

High Frame Rate Video Quality Assessment using VMAF and Entropic Differences

Pavan C Madhusudana¹, Neil Birkbeck², Yilin Wang², Balu Adsumilli², Alan C. Bovik¹

¹The University of Texas at Austin, TX, USA. ²Google, Mountain View, CA, USA

E-mail : pavancm@utexas.edu, {birkbeck, yilin, badsumilli}@google.com, bovik@ece.utexas.edu

Abstract—The popularity of streaming videos with live, high-action content has led to an increased interest in High Frame Rate (HFR) videos. In this work we address the problem of frame rate dependent Video Quality Assessment (VQA) when the videos to be compared have different frame rate and compression factor. The current VQA models such as VMAF have superior correlation with perceptual judgments when videos to be compared have same frame rates and contain conventional distortions such as compression, scaling etc. However this framework requires additional pre-processing step when videos with different frame rates need to be compared, which can potentially limit its overall performance. Recently, Generalized Entropic Difference (GREED) VQA model was proposed to account for artifacts that arise due to changes in frame rate, and showed superior performance on the LIVE-YT-HFR database which contains frame rate dependent artifacts such as judder, strobing etc. In this paper we propose a simple extension, where the features from VMAF and GREED are fused in order to exploit the advantages of both models. We show through various experiments that the proposed fusion framework results in more efficient features for predicting frame rate dependent video quality. We also evaluate the fused feature set on standard non-HFR VQA databases and obtain superior performance than both GREED and VMAF, indicating the combined feature set captures complimentary perceptual quality information.

Index Terms—high frame rate, video quality assessment, full reference, entropy, generalized Gaussian distribution

I. INTRODUCTION

The explosion of internet usage has resulted in videos becoming more mainstream and forming an important component in day-to-day lives. With videos commanding a significant share in internet traffic, considerable effort has been made in improving the video characteristics in terms of dynamic range, spatial resolutions (4K/8K), immersive content (Virtual/Augmented Reality) and extending temporal resolutions by employing higher frame rate (HFR) videos. However, there has been much less emphasis placed on increasing frame rates, with many existing display devices and content compatible only upto 60 frames per second (fps).

With the recent surge in streaming content pertaining to sports, high motion and live action, there has been a renewed interest in HFR videos. It is a common belief that using more number of frames results in smoother variation of motion content, thus higher frame rates results in better quality videos. However, there have been very few works which comprehensively evaluate these notions. Existing legacy Video Quality Assessment (VQA) databases such as LIVE-VQA [1], CSIQ-VQA [2] only contain videos which have reference and

distorted versions at the same frame rate. This restricts the analysis of distortions that arise due to frame rate changes. Recent publications of HFR-VQA databases such as Waterloo-HFR [3], BVI-HFR [4] and LIVE-YT-HFR [5] has rekindled interest in the problem of HFR-VQA.

Perceptual VQA models are an essential component of many video streaming applications such as YouTube, Netflix, Hulu etc. where video quality is objectively evaluated at a large scale. VQA models can be categorized into three divisions: Full Reference (FR) models which have access to pristine high quality reference [6]–[10], Reduced Reference (RR) models which use limited reference information [11], [12] and No Reference (NR) models which use only the distorted version for quality prediction [13]–[15]. In this paper we consider the problem of VQA when reference and distorted sequences can possibly have different frame rates, thus it falls under the category of FR and RR VQA methods.

Prior works addressing the problem of VQA in HFR domain are very limited. In [16] a motion smoothness measure is introduced to quantify quality changes that occur due to frame rate variations. Zhang *et al.* [17] proposed Frame Rate Quality Metric (FRQM) which uses differences between the wavelet coefficients of reference and distorted videos to measure quality, and achieved competitive performance on the BVI-HFR dataset. Although the above methods account for artifacts arising due to frame rate changes, their performance deteriorates in the presence of compression, thus limiting their applicability. In [18] a model based on entropic differences in spatial and temporal band-pass domain was proposed to capture frame rate artifacts in the presence of compression, and achieved superior performance on the LIVE-YT-HFR database.

