Evaluating Foveated Video Quality Using Entropic Differencing

Yize Jin
The University of Texas at Austin
Austin, USA
yizejin@utexas.edu

Anjul Patney
Facebook Reality Labs
Seattle, USA
rwebb@fb.com

Alan Bovik
The University of Texas at Austin
Austin, USA
bovik@ece.utexas.edu

Abstract—**Virtual Reality is regaining attention due to recent advancements in hardware technology. Immersive images / videos are becoming widely adopted to carry omnidirectional visual information. However, due to the requirements for higher spatial and temporal resolution of real video data, immersive videos require significantly larger bandwidth consumption. To reduce stresses on bandwidth, foveated video compression is regaining popularity, whereby the space-variant spatial resolution of the retina is exploited. Towards advancing the progress of foveated video compression, we propose a full reference (FR) foveated image quality assessment algorithm, which we call foveated entropic differencing (FED), which employs the natural scene statistics of bandpass responses by applying differences of local entropies weighted by a foveation-based error sensitivity function. We evaluate the proposed algorithm by measuring the correlations of the predictions that FED makes against human judgements on the newly created 2D and 3D LIVE-FBT-FCVR databases for Virtual Reality (VR). The performance of the proposed algorithm yields state-of-the-art as compared with other existing full reference algorithms. Software for FED has been made available at: http://live.ece.utexas.edu/research/Quality/FED.zip**

Index Terms—**Entropic Differencing, foveated video quality assessment, contrast sensitivity function**

I. INTRODUCTION

In virtual reality (VR), immersive images and videos are important forms of media that carry visual information covering all 360 degrees. Given the highest resolution of the human vision system (HVS) of about 120 pixels per degree (PPD), to reach a desirable viewing experience, the spatial resolution of immersive videos should reach at least 8K. In addition, immersive videos should include high frame rates (HFR) to reduce motion sickness, hence demanding even higher bandwidth consumption.

To reduced the stresses imposed by such bandwidth-hungry immersive VR systems, foveated image / video compression is regaining relevance, due to the availability of consumer eye tracking devices that can be easily integrated into HMDs. Foveated image / video compression exploits the reduced visual acuity in the visual periphery, and assign higher compression ratio / quantization parameters (QP) to higher eccentricity relative to the fovea. Early implementations of the technique appeared more than two decades ago [1]–[4], however, neither

the driving need nor affordable eye tracking devices were mature.

Recently developed foveated compression algorithms, [5], [6], are usually supported by user studies which serve to verify the efficacy of the algorithms. However, although user / subjective evaluations are the most reliable method of quality assessment, they are also known to be expensive and time consuming, making it necessary to devise objective foveated quality assessment algorithms. Towards the development of objective algorithms, the authors of [7], [8] developed 2D and 3D foveated / compressed video quality databases in Virtual Reality, and evaluated a wide variety of quality assessment (QA) algorithms on the databases. Among available full reference (FR) foveated algorithms, Foveated Wavelet Quality Index (FWQI) [9] weights the error between the reference image and the foveated / distorted image in the wavelet domain, using a contrast sensitivity function (CSF) model [4], and a visually detectable noise threshold model [10]. The Foveation-based Content Adaptive SSIM (FA-SSIM) model [11] weights the SSIM index with a foveation-based CSF which includes the retinal velocity and the corner velocity.

An important category of IQA algorithms relies on the statistical regularities of natural images (natural scene statistics – NSS). NSS-based QA was first used in the Information Fidelity Criterion (IFC) [12] and the Visual Information Fidelity (VIF) models [13], where Gaussian scale mixtures (GSM) [14] were used to characterize natural images in the wavelet domain, and image quality is captured by the conditional mutual information between the reference image model and the distorted image model. The creators of the Reduced Reference Entropic Differencing (RRED) [15] and Spatio-Temporal Entropic Differencing (STRRED) [16] models showed that, rather than computing conditional mutual information, which requires full reference information, a reduced amount of reference information can be obtained by computing the average differences of scaled local entropies in the wavelet domain. The authors of Speed-QA [17] showed that local entropic differencing can be applied in the spatial domain, achieving comparable performance to RRED or STRRED with much higher efficiency.

