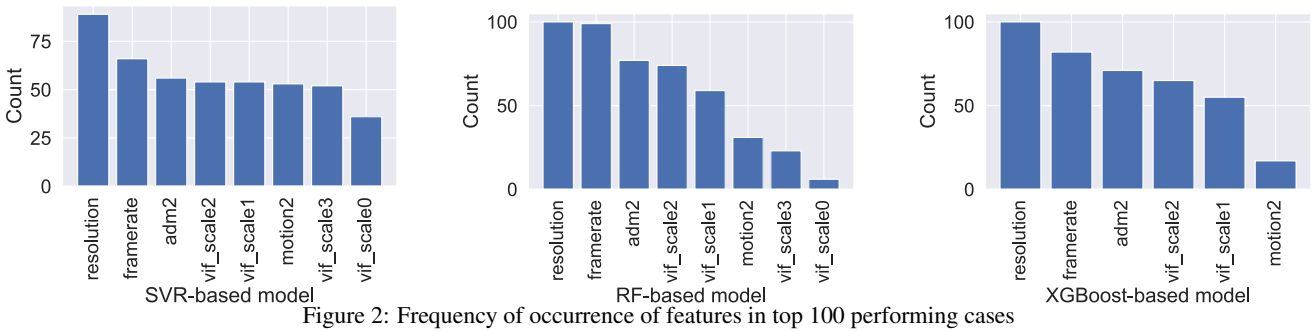


Figure 2: Frequency of occurrence of features in top 100 performing cases

Table 1: Performance comparison between Hybrid-VMAF and other SoA Objective Metrics (considering the best performing feature combination)

Metric	RMSE	PCC	SROCC	Kendall	#Tree	#Feature
VMAF	0.592	0.807	0.811	0.624	NA	NA
BRISQUE	0.641	0.813	0.833	0.646	NA	NA
NIQE	1.006	0.393	0.387	0.265	NA	NA
PSNR	1.004	0.313	0.491	0.352	NA	NA
SSIM	0.871	0.497	0.580	0.418	NA	NA
MS-SSIM	0.832	0.559	0.581	0.421	NA	NA
ADM2	0.598	0.803	0.806	0.615	NA	NA
VIFP	0.789	0.618	0.612	0.449	NA	NA
VMAF (50:50 retraining)	**0.588**	**0.849**	**0.870**	**0.690**	NA	**6**
Hybrid-VMAF (SVR)	**0.397**	**0.939**	**0.929**	**0.774**	NA	**5**
Hybrid-VMAF (RF)	**0.434**	**0.921**	**0.918**	**0.756**	**26**	**4**
Hybrid-VMAF (XGBoost)	**0.433**	**0.924**	**0.927**	**0.772**	**101**	**4**

Table 2: Performance comparison between Hybrid-VMAF and other SoA Objective Metrics (considering lowest number of trees and features with comparable performance as the best case)

Metric	RMSE	PCC	SROCC	Kendall	#Tree	#Feature
VMAF	0.592	0.807	0.811	0.624	NA	NA
BRISQUE	0.641	0.813	0.833	0.646	NA	NA
NIQE	1.006	0.393	0.387	0.265	NA	NA
PSNR	1.004	0.313	0.491	0.352	NA	NA
SSIM	0.871	0.497	0.580	0.418	NA	NA
MS-SSIM	0.832	0.559	0.581	0.421	NA	NA
ADM2	0.598	0.803	0.806	0.615	NA	NA
VIFP	0.789	0.618	0.612	0.449	NA	NA
VMAF (50:50 retraining)	**0.588**	**0.849**	**0.870**	**0.690**	NA	**6**
Hybrid-VMAF (SVR)	**0.438**	**0.930**	**0.913**	**0.744**	NA	**4**
Hybrid-VMAF (RF)	**0.442**	**0.919**	**0.920**	**0.751**	**6**	**4**
Hybrid-VMAF (XGBoost)	**0.438**	**0.921**	**0.925**	**0.769**	**71**	**4**

based models may not implicitly or explicitly include features that are capable of modeling the effect of resolution and framerate scaling on perceived video quality. To overcome such limitations, in this paper we propose a concept to include meta-data like resolution and framerate explicitly as additional features to the existing pixel-based models, to more accurately address their effects and thereby increase the prediction accuracy of corresponding video quality models. Although we consider VMAF to demonstrate the validity of our approach, this approach can also be used for reduced- and no-reference models. We performed several evaluation experiments to analyze the performance of several machine learning algorithm used within our proposed model pipeline. Our results show that the inclusion of meta-data improves the performance in all cases. In addition, the concept also shows that the performance of the existing models can be enhanced in a rather simple manner compared to developing completely new pixel-based models. In future extensions, we will investigate our proposed hybrid model extension considering different application scopes, e.g. considering 360° or gaming video quality prediction.

Acknowledgment

EFRE
EUROPE FOR THURINGIA
EUROPEAN REGIONAL DEVELOPMENT FUND
EUROPEAN UNION

References

[1] 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.

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

[3] S. Fremerey et al. “Subjective Test Dataset and Meta-data-based Models for 360° Streaming Video Quality”. In: *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. 2020, pp. 1–6.

[4] Z. Gao, E. Edirisinghe, and S. Chesnokov. “Image Super-Resolution Using CNN Optimised By Self-Feature Loss”. In: *2019 IEEE International Conference on Image Processing (ICIP)*. 2019, pp. 2816–2820.

[5] 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.

[6] S. Göring, R. Rao, and A. Raake. “nofu - A Lightweight No-Reference Pixel Based Video Quality Model for Gaming Content”. In: *2019 Eleventh International Conference on Quality of*

Multimedia Experience (QoMEX) (QoMEX 2019). Berlin, Germany, June 2019.

