A Full-Reference Video Quality Assessment Method for 4K UHD Video Based on Multi-Feature Fusion

Geng Yi, Shi Ping and Pan Da

February 24, 2023

A Full-reference Video Quality Assessment Method for 4K UHD Video based on Multi-Feature Fusion

1st Geng Yi
School of Information and Communication Engineering
Communication University of China
Beijing, China
gy@cuc.edu.cn

2nd Shi Ping *
School of Information and Communication Engineering
Communication University of China
Beijing, China
shiping@cuc.edu.cn

3rd Pan Da
School of Information and Communication Engineering
Communication University of China
Beijing, China
pdmeng@cuc.edu.cn

***Abstract*—Video quality assessment plays an important role in the quality control of video transmission and the development of video processing equipment and algorithms. With the popularity of UHD TV, the demand for UHD video quality assessment is becoming more and more urgent. In this paper, we propose a method for 4K UHD video quality assessment based on multi-feature fusion (MFF-VQA). First, we select eight frame-level features which could better reflect the perceived video quality through a series of ablation experiments. Then, we present a scheme which can fuse the eight features into a quality score. Experimental results show that, compared with other similar methods, the proposed method can achieve better performance even with lower algorithm complexity and fewer video frames.**

Key Words—Full-reference, video quality assessment, ultra-high-definition, multi-feature fusion

I. INTRODUCTION

Ultra-High Definition (UHD) is a new video format with 4K (3840×2160 pixels) or 8K (7680×4320 pixels) resolution, high dynamic range and wide color gamut etc. In recent years, with the development of UHD technology, UHD video has become more and more popular. It is widely applied in television, education, entertainment, intelligent transportation, industrial manufacturing and so on. Like Standard-Definition (SD) and High-Definition (HD), UHD videos also suffer from distortion and quality degradation in signal acquisition, processing, transmission and storage, due to imperfect imaging systems, signal processing methods, transmission media, and storage devices. In order to better monitor and control the video quality in each transmission link, to develop an efficient and effective video quality assessment algorithm is essential.

Video Quality Assessment (VQA) aims to measure the variation and distortion of video information, and is the key to the development and optimization of video processing algorithms. It is also a classical and challenging problem in the field of computer vision. VQA can be divided into subjective quality assessment and objective quality assessment. Subjective quality assessment is the direct evaluation of video quality by subjects. Although the scoring result obtained by this method are more in line with the subjective feelings of the audience, it also has the disadvantages of heavy workload and long time. The objective quality assessment is to develop video quality metrics, which are calculated by a computer according to a certain algorithm. Objective assessment methods can be divided into three types: Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR), according to the degree of dependence on the reference video. Full-Reference methods often achieve better performance because they can use all the information of the reference video. In many practical applications, FR-VQA can replace the subjective assessment method. Therefore, the FR-VQA has important research and application value.

In recent years, the quality assessment of UHD video has received great attention from many researchers. And the robustness of some metrics developed for SD or HD video have also been validated in UHD video, such as FSIM[1], VIF[2] and SSIM[3]. However, the performance of these metrics varies across different UHD datasets, and none significantly outperforms the others. Some metrics perform better on certain datasets or distortion types, but may perform poorly when generalized to other datasets or distortion types. There are also a few studies using deep neural networks to build quality evaluation models[4]. Although some deep learning methods perform well, they are difficult to deploy effectively in many practical problems. This is because deep learning algorithm models often have a large number of parameters and require enormous computing and storage resources. This paper will focus on FR-VQA and propose a multi-feature fusion method for UHD video. Our main work can by described as follows: the frame-level features which could better reflect the perceived video quality have been chosen and a scheme which can fuse the features into a quality score is presented. Unlike the work in [5],[6] we are not just focus on only one or more metrics, such as sharpness or contrast, but on the overall perceptual quality of the video. To verify the performance of the proposed MFF-VQA model, a comprehensive experimental evaluation is conducted. MFF-VQA outperforms traditional quality metrics, especially in terms of lightweight. It achieves a better performance with fewer video frames.