We propose an NSS-based foveated image quality assessment algorithm which model the bandpass-filtered images as GSM [14], and, for each subband, instead of scaling local

entropies with local scalar factors, we weight the entropy map using a foveation-based error sensitivity function [9], with frequency set to the average frequency of the subband. Then the entropy differences are calculated and averaged across subbands. We have found that the proposed Foveated Entropic Differencing (FED) algorithm achieves state-of-the-art performance on both the 2D and 3D LIVE-FBT-FCVR databases [8] when applied on a frame-by-frame basis.

II. Foveation Based Error Sensitivity

We employ the foveation-based error sensitivity function used in FWQI [9] for given frequency f and eccentricity e:

$$S_f(f,e) = \begin{cases} \frac{CS(f,e)}{CS(f,0)} = \exp(-0.0461f \cdot e), & f \leq f_m(e), \\ 0, & f > f_m(e), \end{cases}$$

(1)

where $CS(f,e)$ is the contrast sensitivity function in [4] defined by the inverse of the contrast threshold CT.

$$CS(f,e) = \frac{1}{CT(f,e)}$$

(2)

$$CT(f,e) = CT_0 \exp\left(\alpha f \frac{e + e_2}{e_2}\right),$$

(3)

where CT is the contrast threshold, CT_0 is a constant minimal contrast threshold, f and e are spatial frequency and the eccentricity, respectively, α is the spatial frequency decay constant, and e_2 is the constant half-resolution eccentricity. We follow the parameter values given in [4]: $CT_0 = 1/64, \alpha = 0.106$, and $e_2 = 2.3$.

In (1) f_m is the combined cutoff frequency for a given eccentricity:

$$f_m(e) = \min(f_c(e), f_d)$$

(4)

where the critical frequency f_c for a given eccentricity e is computed by setting $CT = 1.0$ in (3), which is the maximum possible contrast:

$$f_c(e) = \frac{e_2 \ln(1/CT_0)}{(e + e_2)\alpha}$$

(5)

and f_d is the display Nyquist frequency: $f_d = d/2$ (cicles/degree), where d is the display resolution:

$$d \approx \frac{\pi M v}{180} \text{ (pixels/degree)}$$

(6)

where M is the with of the picture in pixel, v is the viewing distance measured in image width, and can be related to the field of view (FOV) when gazing at the center of the image by:

$$v = \frac{\cot(FOV/2)}{2}$$

(7)

III. The Proposed Algorithm

NSS-based FR / RR IQA algorithms rely on the strong statistical regularities of natural images: bandpass coefficients of natural images follow heavy-tailed distributions [14], and divisively normalizing the bandpass coefficients using local variances causes the empirical distribution strongly tend towards a decorrelated normal Gaussian distribution [18]. The
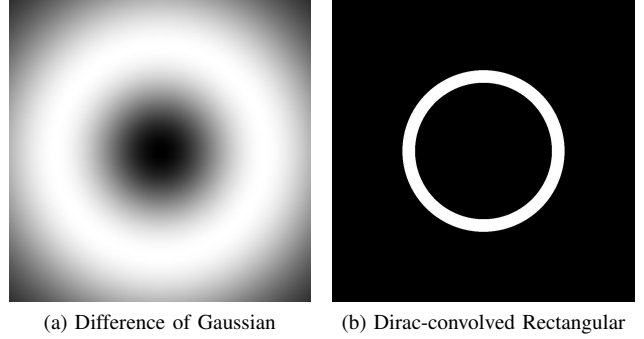

(a) Difference of Gaussian   (b) Dirac-convolved Rectangular

Fig. 1. A DoG and a 2D Dirac-convolved rectangular in the frequency domain. We set $\sigma_1 = r_1$ and $\sigma_2 = r_2$, where σ_1 and σ_2 are the two variances of Gaussians in the DoG, and r_1 and r_2 are the inner and outer radii of the Dirac-convolved rectangular. It can be seen that the DoG has much larger frequency coverage.

authors of [15], [17], [19] have explored DCT decomposition, wavelet decompositions, and spatial predictive coding filters.