In this work we propose a simple fusion framework to combine features from VMAF [19] and GREED [18] models. This principle is motivated from [20] where similar idea was presented to combine VMAF features with SpEED [21] VQA model. VMAF has superior correlation with subjective judgments on existing FR-VQA databases. However VMAF performance is limited when used with videos of different frame rates. Since these videos contain dominant temporal artifacts, lower VMAF performance indicates that the temporal component in VMAF is insufficient to capture frame rate artifacts. On the other hand, GREED achieves good performance when used with videos of different frame rates, suggesting the superiority of temporal features present in GREED. Thus

fusing these features can result in a robust model that can perform well on both HFR and non-HFR VQA databases.

The rest of the paper is organized as follows. In Section II we provide a brief background about GREED and VMAF, and discuss the proposed feature integration procedure. In Section III we report and analyze various experimental results, and provide some concluding remarks in Section IV.

II. BACKGROUND

A. GREED Description

A brief description of GREED [18] model and its components is provided in this section. GREED model is motivated from the statistical deviations observed between the distributions of band-pass coefficients of different frame rates. The video $V(\mathbf{x}, t)$ ($\mathbf{x} = (x, y)$) is subjected to temporal band-pass filtering by a bank of K 1D filters denoted by b_k for $k \in \{1, \dots, K\}$

$$B_k(\mathbf{x}, t) = V(\mathbf{x}, t) * b_k(t) \quad \forall k \in \{1, \dots, K\}, \quad (1)$$

where B_k denotes band-pass response of k^{th} filter. The coefficients of B_k show heavy tailed distribution which can be well modeled using Generalized Gaussian Distribution (GGD) which is given by

$$f(x; \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left(\frac{|x|}{\alpha}\right)^\beta\right)$$

where $\Gamma(\cdot)$ is the gamma function:

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx.$$

Here α and β are parameters of GGD. The band-pass coefficients are partitioned into non-overlapping patches of size $\sqrt{M} \times \sqrt{M}$, which are indexed by $p \in \{1, 2, \dots, P\}$. Let B_{kpt} denote the vector of band-pass coefficients at frame t , patch p and band-pass filter k . The band-pass coefficients B_{kpt} are allowed to pass through a Gaussian channel to model perceptual imperfections such as neural noise [8], [12] resulting in a noisy version \tilde{B}_{kpt} which can also be approximately modeled using a GGD. The main idea here is to compare the entropy $h(\tilde{B}_{kpt})$ of reference and distorted versions for evaluating quality. The expression for GGD entropy of random variable $X \sim GGD(0, \alpha, \beta)$ is given by:

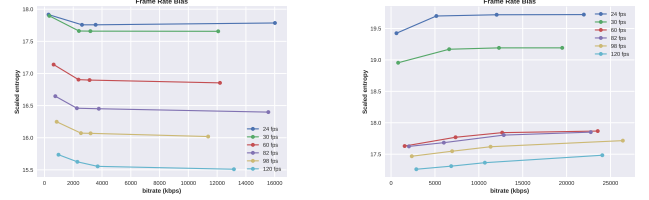
$$h(X) = \frac{1}{\beta} - \log\left(\frac{\beta}{2\alpha\Gamma(1/\beta)}\right). \quad (2)$$

$h(\tilde{B}_{kpt})$ is calculated using (2) for each patch by estimating parameters α and β using kurtosis matching procedure detailed in [22], [23].

Temporal GREED: To make entropies more local and provide numerical stability to regions with small variances, $h(\tilde{B}_{kpt})$ is pre-multiplied by a scaling factor given as:

$$\epsilon_{kpt} = [\log(1 + \sigma^2(\tilde{B}_{kpt}))]h(\tilde{B}_{kpt})$$

Subsampling a signal as well as compression leads to entropy modification. However the entropy change due to frame rate



(a) Content - bouncyball

(b) Content - leaves

Fig. 1. Plot of scaled entropies against bitrate.

