SpEED-QA: Spatial Efficient Entropic Differencing for Image and Video Quality

Christos G. Bampis, Praful Gupta, Rajiv Soundararajan, and Alan C. Bovik

***Abstract*—Many image and video quality assessment (I/VQA) models rely on data transformations of image/video frames, which increases their programming and computational complexity. By comparison, some of the most popular I/VQA models deploy simple spatial bandpass operations at a couple of scales, making them attractive for efficient implementation. Here we design reduced-reference image and video quality models of this type that are derived from the high-performance reduced reference entropic differencing (RRED) I/VQA models. A new family of I/VQA models, which we call the spatial efficient entropic differencing for quality assessment (SpEED-QA) model, relies on local spatial operations on image frames and frame differences to compute perceptually relevant image/video quality features in an efficient way. Software for SpEED-QA is available at: http://live.ece.utexas.edu/research/Quality/SpEED_Demo.zip**

***Index Terms*—Entropic differencing, image and video quality assessment (I/VQA), reduced-reference (RR) models.**

I. Introduction

OBJECTIVE image and video quality assessment (I/VQA) models aim to predict visual quality without the need to collect human subjective scores. These models often rely on statistical regularities (viz., natural scene statistics—NSS) that govern natural images and videos. NSS-derived features may be used to quantify deviations from these statistical properties that are predictive of visual impairments. There are three categories of objective I/VQA models: full-reference (FR), reduced-reference (RR) and no-reference (NR). FR models [1], [2] compare possibly distorted signals against entire reference versions of them. RR models [3]–[7] use only a subset of the reference data to predict quality, whereas NR models use only the distorted image/video to measure quality [8]–[10].

Among these three types of IQA/VQA models, RR approaches have received less attention. However, since FR models produce highly correlated quality maps, therefore, it should be possible to subsample their responses in space and/or time with little degradation in performance [11]. Thus, here we seek to design *efficient* RR models that exploit the redundancies inherent in space-time distortions, by forming predictions of visual quality from reduced sets of more descriptive image/video coefficients.

Most objective I/VQA models use a mapping to a different coordinate domain, e.g., a wavelet decomposition [2], [3], [6], [10], [12]–[14] or discrete cosine transform (DCT) [15]–[17], whereas others achieve high performance using simple and local spatial operations [1], [8].

Given the recently observed robust performance of spatio-temporal RR entropic differencing (ST-RRED) on a wide range of VQA databases [14], [18], we chose to investigate whether it is possible to design QA algorithms similar to ST-RRED, but with reduced complexity, and without sacrificing quality prediction accuracy. In particular, we have developed an RR-I/VQA model called spatial efficient entropic differencing for quality assessment (SpEED-QA), which is an efficient NSS-based model that applies local entropic differencing between test and reference signals in the spatial domain. SpEED-QA is very fast, yet it delivers highly competitive performance against other objective I/VQA quality models. We organize this Letter as follows. Section II discusses previous work on RR I/VQA, whereas Sections III and IV introduce the SpEED-QA suite of I/VQA models. Section V describes experimental results and Section VI concludes with future work.

II. Related Work

RR I/VQA models conduct objective quality prediction by computing a set of perceptually relevant RR features from both reference and distorted image/videos, and then comparing them. For example, in [3], the marginal statistics of wavelet coefficients were modeled by a generalized Gaussian distribution (GGD). The RR GGD models of reference and distorted image frames were compared using the Kullback–Leibler divergence (KLD), yielding visual quality predictions. In [5], the Gaussian scale mixture (GSM) model [19] was used to statistically describe the neighboring wavelet coefficients. Divisive normalization was applied to reduce the statistical dependencies between them, yielding a set of approximately Gaussian-distributed marginal wavelet coefficients.