The proposed FED model is based on the idea of combining measurements of statistical regularities (or lack thereof) with foveation-based error sensitivity functions. A wide range of zero-DC bandpass filters can be used to reveal these statistical regularities. We assume that when restricted to a narrow band in the frequency domain, as shown in Fig. 1b, the NSS of the bandpass coefficients often reflect distortions expressed in the particular frequency band. Therefore, the entropy differences computed from the NSS may be used to capture local perceptual distortions within subbands. A foveation-based error sensitivity function [9] tuned to a specific narrow subband can be applied to weight spatial entropy differences from each such subband to approximately weight the importance of perceptual distortions in the fovea and in the periphery.

A. Modeling Bandpass Coefficients

We define $\{B_k\}_N, k \in \{1, ..., N\}$ to be a filterbank of zero-DC bandpass filters, and $\{\hat{I}_k\}$ corresponding bandpass reponses:

$$\hat{I}_k(i,j) = B_k(i,j) * I(i,j),$$

(8)

where $*$ and $I(i,j)$ are the convolution operator and the input image, respectively. We model a $b \times b$ bandpass response in $\hat{I}_k$ as a GSM vector: $\mathbf{x} \sim z\mathbf{U}$, where $\sim$ denotes equal in distribution, z indicates the mixing multiplier which is independent of $\mathbf{U}$, and $\mathbf{U}$ is a zero-mean Gaussian vector with covariance matrix $\mathbf{C}_u$. The distribution of the GSM vector can be expressed as:

$$f_{\mathbf{x}}(\mathbf{x}) \sim \int \frac{1}{(2\pi)^{d/2}|z^2\mathbf{C}_u|^{1/2}} \exp\left(\frac{-\mathbf{x}^T \mathbf{C_u}^{-1}\mathbf{x}}{z^2}\right) f_z(z)dz$$

(9)

We found the choice of filterbanks also important. One possible choice is Difference of Gaussian (DoG) filters:

$$DoG(r; \sigma_1, \sigma_2) =$$
$$\frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{r^2}{2\sigma_1^2}\right) - \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{r^2}{2\sigma_2^2}\right),$$

(10)

where σ_1 and σ_2 are standard deviations of the Gaussian functions, and r is the distance between the pixel (i, j) and the gaze center (i_0, j_0):

$$r = \sqrt{(i - i_0)^2 + (j - j_0)^2} \qquad (11)$$

However, DoG filters are not desirable because of their large frequency coverage when trying to extract high-frequency information, as shown in Fig. 1a. This is because when computing the CSF of the subband coefficients, it is desired that each bandpass filter only cover a narrow frequency band, so that the frequency band used in the CSF is precise. So we designed a bank of filters in the frequency domain by convolving an isotropic (circular) Dirac delta function with an ideal rectangular filter, as shown in Fig. 1b:

$$\tilde{B}_k(r) = \delta(r - r_k) * rect(r), \qquad (12)$$

where $\tilde{B}_k$ is B_k represented in the frequency domain, r_k is the center frequency of the kth subband, $rect(r)$ is a rectangular function with width parameter r_b:

$$rect(r) = \begin{cases} 1, & |r| < r_b, \\ 0, & \text{otherwise} \end{cases} \qquad (13)$$

The convolution in (12) can be easily computed as:

$$\tilde{B}_k(r) = \begin{cases} 1, & |r - r_k| < r_b, \\ 0, & \text{otherwise} \end{cases} \qquad (14)$$

B. Foveated Entropic Differencing

Let C_{pqkr} and C_{pqkd} be the $b \times b$ bandpass coefficients of the reference and distorted luminance image block (p, q) using filter B_k respectively. The bandpass coefficients are described by the GSM model:

$$C_{pqkr} = Z_{pqkr} U_{pqkr}, C_{pqkd} = Z_{pqkd} U_{pqdk}, \qquad (15)$$

where Z_{pqkr} and Z_{pqkd} are random scalar variables and $U_{pqkr} \sim \mathcal{N}(0, \mathbf{K}_{U_{kr}}$ and $U_{mkd} \sim \mathcal{N}(0, \mathbf{K}_{U_{kd}})$ are Gaussian random vectors. By passing the bandpass coefficients through a hypothetical neural noise model, we have:

$$C'_{pqkr} = C_{pqkr} + W_{pqkr}, \ C'_{pqkd} = C_{pqkd} + W_{pqkd}, \qquad (16)$$