dominates that occurring due to compression as can be observed from Fig. 1 where entropy values remain nearly same regardless of compression factor. This is referred to as frame rate bias of entropy and this makes entropy difference inadequate to account for compression distortions. To overcome this, reference video R is temporally downsampled to have the same frame rate of distorted video D resulting in another sequence called Pseudo Reference PR . Given these signals, Temporal GREED is defined as

$$\text{TGREED}_{kt} = \frac{1}{P} \sum_{p=1}^P \left| \left(1 + |\epsilon_{kpt}^D - \epsilon_{kpt}^{PR}| \right) \frac{\epsilon_{kpt}^R + 1}{\epsilon_{kpt}^{PR} + 1} - 1 \right| \quad (3)$$

TGREED can be interpreted as a weighted entropic difference, where the ratio term which is dependent on frame rate of R and D acts as a weighting factor. From Fig. 1 it can be observed that the entropy separation among various frame rates is dependent on the content. This makes the ratio term sensitive to content. The absolute difference term captures compression artifacts since PR is lossless version of D at the same frame rate.

Spatial GREED: For capturing spatial distortions, each frame is subjected to spatial band-pass filtering by local Mean Subtracted (MS) filtering scheme similar to [21]. Similar to temporal band-pass coefficients, MS coefficients also can be modeled using GGD and scaled entropy can be calculated as

$$\theta_{pt} = [\log(1 + \sigma^2(\tilde{V}_{pt}^{MS}))]h(\tilde{V}_{pt}^{MS})$$

where \tilde{V}_{pt}^{MS} denotes MS coefficients of frame t and patch p passed through neural noise. There does not exist any frame rate bias since spatial entropies were calculated using only single frames and are frame rate agnostic. Spatial GREED (SGREED) is defined as :

$$\text{SGREED}_t = \frac{1}{P} \sum_{p=1}^P |\theta_{pt}^D - \theta_{pt}^R|. \quad (4)$$

B. VMAF description

VMAF uses multiple elementary quality indices and fuses them using a Support Vector Regressor (SVR) [19]. The first step in VMAF is extracting the following features : DLM [24], VIF [8] and absolute luma differences between consecutive frames (Temporal Information - TI). The DLM feature captures detail losses while VIF measures image fidelity based on

Algorithm 1 GREED-VMAF Algorithm

Input: reference video R , distorted video D **Output:** GREED-VMAF features

- 1: Temporal subsample R to get PR
 - 2: Calculate DLM(PR,D)
 - 3: Calculate VIF(PR,D) - 4 spatial scales
 - 4: Scale $S = \{4, 5\}$, band-pass filter bank $b_k : k \in \{1, \dots, 7\}$
 - 5: **for each** $s \in S$ **do**
 - 6: Calculate SGREED from (4)
 - 7: **for each** b_k **do**
 - 8: Calculate TGREED $_k$ from (3)
 - 9: **end for**
 - 10: **end for**
 - 11: Concatenate DLM, VIF, SGREED, TGREED to obtain GREED-VMAF features.
-

Natural Scene Statistics (NSS). VIF is computed at four scales resulting four VIF features. TI aims to capture the temporal artifacts which are quantified through luma differences. This results in VMAF containing six features overall. TI is the only feature in VMAF that accounts for temporal quality evaluation.

C. Feature Integration

Although GREED achieves high correlations against human judgements on LIVE-YT-HFR database, the performance is sensitive to the differences between the frame rates of reference and distorted sequences, where lower frame rate videos have lower correlations when compared to high frame rate videos. This behavior can be observed in Table II. Another shortcoming of GREED is that it underperforms in the presence of other types of distortions apart from frame rate and compression artifacts. This is evidenced in Table III where GREED is evaluated on databases such as LIVE-VQA [1], CSIQ-VQA [2] etc. VMAF also has a drawback in terms of capturing temporal distortions as TI feature is a basic measure and complex temporal artifacts such as ghosting, flickering etc. are not effectively represented by TI.