A high-performing RR image quality model, called RRED [6], was successfully extended to VQA in [14]. This model, which does not require motion computation, is called ST-RRED. The RRED framework computes and compares the subsampled conditional entropies of the wavelet coefficients of image/video frames and frame differences. Building on the NSS-based I/VQA work in [2] and [13], a GSM-based information-theoretic model was applied to wavelet coefficients assumed to have been subjected to linear distortion followed by a neural noise process. In ST-RRED, the absolute differences of the local entropies of reference and distorted image/video frames (or

Manuscript received May 8, 2017; revised July 8, 2017; accepted July 10, 2017. Date of publication July 13, 2017; date of current version July 27, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Charles Kervrann. *(Corresponding author: Christos G. Bampis.)*

C. G. Bampis, P. Gupta, and A. C. Bovik are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA (e-mail: bampis@utexas.edu; praful_gupta@utexas.edu; bovik@ece.utexas.edu).

R. Soundararajan is with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangaluru 560012, India (e-mail: rajivs@ece.iisc.ernet.in).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2726542

frame differences), conditioned on the locally estimated variance field, are used to characterize image and video distortions. Then, subband weights are applied to the computed spatial and temporal entropies, which are combined in a product form. The method developed here is based on the RRED framework, and may be viewed as a highly efficient space-domain counterpart.

Similar ideas have also been applied in the DCT domain. In [17], Ma *et al.* modeled image DCT coefficients as GGD, then applied the city-block distance to compare between reference and distorted RR parameters. A learning-based approach was developed in [7], where orientation-selective descriptors were used to predict image quality. A conceptually different approach in [20] is based on the free energy principle.

III. SpEED-IQA: Image Quality Assessment Model

A. Spatial Modeling

Many objective image quality models rely on two ideas. First, a band-pass transformation applied on an image or video frame, such as a DCT [21] or wavelet decomposition [19], [22] yields coefficients that follow a heavy-tailed distribution. Second, the statistics of divisively normalized spatial band-pass image responses tend toward an uncorrelated Gaussian distribution [8], [23]. Distortions modify the Gaussian characteristic, hence objective quality models aim to quantify these statistical irregularities. The divisive normalization step broadly resembles local gain-control mechanisms in biological visual systems [24], [25]. These ideas are effectively modeled in the spatial domain using simple filters [23] in NR algorithms such as BRISQUE [8] and NIQE [26] to achieve high quality prediction performance with very low computational complexity.

RR models also make use of these properties, but they usually deploy mappings to different coordinate domains, like the wavelet decomposition used in RRED [6]. Motivated by the efficiency and successes of space-domain approaches to IQA [8] and the high-performance and theoretical soundness of the RRED framework, we sought to develop an efficient spatial approach to RR quality assessment.

We begin by revisiting NSS regularities in the spatial domain. While GSM models have been successfully used to model natural images in the wavelet/bandpass domain [19], [27], we model blocks of good quality luminance images that have been locally mean subtracted (MS) as following a GSM model. Let $\mathbf{I_{MS}} = (\mathbf{I} - \boldsymbol{\mu})$ be a $M \times N$ image of local MS coefficients, where

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} \mathbf{I}_{k,l}(i,j), \ \boldsymbol{\mu} = [\mu_{i,j}] \tag{1}$$

where $\mathbf{I}$ is the image in the pixel domain, $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$ are spatial indices, M, N are the image height and width, respectively, $w = \{w_{k,l} | k = -K, \ldots, K, l = -L, \ldots L\}$ is a two-dimensional circularly symmetric Gaussian weighting function sampled out to three standard deviations and rescaled to unit volume and $K = L = 3$.