The rest of this paper is organized as follows. In Section II, the existing open UHD datasets are investigated in detail. Section III describes the selection of eight features. Section IV presents the MFF-VQA model, and the performance comparisons are conducted in Section V. Conclusions and the potential future extensions are discussed in Section VI.

II. EXISTING OPEN UHD DATASETS

Video quality assessment datasets are generally constructed using subjective scores, such as Mean Opinion

*Corresponding author： Shi Ping
shiping@cuc.edu.cn

TABLE I. THE SUMMARY OF EXISTING OPEN VIDEO DATASETS

Dataset	Year	Camera	Raw Video	Distorted video	Resolution	Frame Rate (fps)	Score Type
[14]	2013	Sony F65	15	—	3840x2160	30	—
[15]	2017	RED EPIC-M	10	240	5120×2700	30	DMOS
[16]	2019	—	17	756	3840x2160	15-60	MOS
[17]	2020	Sony F65	16	—	3840×2160	50/120	—
[18]	2021	SONY Z280 etc.	—	206	3840×2160	50	MOS

Scores(MOS) or Difference of Mean Opinion Scores(DMOS). Traditional video datasets are usually shot with professional equipment to create a set of reference videos, and then manually process these reference videos to get distorted videos. In addition to this method, 4K UHD video can also be made by interpolating low-resolution videos[7] or generate by cropping 5k videos[8]. In pursuit of higher picture quality, some large platforms have also launched various social, animation, and popular science 4K videos [9]-[12]. The work in [13] simulated the most common types of video distortion in real-life, and constructs a wild UHD video dataset, which provides a more comprehensive research profile for UHD video quality assessment.

SJTU Media Lab produced a dataset for 4K video sequences [14]. The 4K dataset consists of fifteen 2160p sequences, shot at 30 frames per second(fps) using a Sony F65 camera. They come in two formats: 10-bit YUV 4:4:4 and 8-bit YUV 4:2:0. The raw video data is recorded in Sony RAW 16-bit MAF format at 60 fps.

MCML video dataset consists of 10 original reference videos and 240 distorted videos [15]. The videos are all 10 seconds long and run at 30fps. The resolution of the original videos is 5K (5120×2700 pixels), and 4K resolution videos are generated by cropping. The video content includes flowers and trees, parks, dolls, animations, text, etc. Two spatial resolutions (4K UHD and FHD), three compression techniques (HEVC, AVC, and VP9), and wide bitrate ranges are included in the experiment. The dataset also contains the DMOS corresponding to the video, as well as the score range from 0 to 10.

AVT-VQDB-UHD-1 video dataset consists of 17 raw videos and 756 distorted videos [16]. Video lengths vary from 8 to 10 seconds. Video content includes animation, characters, dark scene, indoor scenes, etc. Videos are compressed using three different codecs, AVC, HEVC and VP9, at resolutions from 360p to 2160p, and framerates from 15fps to 60fps. The dataset also contains MOS [0,5] corresponding to distorted videos and associated confidence intervals (CIs).

Ultra-video Group (UVG) dataset consists of sixteen 4K (3840×2160) raw videos [17]. These natural sequences are captured at 50 fps or 120 fps and stored online in the original 8-bit and 10-bit 4:2:0 YUV format. Each video has its corresponding spatiotemporal variation, rate distortion characteristics, and HEVC and VVC compression distortion video.

DVL2021 dataset [18] contains 206 4K UHD video sequences, all shot in the wild with different types of camera. Each sequence was captured at 50 fps and stored in the original 10-bit 4:2:0 YUV format for 10 seconds.

The details of the aforementioned datasets are briefly summarized in Table I. The datasets in [14] and [17] do not provide subjective assessment scores, and the transmission distortion of UHD is not considered in [18]. While [15] and [16] contain rich themes, a large number of distorted videos, typical UHD distortion types. Therefore, they can provide a good basis for objective quality assessment.