To overcome these limitations, we propose to combine the features from both GREED and VMAF to exploit the advantages offered by both models. Since GREED achieves superior performance for temporal artifacts, we replace the TI feature in VMAF by all GREED features. We retain the 5 spatial VMAF features (DLM and VIF from 4 scales) as they sufficiently capture spatial quality. GREED features are computed at 2 scales $s = 4, 5$ where entropic differences are computed after downscaling both spatial dimensions by 2^s times. For calculating TGREED Biorthogonal-2.2 band-pass filter with 3 levels of wavelet packet decomposition [25] was used. Since low-pass subband is ignored, this results in 7 TGREED features for each scale from respective temporal subbands and a total of 14 TGREED features from both scales. Although spatial components of VMAF account for spatial distortions, including SGREED features was experimentally observed to further improve performance. Thus the proposed model GREED-VMAF consists a total of 21 features alto-

TABLE I
PERFORMANCE COMPARISON OF GREED AGAINST DIFFERENT FR ALGORITHMS ON THE LIVE-YT-HFR DATABASE. THE REPORTED NUMBERS ARE MEDIAN VALUES FROM 200 ITERATIONS OF RANDOMLY CHOSEN TRAIN-TEST SETS. IN EACH COLUMN FIRST AND SECOND BEST MODELS ARE BOLDFACED.

	SROCC \uparrow	KROCC \uparrow	PLCC \uparrow	RMSE \downarrow
PSNR	0.7802	0.5934	0.7481	7.75
SSIM [6]	0.5566	0.4042	0.5418	9.99
MS-SSIM [7]	0.5742	0.4135	0.5512	10.01
ST-RRED [12]	0.6394	0.4516	0.6073	9.58
SpEED [21]	0.6051	0.4437	0.5206	10.28
VMAF [19]	0.7782	0.5918	0.7419	8.10
GREED	0.8822	0.7046	0.8869	5.48
GREED-VMAF	0.8658	0.6840	0.8723	5.89

gether. The entire feature extraction procedure for GREED-VMAF is summarized in Algorithm 1.

III. EXPERIMENTS AND RESULTS

A. Experimental Settings

Compared Methods : We choose 3 FR-IQA methods: PSNR, SSIM [6] and MS-SSIM [7], and 3 FR-VQA models: ST-RRED [12], SpEED [21] and VMAF [19] for performance comparison. The above mentioned models require reference and distorted versions to have same frame rates, thus for the cases where the frame rates were different the distorted video was temporally upsampled to match the frame rate of reference video by frame duplication. Although the frame rates can be matched by temporally downsampling the reference video as well, we avoid this option as it can potentially introduce temporal artifacts in the reference video which is not desirable.

Evaluation Criteria : VQA models are compared in terms Spearman's rank order correlation coefficient (SROCC), Kendall's rank order correlation coefficient (KROCC), Pearson's linear correlation coefficient (PLCC), and root mean squared error (RMSE). PLCC and RMSE are calculated after passing the predicted scores through a four-parameter logistic non-linearity as described in [26]:

$$Q(x) = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \exp\left(-\left(\frac{x - \beta_3}{|\beta_4|}\right)\right)}. \quad (5)$$

B. Performance Analysis on LIVE-YT-HFR Dataset

Since GREED-VMAF is a training based model, the LIVE-YT-HFR database is divided into 70%, 15% and 15% subsets corresponding to training, validation and testing respectively. The validation subset was used to determine hyperparameters of SVR. Additionally we ensured that there was no content overlap between the subsets. To avoid any performance bias towards the choice of training set, the experiment was repeated 200 times with random train-test splits and median performance is reported. Since only the spatial features from VMAF are employed in GREED-VMAF, and since VMAF framework requires the compared videos to have same frame rate, we match the frame rate by temporally subsampling the

TABLE II

PERFORMANCE COMPARISON OF GREED AGAINST VARIOUS FR METHODS FOR INDIVIDUAL FRAME RATES ON THE LIVE-YT-HFR DATABASE. THE REPORTED NUMBERS ARE MEDIAN VALUES OVER 200 ITERATIONS OF RANDOMLY CHOSEN TRAIN-TEST SETS. IN EACH COLUMN THE BEST AND SECOND BEST VALUES ARE MARKED IN BOLDFACE.