where $W_{pqkr} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}_N)$ and $W_{pqkd} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}_N)$. Since bandpass coefficients tends towards a multivariate Gaussian distribution when conditioned on realizations of Z_{pqkr} and Z_{pqkd}, conditional local entropies are computed as:

$$h(C'_{pqkr}|z_{pqkr}) = const. + \frac{1}{2} \log[|z_{pqkr}^2 \mathbf{K}_{U_{kr}} + \sigma_w^2 \mathbf{I}_N|] \quad (17)$$

$$h(C'_{pqkd}|z_{pqkd}) = const. + \frac{1}{2} \log[|z_{pqkd}^2 \mathbf{K}_{U_{kd}} + \sigma_w^2 \mathbf{I}_N|] \quad (18)$$

In [15], [17] and [16], the local entropies are scaled by scalar factors $\gamma_{pqkr} = \log(1 + s_{pqkr}^2)$ and $\gamma_{pqkd} = \log(1 + s_{pqkd}^2)$, here instead we weight the local entropies using the foveation based error sensitivity function. Specifically, let f_k be the average

frequency of the filter B_k, e be the eccentricity in degrees of the block (p, q), which is computed as follows:

$$e(p, q) = \tan^{-1}\left(\frac{\sqrt{(bp - i_0)^2 + (bq - j_0)^2}}{vM} \right) \qquad (19)$$

where (i_0, j_0) is the gazing center. We first normalize the foveation-base error sensitivity function (Equation (1)) as:

$$S_k^{\mathcal{N}}(p, q) = \frac{S_f(f_k, e(p, q))}{\sum_{p,q} S_f(f_k, e(p, q))} \qquad (20)$$

where we have assumed equal contribution of each subband. Then the foveated entropic differencing map for subband k is defined as:

$$D_k(p, q) = S_k^{\mathcal{N}}(p, q) \cdot (h(C'_{pqkr}|z_{pqkr}) - h(C'_{pqkd}|z_{pqkd})), \qquad (21)$$

and the final quality prediction score is computed by taking the summation of the absolute entropy differences:

$$FED = \sum_{k,p,q} |D_k(p, q)| \qquad (22)$$

IV. EXPERIMENTS

A. Evaluation Framework

The algorithms were tested and compared on the LIVE-FBT-FCVR databases [8], where subjective opinion scores (MOS) and difference MOS (DMOS) were collected for 180 foveated / compressed images and 10 reference (190 total) 8K immersive videos.

The foveation distortions in the subjective study of [8] were created in real time. During the playback, the system reads three immersive (equirectangular) videos, each video uniformly pre-compressed with one of 5 QP levels, and creates the foveation distortion by combining three videos with descending quality order from the fovea to the periphery, given gazing data obtained from the eye tracker.

To recover the foveation experience, we adopt a viewport-based assessment framework as also mentioned in [7], where for each immersive video we sample 18 viewports videos, and set the resolution of each viewport to 1024x1024, and the FOV to be 90°. The viewports distribute uniformly on a sphere in terms of longitude and latitude.

B. Parameter Settings

The display resolution d and the display Nyquist frequency f_d can then be computed as:

$$d \approx \frac{\pi M v}{180} = 8.94 \text{ (pixels/degree)}, \qquad (23)$$

$$f_d = 4.47 \text{ (cycles/degree)} \qquad (24)$$

We then divided the frequency band from d to 0 uniformly into n subbands, and set $n = 12$. We set the block size b to 4 and the noise level σ_w to 0.1.

TABLE I
PERFORMANCE OF ALGORITHMS ON 2D AND 3D LIVE-FBT-FCVR
DATABASES. THE BEST PERFORMING ALGORITHM IS BOLDFACED.

IQA Models	Performances			
	The 2D database		The 3D database	
	PLCC↑	SROCC↑	PLCC↑	SROCC↑
PSNR [20]	0.6941	0.6954	0.4379	0.4418
SSIM [20]	0.7260	0.7191	0.4184	0.4429
MS-SSIM [21]	0.7288	0.7243	0.5337	0.5531
VIF [13]	0.8102	0.8068	0.6536	0.6765
S-RRED [16]	0.7896	0.7885	0.5604	0.5744
Speed-IQA [17]	0.7760	0.7866	0.5084	0.4959
FSIM [22]	0.7712	0.7808	0.5904	0.6273
FWQI [9]	0.7906	0.7848	0.8041	0.7841
FASSIM [11]	0.7573	0.7418	0.7549	0.7401
FovFilt*	0.8484	0.8358	0.6000	0.6224
FED†	**0.8971**	**0.8954**	**0.8276**	**0.8364**

*Foveated filtering is simply applied before SpEED-IQA.