III. FEATURES OF MFF-VQA

In order to screen out the features that can better reflect the perceptual video quality, we conducted a large number of ablation experiments. Finally, we choose three types of features, which are brightness gradient similarity, chroma similarity and mutual information. We will describe these features in detail next.

A. Gradient similarity

Brightness is one of the most basic features of the human visual system, and people's perception of brightness is mainly based on the sensitivity of brightness changes. In general, the human eye is less sensitive to noise clinging to areas of high brightness. This means that if an image has a higher background brightness, it can embed more additional information. Continuous image gradient changes can reflect the human eye's perception of brightness detail levels, and there have been many studies on the application of the image gradient function to quality evaluation [19]-[22]. Noland et al. attempted to get the perceived brightness of an HDR image from the pixel values, and test results showed that the most effective measure is to average the pixel brightness [23]. Therefore, we take gradient similarity as one of the features of proposed model (1), and finally get the gradient similarity feature of the video by calculating the average value.

$$G = \frac{1}{NL}\sum_{y=1}^{N}\sum_{x=1}^{L}\frac{2GM_S \cdot GM_D + T_1}{GM_S^2 + GM_D^2 + T_1} \quad (1)$$

Where GM_S, GM_D represent the magnitude of the gradient of the source video frame and the distorted video frame at *(x, y)*, respectively, and the gradient calculation uses the Scharrr operator. L and N are the number of horizontal and vertical pixels of the video frame. *T1* is a constant, and takes the value 160.

B. Chroma similarity

Chroma provides more visual detail, helping to quickly identify and confirm objects during human eye perception. In quality assessment, we usually use chroma as a complement to brightness. Therefore, we apply chroma similarity as a primary feature to the MFF-VQA model:

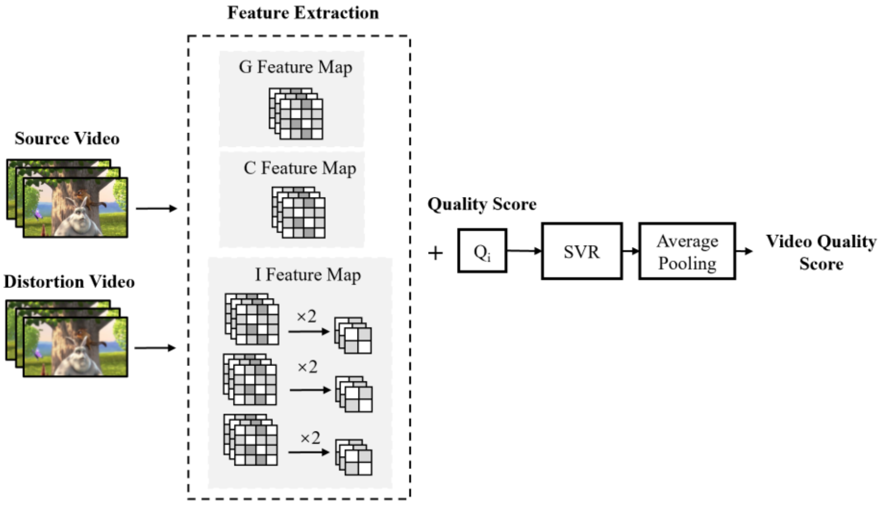

Fig. 1. Block diagram of MFF-VQA.

$$C = \frac{1}{NL}\sum_{y=1}^{N}\sum_{x=1}^{L} S_U \cdot S_V \quad (2)$$

S_U and S_V are U and V component similarity values for video frames.

C. Mutual information

The visual features of the human eye are like a multi-channel model, or rather it has a muti-frequency channel decomposition. In other words, the images in the retina are broken down into certain frequency bands, which are equally wide on a logarithmic scale. This also lays the foundation for us to use mutual information features to measure the human eye vision system, and this idea has also been applied in VIF, VMAF[25].