	24 fps		30 fps		60 fps		82 fps		98 fps		120 fps		Overall	
	SROCC \uparrow	PLCC \uparrow	SROCC \uparrow	PLCC \uparrow	SROCC \uparrow	PLCC \uparrow	SROCC \uparrow	PLCC \uparrow	SROCC \uparrow	PLCC \uparrow	SROCC \uparrow	PLCC \uparrow	SROCC \uparrow	PLCC \uparrow
PSNR	0.541	0.487	0.564	0.535	0.753	0.694	0.771	0.771	0.821	0.783	0.741	0.736	0.780	0.748
SSIM [6]	0.266	0.222	0.283	0.189	0.382	0.302	0.371	0.362	0.537	0.497	0.867	0.833	0.556	0.541
MS-SSIM [7]	0.305	0.260	0.296	0.238	0.416	0.338	0.439	0.393	0.578	0.561	0.706	0.696	0.574	0.551
ST-RRED [12]	0.305	0.275	0.296	0.206	0.612	0.613	0.584	0.513	0.650	0.604	0.755	0.696	0.639	0.607
SpEED [21]	0.432	0.273	0.410	0.233	0.439	0.292	0.546	0.390	0.578	0.471	0.758	0.739	0.605	0.520
VMAF [19]	0.250	0.368	0.362	0.471	0.630	0.680	0.734	0.793	0.860	0.868	0.818	0.816	0.779	0.742
GREED	0.727	0.822	0.702	0.843	0.732	0.840	0.818	0.896	0.864	0.891	0.888	0.895	0.882	0.887
GREED-VMAF	0.748	0.805	0.743	0.833	0.773	0.836	0.786	0.880	0.860	0.899	0.881	0.903	0.865	0.872

TABLE III

SROCC PERFORMANCE COMPARISON ON MULTIPLE VQA DATABASES. THE REPORTED NUMBERS ARE MEDIAN VALUES FROM EVERY POSSIBLE COMBINATION OF TRAIN-TEST SPLITS WITH 80% OF CONTENT USED FOR TRAINING. IN EACH COLUMN THE BEST AND SECOND BEST MODELS ARE BOLDFACED.

	LIVE-VQA	LIVE-mobile	CSIQ-VQA
PSNR	0.711	0.788	0.579
SSIM [6]	0.802	0.798	0.705
MS-SSIM [7]	0.830	0.800	0.757
ST-RRED [12]	0.826	0.882	0.813
SpEED [21]	0.801	0.886	0.743
VMAF [19]	0.794	0.897	0.618
GREED	0.750	0.863	0.780
GREED-VMAF	0.784	0.904	0.907

reference video (PR video) instead of temporally upsampling distorted version as we observed the spatial artifacts were better captured by employing this design. SVR with linear kernel was used to learn mapping from features to quality scores.

In Table I the performance of GREED-VMAF against other models is compared. GREED and GREED-VMAF outperform other models by a large margin. In Table II the LIVE-YT-HFR database is subdivided into sets containing videos of same frame rate, and the performance is individually analyzed on each frame rate. Here as well GREED and GREED-VMAF have superior correlation when compared to other models across each frame rate. Although we observe a small drop in performance of GREED-VMAF in Table I, its performance is superior to GREED for individual frame rates as shown in Table II, particularly in low frame rate sets such as 24 and 30 fps. This makes performance of GREED-VMAF be less sensitive to the choice of distorted frame rate and provide more stable quality predictions than GREED.

C. Performance Comparison on other VQA Databases

In this subsection we evaluate the generalizability of GREED-VMAF features by comparing its performance on three popular VQA databases LIVE-VQA [1], LIVE-mobile [27] and CSIQ-VQA [2]. These databases contain both the reference and distorted videos at the same frame rate, thus the TGREED term in expression (3) will only depend on the absolute difference of scaled entropies since the ratio term simplifies to one. We divide the database into sets of 80% and 20% for training and testing respectively. We also ensured

that there does not exist any overlap of contents between training and testing subsets. Further, to avoid any performance bias towards the choice of train-test contents, we repeat the above procedure for all possible train-test combinations, and the median SROCC performance is reported in Table III. For these datasets, SVR with radial basis function (RBF) kernel was used to learn mapping from features to quality scores. From the Table it maybe observed that GREED-VMAF outperforms GREED across all databases indicating that the combined set of features are more generalizable than GREED. Notably GREED-VMAF outperforms VMAF and achieves top performance among the compared models on LIVE-mobile and CSIQ-VQA datasets. Since GREED-VMAF performance is higher than both GREED and VMAF, it can be inferred that the respective individual features capture complementary perceptual quality information, and thus fusing them leads to better correlation with subjective judgments.