We model a $r \times r$ block of the bandpass responses $\mathbf{I_{MS}}$ as a GSM vector, i.e., a vector expressed as $\mathbf{x} \equiv z\mathbf{U}$, where $\equiv$ denotes equality in probability distribution, $\mathbf{U}$ is an underlying zero-mean Gaussian vector with covariance matrix $\mathbf{C}_u$, and z is a positive scalar random variable called the mixing multiplier. The probability density of the GSM vector is given by

$$p_{\mathbf{x}}(\mathbf{x}) = \int \frac{1}{(2\pi)^{d/2} |z^2 C_u|^{1/2}} \exp\left(\frac{-\mathbf{x}^T C_u^{-1} \mathbf{x}}{z^2}\right) p_z(z) dz.$$

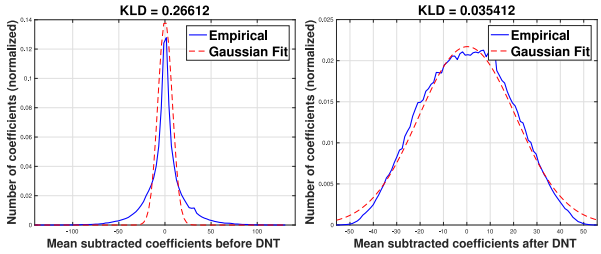

Fig. 1. Marginal histograms of normalized and unnormalized MS field and KLD values of their Gaussian fits.

The maximum likelihood estimate [27] of the scalar multiplier $\hat{z}$ is computed per block using the block covariance matrix $\mathbf{C}_u$, which is estimated using neighboring responses from the MS image $\mathbf{I_{MS}}$. Using $\hat{z}$ to divisively normalize the MS coefficients yields Gaussianized variates $\tilde{x} = x/\hat{z}$ (see Fig. 1). This divisive normalization decreases the mutual information between MS coefficients and is more precise than the variance normalization used in BRISQUE [8], and the weighted sum of squares normalization used in [28]–[30].

B. Single-Scale SpEED-IQA

Here we describe the simplest, single-scale SpEED-IQA index, whereby the computed conditional entropies of spatially corresponding MS block coefficients from the reference and distorted images are compared. By scale, we refer to the size (relative to the original image resolution) that the model is applied to. Let C_{mkr} be the MS coefficients of the reference luminance image in block m and scale k, and C_{mkd} the corresponding MS coefficients for the distorted image. The GSM model in the spatial domain is then described by

$$C_{mkr} = S_{mkr} U_{mkr}, \ C_{mkd} = S_{mkd} U_{mkd} \tag{2}$$

where S_{mkr} and S_{mkd} are non-negative random variables independent of $U_{mkr} \sim \mathcal{N}(0, \mathbf{K}_{U_{kr}})$ and $U_{mkd} \sim \mathcal{N}(0, \mathbf{K}_{U_{kd}})$, respectively. To model neural noise along the visual pathway, these coefficients are further passed through an additive noise channel yielding

$$C'_{mkr} = C_{mkr} + W_{mkr}, \ C'_{mkd} = C_{mkd} + W_{mkd} \tag{3}$$

where $W_{mkr} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}_N)$ and $W_{mkd} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}_N)$.

Conditioning the MS coefficients C'_{mkr} and C'_{mkd} on realizations of S_{mkr} and S_{mkd} yields Gaussian random vectors, and allows for the computation of local entropies h as follows:

$$h(C'_{mkr} | S_{mkr} = s_{mkr}) = \frac{1}{2} \log[(2\pi e)^N |s_{mkr}^2 \mathbf{K}_{U_{kr}} + \sigma_w^2 \mathbf{I}_N|]$$

$$h(C'_{mkd} | S_{mkd} = s_{mkd}) = \frac{1}{2} \log[(2\pi e)^N |s_{mkd}^2 \mathbf{K}_{U_{kd}} + \sigma_w^2 \mathbf{I}_N|].$$