Images, as an independent information system, the larger the entropy of its information, the more complex the system and the more difficult it is to predict. The information entropy of the image can reflect the amount of information in the image. The larger the information entropy, the richer the detailed information. Traditional quality assessment is to measure how much information is lost when the external light image enters the human eye. Thus, most previous quality evaluations have focused on the amount of loss between the image information perceived by the human eye and the source image[26][27]. However, as the UHDTV video system almost achieves the ultimate quality, fidelity-based evaluation is insufficient [28].

In this paper, the model we proposed also includes the source video mutual information features and the distorted video mutual information features, which are intuitive sources of information for the human eye. Equation (3) calculates the mutual information between the source video frame information and the corresponding information perceived by the human eye. Equation (4) represents the distorted video mutual information, and equation (5) is the ratio of the first two. We input I_S, I_D , I and three of their down-sampled features, into the algorithm network as mutual information features of visual perception.

$$I_S = \sum_{x=1}^{N}\sum_{y=1}^{L} log_2\left(1+\frac{\widehat{Cov}(Y_S,Y_S)}{\sigma_n^2}\right) \quad (3)$$

$$I_D = \sum_{i=1}^{N}\sum_{k=1}^{L} log_2\left(1+\frac{M^2\widehat{Cov}(Y_S,Y_S)}{\sigma_d^2+\sigma_n^2}\right) \quad (4)$$

$$I = \frac{I_D}{I_S} \quad (5)$$

Where Y_S, Y_D are the grayscale information of the source video and the distorted video, respectively. σ_n^2 is the visual noise variance and takes the value 2. The parameters M and distortion variance σ_d^2 in equation (4) are calculated by (6) and (7):

$$M = \widehat{Cov}(Y_S,Y_D)\widehat{Cov}(Y_S,Y_S)^{-1} \quad (6)$$

$$\sigma_d^2 = \widehat{Cov}(Y_D,Y_D) - M\widehat{Cov}(Y_S,Y_D) \quad (7)$$

The mutual information calculation method we used is simpler and lighter than VIF. Its good performance has also been verified in the experimental part.

IV. MFF-VQA MODEL

A. Feature fusion

We calculate the above eight features on each video frame to obtain their corresponding feature maps, and then average the feature maps to get eight feature values to form a one-dimensional vector F as the training feature vector of the video frame. The one-dimensional vector features and video subjective evaluation score of the same video are input into support vector regression (SVR) for training and fractional regression. Finally, the video quality score is obtained by averaging the predicted scores of each frame output by SVR. Fig. 1 shows the overall process of the proposed method.

B. Support Vector Regression

Support Vector Regression (SVR) excels at high-dimensional regression problems and previous quality

TABLE II. PERFORMANCE COMPARISON OF 7 MODELS ON MCML DATASET

	MCML			AVT		
Metric	SROCC	PLCC	RMSE	SROCC	PLCC	RMSE
PSNR	0.8711	0.8541	1.3530	0.6700	0.6807	0. 8021
FSIM	0.9375	0.9408	0.8684	0.7410	0.8201	0.7273
VIF	0.9254	0.9495	0.7997	0.7866	0.7911	0.7090
SSIM	0.8951	0.8826	1.1331	0.7765	0.7573	0.8063
MS-SSIM	0.9275	0.9063	1.0802	0. 8091	0.8147	0.7163
IFC	0.8568	0.9133	1.0374	0.6107	0.6657	0. 8012
MFF-VQA	**0.9530**	**0.9764**	**0.5137**	**0.8702**	**0.8621**	**0. 5874**

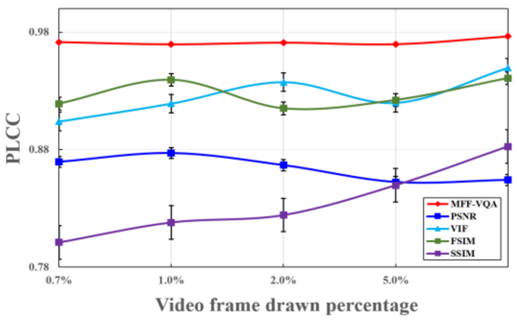

Fig. 2. Mean PLCC of quality assessment methods as a function of the percentage of video frame drawn on MCML dataset.