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

[8] S. Göring, R. R. R. Rao, and A. Raake. "Prenc — Predict Number of Video Encoding Passes with Machine Learning". In: *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. 2020, pp. 1–6.

[9] ITU-T. *J.143 : User requirements for objective perceptual video quality measurements in digital cable television*. Tech. rep. International Telecommunication Union, 2000.

[10] ITU-T. *P.1401 : Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*. Tech. rep. Int. Telecommunication Union, 2014.

[11] 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.

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

[13] S. Li et al. "Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments". In: *IEEE Transactions on Multimedia* 13.5 (2011), pp. 935–949.

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

[15] P. Madhusudana et al. *Subjective and Objective Quality Assessment of High Frame Rate Videos*. July 2020.

[16] A. Mittal, A. K. Moorthy, and A. C. Bovik. "No-Reference Image Quality Assessment in the Spatial Domain". In: *IEEE Transactions on Image Processing* 21.12 (2012), pp. 4695–4708.

[17] A. Mittal, R. Soundararajan, and A. C. Bovik. "Making a "Completely Blind" Image Quality Analyzer". In: *IEEE Signal Processing Letters* 20.3 (2013), pp. 209–212.

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

[19] Netflix. *VMAF 4K included*. [Online; 07.09.2018]. 2018.

[20] A. Raake et al. "A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1". In: *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. 2017, pp. 1–6.

[21] A. Raake et al. "Multi-Model Standard for Bitstream-, Pixel-Based and Hybrid Video Quality Assessment of UHD/4K: ITU-T P.1204". In: *IEEE Access* 8 (2020), pp. 193020–193049.

[22] 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.

[23] R. Rao et al. "AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1". In: *21st IEEE International Symposium on Multimedia (IEEE ISM)*. 2019, pp. 1–8.

[24] R. Rao et al. "Bitstream-based Model Standard for 4K/UHD: ITU-T P.1204.3 – Model Details, Evaluation, Analysis and Open Source Implementation". In: *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. Athlone, Ireland, May 2020.

[25] W. Robitza, M. Garcia, and A. Raake. "A modular HTTP adaptive streaming QoE model — Candidate for ITU-T P.1203 ("P.NATS")". In: *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. 2017, pp. 1–6.

[26] 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.

[27] H. R. Sheikh and A. C. Bovik. "Image information and visual quality". In: *IEEE Transactions on image processing* 15.2 (2006), pp. 430–444.

[28] A. Takahashi, D. Hands, and V. Barriac. "Standardization activities in the ITU for a QoE assessment of IPTV". In: *IEEE Communications Magazine* 46.2 (2008), pp. 78–84.

[29] Z. Wang, E. P. Simoncelli, and A. C. Bovik. "Multiscale structural similarity for image quality assessment". In: *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*. Vol. 2. IEEE. 2003, pp. 1398–1402.

[30] Z. Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE transactions on image processing* 13.4 (2004), pp. 600–612.

[31] F. Yang and S. Wan. "Bitstream-based quality assessment for networked video: a review". In: *IEEE Communications Magazine* 50.11 (2012), pp. 203–209.

[32] T. Zinner et al. "Impact of frame rate and resolution on objective QoE metrics". In: *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*. 2010, pp. 29–34.

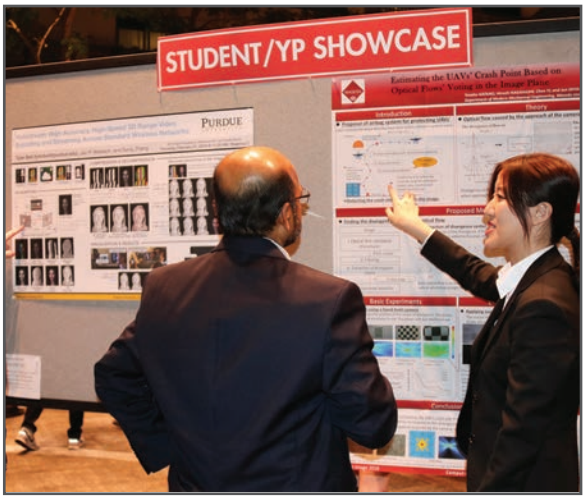
JOIN US AT THE NEXT EI!

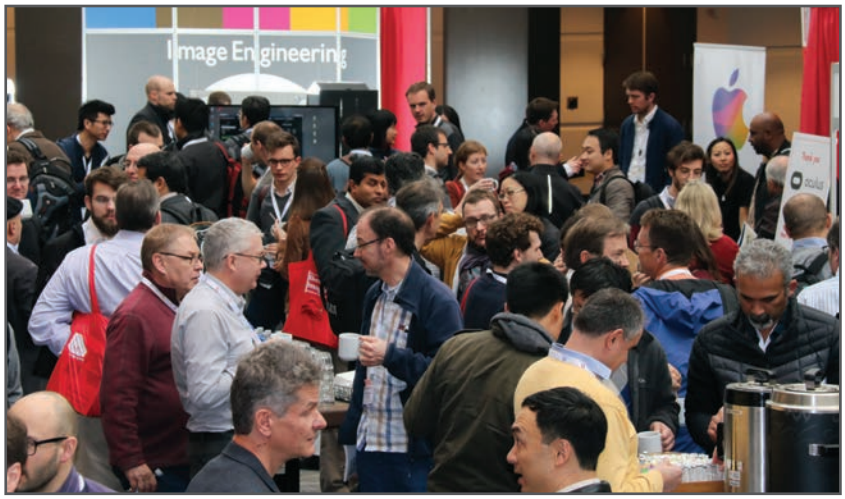
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org