These entropies are further locally weighted [6] by the scalar factors $\gamma_{mkr} = \log(1 + s_{mkr}^2)$ and $\gamma_{mkd} = \log(1 + s_{mkd}^2)$

$$\alpha_{mkr} = \gamma_{mkr} h(C'_{mkr} | S_{mkr} = s_{mkr}) \tag{4}$$

$$\alpha_{mkd} = \gamma_{mkd} h(C'_{mkd} | S_{mkd} = s_{mkd}). \tag{5}$$

This weighting approach enhances the sensitivity of SpEED to spatio-temporal details. The neural noise variance σ_w^2 was fixed to the same small value of 0.1 as was used in [2], [6], [13], [14], following the same rationale as detailed there. After calculating the conditional block entropies of the reference and distorted

TABLE I
IMAGE RESULTS ON LEADING IQA DATABASES

IQA model	LIVE	CSIQ	TID13	TID08	Overall
PSNR	0.8756	0.8058	0.6468	0.5559	0.6740
SSIM [1]	0.9479	0.8756	0.7450	0.7784	0.7979
FSIM [35]	0.9634	0.9242	0.8015	0.8805	0.8602
MS-SSIM [31]	0.9513	0.9133	0.7859	0.8542	0.8428
VIF [2]	0.9636	0.9194	0.6770	0.7491	0.7664
RRED [6]	0.9429	0.9185	0.7726	0.8323	0.8305
SpEED-IQA	0.9430	0.9252	0.7774	0.8419	0.8362
SpEED-IQA-O3	0.9397	0.9211	0.8331	0.8435	0.8617
MS-SpEED-IQA	0.9548	0.9344	0.7906	0.8630	0.8508

(luminance) images, the differences between the entropies of the corresponding blocks are computed and then averaged over all blocks, thereby yielding a single image quality score

$$\text{SpEED-IQA}_k = \frac{1}{M_k}\sum_{m=1}^{M_k} |\alpha_{mkr} - \alpha_{mkd}| \tag{6}$$

where k denotes the scale used in computation and M_k denotes the number of scalars (one for each block) used from the reference and distorted luminance images.

The RR nature of the proposed model is similar to the wavelet-domain RRED model: Only one entropy value per reference and test block is transmitted and differenced. If the order of differencing and averaging is reversed, then a "single-number" (SN) algorithm results, whereby QA is conducted using a single entropy value from the reference image [6]. The entire procedure can be similarly applied to any RGB or opponent-space color channel in place of the luminance channel. This choice will generally be application-specific, although we have found a particular choice to work well (see Section IV-A and Table I).

An effective multiscale version of SpEED-QA may be devised by down-sampling each image (possibly multiple times), then applying single-scale SpEED-IQA to each subsampled block. The values computed at each scale can be combined, e.g., by using the perceptually optimized exponent weights used in MS-structural similarity (SSIM) [31]. This multiscale approach (MS-SpEED-IQA) increases the number of scalars and the execution time, but its prediction performance is even more competitive (see Table I).

IV. SPEED-VQA: VIDEO QUALITY ASSESSMENT MODEL

As we intimated earlier, similar ideas can be applied to the VQA problem. The spatial SpEED-VQA index is obtained by applying SpEED-IQA on every frame of the video sequence. To capture temporal distortions, rather than relying on expensive motion-estimation approaches [32], [33], SpEED-VQA computes conditional entropies on frame-differenced blocks (see Fig. 2). Given each pair of successive luminance frames indexed $i, i+1$, we calculate their differences. Next, we apply the same block-based decorrelation and conditional entropy calculations as on the original luminance frames. As shown in [14], bandpass frame differences nicely follow a GSM model. Thus, we compute the block conditional entropies of the frame-differenced MS coefficients on the reference and distorted video; then difference and (average) pool them. Finally, by computing the product between spatial SpEED-VQA and the corresponding temporal score just arrived at, a SpEED-VQA score is obtained.