TABLE III. COMPARISON OF COMPUTATION TIME AMONG 7 MODELS ON MCML DATASET

Metric	**Time (sec/video frame)**
PSNR	1.5531
FSIM	1.4208
VIF	2.1406
SSIM	1.1461
MS-SSIM	1.6922
IFC	16.2752
MFF-VQA	**0.8563**

assessment algorithms, so we ended up choosing SVR as a machine learning tool. The quality score of each frame output by SVR are averaged to obtain the final video quality score. To implement the SVR, we used the libSVM package [29] with radial basis function kernel, whose parameters are estimated by cross-validation during training. The dataset is split into 80% and 20% for training and testing, respectively. The training and testing subsets do not overlap in testing. To ensure that the results do not depend on a specific train-test separation, we repeated the test 500 times by randomly splitting the dataset into a pair of training and testing sets.

Spearman Rank-order Correlation Coefficient (SROCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Square Error (RMSE),are used to measure the correlation between predicted scores and MOS scores after nonlinear regression. SROCC is mainly used to describe whether the scores predicted by the quality assessment model is consistent with the monotonicity of MOS. The higher the absolute score of SROCC, the closer the prediction of the model is to reality. PLCC is mainly used to indicate the degree of linear correlation between the predicted scores and MOS. Likewise, the larger the absolute value of PLCC, the better. RMSE is used to measure the deviation between the predicted scores and MOS, and smaller values are better.

A monotonic logistic function with five parameters to fit the predicted MFF-VQA score:

$$Q_j' = \beta_2 \left(\frac{1}{2} - \frac{1}{1+e^{\beta_2(x-\beta_3)}}\right) + \beta_4 Q_j + \beta_5 \tag{8}$$

Where nonlinear least squares optimization is used, β_i,*i=1,2,…,5* are the fitting parameters of visual perception Q_j and Q_j' are the original subjective scores and the fitting scores, respectively. Before experimenting with SVR, we need to linearly scale the fractions obtained from each quality feature to the same range [0,1] to avoid the dominance of the larger values in the overall quality index fusion. Another benefit of dong this is that it avoids numerical difficulties in the calculation.

V. PREFERENCE ASSESSMENT

A. *Performance Comparison*

We compared the MFF-VQA model with the 6 existing FR objective quality assessment models, namely PSNR, FSIM, VIF, SSIM, MS-SSIM [30] and IFC [31]. The results are displayed in Table II, where the optimal performance algorithm method is highlighted in bold. Although the videos and statistical distributions of the MCML dataset differ from those of the AVT dataset, the MFF-VQA is more reliable than the other methods. In contrast, AVT dataset are more challenging.

B. *Performance of Video Frames Extract Rate*

Extracting a small amount of video frames at equal intervals instead of full video frames to evaluate video quality can greatly reduce the computation effort of the algorithm in practical applications, which is especially important in UHD VQA. The performance of each quality evaluation model at different frame extract ratios is investigated. As shown in Fig. 2, the performance of the MFF-VQA model is hardly affected by the percentage reduction in video frames extract. It still maintains higher performance when extracting a few video frames.

C. Comparison of Calculation Complexity

The computational complexity of the 7 quality assessment models is compared in Table III. The measurement is based on the computation time required to evaluate a video frame of size 3840×2160 by using a computer with Intel(R) Core(TM) i7-10750H CPU (12 cores, 2.60GHz). It can be seen that with the exception of IFC, 6 of the other 7 metrics require less than 3 seconds to complete the assessment. IFC still needs about 17 seconds to complete the job. The method we proposed is the only one that can complete the quality evaluation in 1 second. Compared with the well-performing FSIM and VIF methods, the MFF-VQA method saves nearly half of the computation time.