†The number of subband is set to 12.

TABLE II
PERFORMANCE WHEN USING DIFFERENT NUMBERS OF SUBBANDS ON THE
2D LIVE-FBT-FCVR DATABASE. THE BEST PERFORMING ALGORITHM IS
BOLDFACED.

#Subbands	Performances			
	PLCC↑	SROCC↑	KROCC↑*	RMSE↓
6	0.8701	0.8736	0.6900	4.73
8	0.8860	0.8823	0.7030	4.57
10	0.8930	0.8884	0.7109	4.43
12	**0.8971**	**0.8954**	**0.7196**	**4.35**

*Kendall's rank order correlation coefficient.

TABLE III
PERFORMANCE WHEN USING DIFFERENT FILTERBANKS ON THE 2D
LIVE-FBT-FCVR DATABASE. THE BEST PERFORMING ALGORITHM IS
BOLDFACED.

Filterbanks	Performances			
	PLCC↑	SROCC↑	KROCC↑*	RMSE↓
DoG	0.7768	0.7585	0.5720	6.18
Triangular†	0.8399	0.8302	0.6432	5.35
Rectangular†	**0.8971**	**0.8954**	**0.7196**	**4.35**

*Kendall's rank order correlation coefficient.

†The function is convolved by n Dirac delta function.

C. Overall Performance

We compared FED with several leading existing IQA algorithms on both the 2D and 3D LIVE-FBT-FCVR databases: PSNR [20], SSIM [20], MS-SSIM [21], VIF [13], S-RRED [16], SpeedIQA [17], FSIM [22], FWQI [9], and FA-SSIM [11]. For each immersive video, each algorithm was evaluated on every viewport video on a frame-by-frame basis, and the scores obtained on each viewport frame were then averaged across all frames and 18 viewport videos.

Pearson's linear correlation coefficient (PLCC), Spearman's rank order correlation coefficient (SROCC), Kendall's rank order correlation coefficient (KROCC), and root mean square (RMSE) between the DMOS and algorithm predictions are used as the criteria for evaluating the performance of the algorithms. Before computing PLCC and RMSE, a four-parameter logistic non-linearity was employed [23]:

$$Q(x) = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \exp(-\frac{x - \beta_3}{|\beta_4|})} \quad (25)$$

For foveated QA, one intuitive solution would be first apply a foveated filter to both reference and distorted images, and use traditional QA methods to evaluated the filtered image pairs. So we also included as a baseline algorithm the method of using SpEED-IQA on foveated reference-distorted image pairs. We used the same foveated filtering method as in [24].

The results are shown in TABLE I. The proposed FED model outperformed existing FR models, foveated and non-foveated, by a large margin on both 2D and 3D databases. In addition, the FED model also outperformed the intuitive foveated filtering model (FovFilt), showing that it can effectively capture foveation distortions.

D. Number of Subbands

We also studied the effects of using different numbers $(n = 6, 8, 10, 12)$ of subbands, as shown in TABLE II. It can be seen that the performance is improved when using more subbands, since frequency information is more precisely represented, and foveation distortions are better captured. We

also found that increasing the subband number over 12 leads to limited performance gain but higher complexity.

E. Using Different Filterbanks

In addition to Dirac-convolved "ideal" low-pass filterbanks, we also tested DoG filterbanks, and a filterbank composed of isotropic Dirac-convolved triangular functions in the frequency domain. A rotationaly symmetric Dirac-convolved triangular frequency band introduces less Gibbs phenomena than using a rectangular frequency band, but also removes information in each subband. A Dirac-convolved triangular function is defined by:

$$\tilde{B}_k(r) = \begin{cases} \frac{r_b - |r - r_k|}{r_b}, & |r - r_k| < r_b, \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

where r is defined as in (11), and r_k is the center frequency. Their performances are compared in TABLE III. It can be seen that, although there are severe spatial Gibbs phenomena when using Dirac-convolved rectangular functions, the statistics of the corresponding bandpass coefficients are less affected when capturing perceptual distortions. However, the results suggest that, for foveated quality assessment, large frequency bands (DoGs) and loss of information in each frequency band (isotropic Dirac-convolved triangular functions) may cause degraded performance.