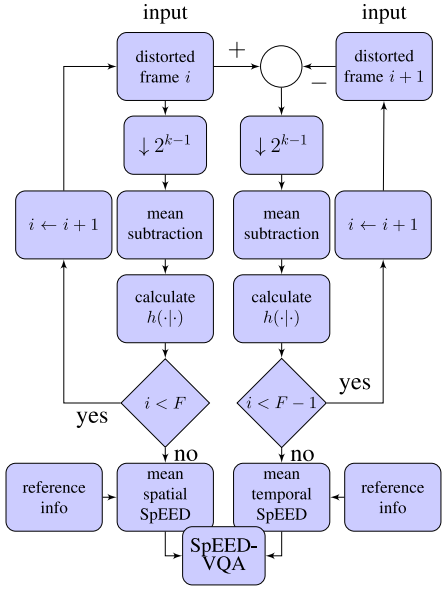

Fig. 2. Overview of SpEED-VQA.

TABLE II
NUMBER OF SCALARS AND EXECUTION TIME COMPARISONS BETWEEN RRED AND SPEED-IQA

IQA model	k	# Scalars	Time	SROCC
RRED	2	64L/2304	0.3138	0.8305
RREDopt	2	64L/2304	0.0842	0.8305
SpEED-IQA	2	64L/2304	0.0587	0.8128
SpEED-IQA	**3**	**16*L*/2304**	**0.0196**	**0.8362**

A weighted (by the number of images) average of time and SROCC was computed across all four image databases. The proposed algorithm is denoted by bold.

In our experiments, we found SpEED-VQA to be most effective when applied at the coarsest scale (after downsampling by a factor of 2 four times), i.e., $k = 5$. A similar observation was made in [14] and may be related to the motion down-shifting phenomena [34].

V. EXPERIMENTS

A. IQA Experiments

We selected several IQA models for comparison: PSNR [1], SSIM [1], MS-SSIM [31], FSIM [35], RRED [6] and the proposed SpEED-IQA models (using a 3×3 block size), and applied all of them on four widely used databases: LIVE [36], CSIQ [37], TID13 [38], and TID08 [39]. As shown in Table I, the proposed SpEED-IQA model performed competitively with RRED, and outperformed the FR models SSIM and visual information fidelity (VIF). Multiscale SpEED-IQA further improved on performance, matching MS-SSIM. It is interesting that using the O3 color channel [40] greatly improved performance on the distortion-specific TID13 database; but not against luminance results in other databases. This may be due to the large number of peculiar color distortions in TID13.

Most importantly, SpEED-IQA performed as well as its wavelet domain precursor (RRED), but with much more computational efficiency. To investigate this claim, Table II shows the data rates (number of scalars in terms of image size L) and execution time (in seconds) required for RRED and SpEED-IQA.

Fig. 3. Prediction of different distortion levels: Left: differential mean opinion score (DMOS) = 68.0166, SpEED-IQA prediction = 17.6358, MS-SSIM prediction = 0.8609; Right: DMOS = 51.4067, SpEED-IQA prediction = 7.9315, MS-SSIM prediction = 0.8220. Note that DMOS increases with worsening perceptual quality over $[0,\ 100]$.

TABLE III
PERFORMANCE OF SPEED-IQA OVER DIFFERENT SCALES

Scale	# Scalars	LIVE	CSIQ	TID13	TID08	Overall
1	L/9	0.8969	0.7643	0.5529	0.5623	0.6287
2	L/36	0.9422	0.9125	0.7466	0.8149	0.8128
3	**L/144**	**0.9430**	**0.9252**	**0.7774**	**0.8419**	**0.8362**
4	L/576	0.9213	0.8629	0.7291	0.7454	0.7766
5	L/2304	0.8838	0.7537	0.6401	0.5887	0.6733

The best result is denoted by bold.

We also reimplemented RRED so that only the best performing subband (band 16) was computed (we refer to this version as RREDopt). An example of the ability of SpEED-IQA to closely predict the monotonic relationship between objective and subjective scores is shown in Fig. 3.