IV. CONCLUSION

We proposed a simple extension to existing VQA models GREED and VMAF by combining the spatial features of VMAF with spatio-temporal features of GREED. This models exploits the advantages of both the models: can be used when reference and distorted videos have different frame rates, with no additional temporal pre-processing and can effectively capture other artifacts apart from frame rate and compression artifacts. We conducted a holistic evaluation of the proposed model on LIVE-YT-HFR database and observed that GREED-VMAF provides better correlation with perceptual judgments when videos of fixed frame rates were analyzed. We also evaluated the generalizability of GREED-VMAF on multiple fixed frame rate VQA databases and observed to have superior performance when compared to both GREED and VMAF. As part of future work we wish to further evaluate the generalizability of GREED-VMAF on additional VQA datasets such as NFLX, EPFL-Polimi etc. which is not reported in this work.

REFERENCES

- [1] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, June 2010.
- [2] P. V. Vu and D. M. Chandler, "Vis3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *Journal of Electronic Imaging*, vol. 23, no. 1, p. 013016, 2014.

- [3] R. M. Nasiri, J. Wang, A. Rehman, S. Wang, and Z. Wang, "Perceptual quality assessment of high frame rate video," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2015, pp. 1–6.
- [4] A. Mackin, F. Zhang, and D. R. Bull, "A study of high frame rate video formats," *IEEE Trans. Multimedia*, vol. 21, no. 6, pp. 1499–1512, 2018.
- [5] P. C. Madhusudana, X. Yu, N. Birkbeck, Y. Wang, B. Adsumilli, and A. C. Bovik, "Subjective and objective quality assessment of high frame rate videos," in *arXiv preprint arXiv:2007.11634*, 2020.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.
- [7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Asilomar Conf. Signals Syst. Comput.*, vol. 2, Nov 2003, pp. 1398–1402 Vol.2.
- [8] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [9] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [10] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb 2010.
- [11] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.
- [12] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, April 2013.
- [13] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [14] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Sig. Process. Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [15] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, March 2014.
- [16] R. M. Nasiri, Z. Duanmu, and Z. Wang, "Temporal motion smoothness and the impact of frame rate variation on video quality," in *IEEE Int'l Conf. Image Process.*, 2018, pp. 1418–1422.
- [17] F. Zhang, A. Mackin, and D. R. Bull, "A frame rate dependent video quality metric based on temporal wavelet decomposition and spatiotemporal pooling," in *IEEE Int'l Conf. Image Process.*, Sep. 2017, pp. 300–304.
- [18] P. C. Madhusudana, N. Birkbeck, Y. Wang, B. Adsumilli, and A. C. Bovik, "St-greed: Space-time generalized entropic differences for frame rate dependent video quality prediction," 2020.
- [19] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," <http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html>.
- [20] C. G. Bampis, Z. Li, and A. C. Bovik, "Spatiotemporal feature integration and model fusion for full reference video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2256–2270, 2018.
- [21] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "SpEED-QA: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1333–1337, Sep. 2017.
- [22] H. Soury and M.-S. Alouini, "New results on the sum of two generalized Gaussian random variables," in *IEEE Global Conf. on Signal and Information Process.*, 2015, pp. 1017–1021.
- [23] X. Pan, X. Zhang, and S. Lyu, "Exposing image splicing with inconsistent local noise variances," in *IEEE International Conference on Computational Photography (ICCP)*, 2012, pp. 1–10.
- [24] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image quality assessment by separately evaluating detail losses and additive impairments," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 935–949, 2011.
- [25] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 713–718, 1992.
- [26] VQEG, "Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment," 2000.
- [27] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. De Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652–671, 2012.