V. CONCLUSION

We propose a foveated video quality model called FED that combines NSS-based QA features with a foveation-based error sensitivity function, based on the assumption that bandpass filtering restricts spatial entropy differences to capture perceptual distortions within a given frequency subband. We found that FED yields SOTA performance on the 2D and 3D LIVE-FBT-FCVR databases.

REFERENCES

[1] P. L. Silsbee, A. C. Bovik, and Dapang Chen, "Visual pattern image sequence coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 4, pp. 291–301, 1993.

[2] T. H. Reeves and J. A. Robinson, "Adaptive foveation of mpeg video," in *Proceedings of the Fourth ACM International Conference on Multimedia*, New York, NY, USA, 1997, MULTIMEDIA '96, p. 231–241, Association for Computing Machinery.

[3] Philip Kortum and Wilson S. Geisler, "Implementation of a foveated image coding system for image bandwidth reduction," in *Human Vision and Electronic Imaging*, Bernice E. Rogowitz and Jan P. Allebach, Eds. International Society for Optics and Photonics, 1996, vol. 2657, pp. 350 – 360, SPIE.

[4] Wilson S. Geisler and Jeffrey S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," in *Human Vision and Electronic Imaging III*, Bernice E. Rogowitz and Thrasyvoulos N. Pappas, Eds. International Society for Optics and Photonics, 1998, vol. 3299, pp. 294 – 305, SPIE.

[5] Gazi Karam Illahi, Thomas Van Gemert, Matti Siekkinen, Enrico Masala, Antti Oulasvirta, and Antti Ylä-Jääski, "Cloud gaming with foveated video encoding," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 1, Feb. 2020.

[6] Jihoon Ryoo, Kiwon Yun, Dimitris Samaras, Samir R. Das, and Gregory Zelinsky, "Design and evaluation of a foveated video streaming service for commodity client devices," New York, NY, USA, 2016, MMSys '16, Association for Computing Machinery.

[7] Yize Jin, Meixu Chen, Todd Goodall Bell, Zhaolin Wan, and Alan Bovik, "Study of 2D foveated video quality in virtual reality," in *Applications of Digital Image Processing XLIII*. International Society for Optics and Photonics, 2020, vol. 11510, pp. 18 – 26, SPIE.

[8] Yize Jin, Meixu Chen, Todd Goodall, Anjul Patney, and Alan C. Bovik, "Subjective and objective quality assessment of 2d and 3d foveated video compression in virtual reality," *IEEE Transactions on Image Processing*, submitted.

[9] Zhou Wang, Alan Conrad Bovik, Ligang Lu, and Jack L. Kouloheris, "Foveated wavelet image quality index," in *Applications of Digital Image Processing XXIV*, Andrew G. Tescher, Ed. International Society for Optics and Photonics, 2001, vol. 4472, pp. 42 – 52, SPIE.

[10] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164–1175, 1997.

[11] S. Rimac-Drlje, G. Martinović, and B. Zovko-Cihlar, "Foveation-based content adaptive structural similarity index," in *2011 18th International Conference on Systems, Signals and Image Processing*, 2011, pp. 1–4.

[12] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[13] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[14] Martin J. Wainwright and Eero P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," Cambridge, MA, USA, 1999, NIPS'99, p. 855–861, MIT Press.

[15] R. Soundararajan and A. C. Bovik, "Rred indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.

[16] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 4, pp. 684–694, 2013.

[17] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "Speed-qa: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1333–1337, 2017.

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

[19] R. Reininger and J. Gibson, "Distributions of the two-dimensional dct coefficients for images," *IEEE Transactions on Communications*, vol. 31, no. 6, pp. 835–839, 1983.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[21] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, 2003, vol. 2, pp. 1398–1402 Vol.2.

[22] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[23] VQEG, "Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment," .

[24] C. Bradley, J. Abrams, and W. S. Geisler, "Retina-v1 model of detectability across the visual field," *Journal of Vision*, vol. 14, no. 12, pp. 1534–7362, 2014.