Table II illustrates the major advantages of SpEED-IQA: drastically reduced execution time, using only one fourth of the image data as compared to RRED. SpEED performs very well at lower scales than RRED, which contributes to its lower computational times. Furthermore, its performance is best at scale $k = 3$, whereas at scale 2 it offers slightly lower performance. The full multiscale version of SpEED-IQA requires more computation and a larger number of scalars. However, we have found that the finest scale does not contribute much to performance and it has a very small weight. Hence, it is possible to reduce the complexity of MS-SpEED-IQA by removing the finest scale without a noticeable loss in predictive performance.

We also studied the performance of SpEED-IQA across scales, ranging from scale 1 (no down-sampling) down to scale 5 (a down-sampling of the image by 2 repeated four times). Table III shows that SpEED-IQA performs best when $k = 3$, whereas RRED performs best using scale $k = 2$ [6].

B. VQA Experiments

We again selected several leading VQA models for comparison: PSNR, SSIM [1], MS-SSIM [31], SSIM+ [41], ST-RRED [14], and SpEED-VQA using 5×5 blocks), and applied them on four video quality databases: LIVE-VQA [42], LIVE-Mobile [43], Waterloo [44], and the recently designed LIVE-NFLX Video QoE DB [45]. The latter two databases include both quality changes and rebuffering events; hence, are highly relevant to video streaming applications. On these two databases, we applied the VQA models only on the uninterrupted video frames. On all databases and for all VQA models, only the luminance frames were used. SSIM+ is not publicly available; hence, we report its performance only on the Waterloo DB. Table IV shows that, except on LIVE-VQA, SpEED-VQA performs at par with ST-RRED, and significantly better than the other models. Recently, we conducted a broad survey of the predictive performance of ST-RRED [14], [18] on a wide variety of image/video quality databases and observed consistent out-performance by ST-RRED. This level of performance is closely matched by SpEED-VQA.

TABLE IV
VIDEO RESULTS ON LEADING VQA DATABASES

IQA/VQA	LIVE-VQA	LIVE-Mobile		Waterloo	LIVE-NFLX
		M	T		
PSNR	0.5233	0.6870	0.5895	0.6715	0.5257
SSIM	0.6947	0.7543	0.6608	0.8177	0.7230
MS-SSIM	0.7364	0.7507	0.6354	0.7928	0.6979
SSIM+	-	-	-	0.8024	-
ST-RRED	0.8052	0.8925	0.9122	0.8196	0.7257
SpEED-VQA	0.7744	0.8971	0.9123	0.8237	0.7273

We separately report the performance on the mobile (M) and tablet (T) subsets of the LIVE-Mobile database.

TABLE V
WEIGHTED AVERAGE (BY THE NUMBER OF VIDEOS) PERFORMANCE ON LIVE VQA, WATERLOO AND LIVE MOBILE, NUMBER OF SCALARS AND EXECUTION TIMES REQUIRED BY LEADING VQA MODELS

IQA model	# Scalars	SROCC	Time
PSNR	*L*	0.6254	0.0073
SSIM	*L*	0.7455	0.0060
MS-SSIM	5*L*	0.7440	0.1320
ST-RRED	*L*/576	0.8493	2.8069
ST-RREDopt	*L*/576	0.8493	0.2895
ST-RREDopt (SN)	1	0.7576	0.2895
SpEED-VQA	*L*/6400	0.8438	0.0451
SpEED-VQA (SN)	1	0.7847	0.0451

L refers to the size of each video frame.

For a fair computational comparison between ST-RRED and SpEED-VQA, we derived a computationally efficient version of ST-RRED, which we call ST-RREDopt. In ST-RREDopt, only the best performing subband of ST-RRED (band 4, which corresponds to $k = 4$ and the vertical orientation) is computed and used for quality assessment. Therefore, computing the full steerable filterbank is avoided.

Table V also illustrates the advantages of the proposed model: it requires less reference video data, is very fast, and delivers quality prediction performance very close to that of ST-RRED. Notably, the SN performance of the proposed model is better than that of ST-RRED. As shown in Tables II and V, SpEED-QA performs as well as or better than (ST-)RREDopt but is much faster. The main reason for the complexity reduction in SpEED-QA is that it can operate at a lower scale than (ST-)RRED, and it requires less reference video data. Furthermore, the simple spatial filters employed by SpEED-QA may be easily implemented and optimized.

VI. FUTURE WORK

SpEED-QA is a very efficient spatial-domain approach to RR IQA and VQA. The main advantages are simplicity of concept and implementation, very efficient computational complexity, and very high quality prediction performance on both images and videos. We envision deploying similar "conditional" approaches to I-VQA in the NR application context.

REFERENCES

[1] 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, Apr. 2004.

[2] H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessment," in *Proc. 1st Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, 2005, pp. 23–25.

[3] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Proc. SPIE*, vol. 5666, pp. 149–159, 2005.

[4] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.

[5] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[6] R. Soundararajan and A. C. Bovik, "R-RED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.

[7] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4602–4613, Nov. 2015.

[8] 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.

[9] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, Jan. 2016.

[10] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[11] C. G. Bampis, T. R. Goodall, and A. C. Bovik, "Sampled efficient full-reference image quality assessment models," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, 2016, pp. 561–565.

[12] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[13] 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.

[14] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 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, Mar. 2014.

[16] X. Li, Q. Guo, and X. Lu, "Spatiotemporal statistics for video quality assessment," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3329–3342, Jul. 2016.

[17] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.

[18] "On the robust performance of the ST-RRED video quality predictor." 2017. [Online]. Available: http://live.ece.utexas.edu/research/Quality/ST-RRED/

[19] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Neural Inf. Process. Syst.*, pp. 855–861, 1999.

[20] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.

[21] R. Reininger and J. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Trans. Commun.*, vol. 31, no. 6, pp. 835–839, Jun. 1983.

[22] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Math. Imag. Vis.*, vol. 18, no. 1, pp. 17–33, 2003.

[23] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[24] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, vol. 9, no. 2, pp. 181–197, 1992.

[25] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vis. Res.*, vol. 38, no. 5, pp. 743–761, 1998.

[26] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[27] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[28] J. Malo, I. Epifanio, R. Navarro, and E. P. Simoncelli, "Nonlinear image representation for efficient perceptual coding," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 68–80, Jan. 2006.

[29] O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neurosci.*, vol. 4, no. 8, pp. 819–825, 2001.

[30] S. Lyu and E. P. Simoncelli, "Statistically and perceptually motivated nonlinear image representation," *Proc. SPIE*, vol. 6492, 2007, Art. no. 649 207.

[31] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, pp. 1398–1402, vol. 2, 2003.

[32] 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.

[33] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process., Honolulu, HI, USA*, Jun. 2007, pp. I-869–I-872.

[34] D. C. Burr and J. Ross, "Contrast sensitivity at high velocities," *Vis. Res.*, vol. 22, no. 4, pp. 479–484, 1982.

[35] 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, Aug. 2011.

[36] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[37] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, 2010, Art. no. 011 006.

[38] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Eur. Workshop Vis. Inf. Process.*, 2013, pp. 106–111.

[39] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.

[40] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[41] A. Rehman, K. Zeng, and Z. Wang, "Display device-adapted video quality-of-experience assessment," *Proc. SPIE*, vol. 9394, 2015, Art. no. 939406.

[42] 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, Jun. 2010.

[43] 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 Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652–671, Oct. 2012.

[44] Z. Duanmu, Z. Kai, K. Ma, A. Rehman, and Z. Wang, "A quality-of-experience index for streaming video," *IEEE Sel. Topics Signal Process.*, 2016, vol. 11, no. 1, pp. 154–166, Feb. 2017.

[45] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Temporal effects on subjective video quality of experience," *IEEE Trans. Image Process.*, to be published.