VI. CONCLUSION

UHD video often involves a number of compression and decompression in the transmission and storage process, due to its huge amount of data. The usefulness of existing algorithms in quickly identifying video quality remains to be considered. In this paper, a multi-feature fusion video quality assessment algorithm is proposed, focusing on the visual features that most intuitively affect the viewer's experience when watching UHD video. After rigorous experimental verification, it is proved that the MFF-VQA model has better performance and lightweight advantage than traditional algorithms on UHD video. Furthermore, it is found that the MFF-VQA maintains high performance while extracting very few video frames, which helps to apply our model to various applications such as real-time quality assessment, quality correction, etc. In the future, we hope to extend our current work in two directions. First, a more targeted UHD quality assessment can be formulated based on the experimental results of this paper. Second, a breakthrough can be made in the lightweight of UHD quality assessment algorithms.

ACKNOWLEDGMENT

This work was supported by“ National Key R&D Program of China”(2019YFB1405902).

REFERENCES

[1] Lin Zhang, Lei Zhang, Xuanqin Mou, and D. Zhang, “FSIM: A Feature Similarity Index for Image Quality Assessment,” *IEEE Trans. on Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[2] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. on Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Trans. on Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[4] M. Cheon, S.-J. Yoon, B. Kang, and J. Lee, “Perceptual Image Quality Assessment with Transformers,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, TN, USA, pp. 433–442, Jun. 2021.

[5] H. Kim, J. Kim, T. Oh, and S. Lee, “Blind Sharpness Prediction for Ultrahigh-Definition Video Based on Human Visual Resolution,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 5, pp. 951–964, May 2017.

[6] H. Kim, S. Ahn, W. Kim, and S. Lee, “Visual Preference Assessment on Ultra-High-Definition Images,” *IEEE Trans. on Broadcast.*, vol. 62, no. 4, pp. 757–769, Dec. 2016.

[7] A. Mackin, M. Afonso, F. Zhang, and D. Bull, “A Study of Subjective Video Quality at Various Spatial Resolutions,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, pp. 2830–2834, Oct. 2018,.

[8] M. Cheon and J.-S. Lee, “Subjective and Objective Quality Assessment of Compressed 4K UHD Videos for Immersive Experience,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1467–1480, Jul. 2018.

[9] “Vimeo: Watch, upload, and share HD and 4K videos with no ads.” [Online]. Available: https://vimeo.com.

[10] “YouTube.” [Online]. Available: https://www.youtube.com/.

[11] “ Can I stream Netflix in Ultra HD?,” Help Center. [Online]. Available: https://help.netflix.com/en/node/13444.

[12] “Amazon.com: Amazon Video: Amazon Video.” [Online]. Available: https://www.amazon.com/Amazon Video/b/ref=topnav_storetab_atv/ 16618650498673925?_encoding=UTF8&node=2858778011.

[13] F. Xing, Y.-G. Wang, H. Wang, J. He, and J. Yuan, “DVL2021: An ultra high definition video dataset for perceptual quality study,” *Journal of Visual Communication and Image Representation*, vol. 82, p. 103374, Jan. 2022.

[14] Li Song, Xun Tang, Wei Zhang, Xiaokang Yang, and Pingjian Xia, “The SJTU 4K video sequence dataset,” in *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt am Wörthersee, Austria, pp. 34–35, Jul. 2013.

[15] M. Cheon and J.-S. Lee, “Subjective and Objective Quality Assessment of Compressed 4K UHD Videos for Immersive Experience,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1467–1480, Jul. 2018.

[16] R. R. Ramachandra Rao, S. Goring, W. Robitza, B. Feiten, and A. Raake, “AVT-VQDB-UHD-1: A Large Scale Video Quality Dataset for UHD-1,” in *2019 IEEE International Symposium on Multimedia (ISM)*, San Diego, CA, USA, pp. 17–177, Dec. 2019.

[17] A. Mercat, M. Viitanen, and J. Vanne, “UVG dataset: 50/120fps 4K sequences for video codec analysis and development,” in *Proceedings of the 11th ACM Multimedia Systems Conference*, Istanbul Turkey, pp. 297–302, May 2020.

[18] F. Xing, Y.-G. Wang, H. Wang, J. He, and J. Yuan, “DVL2021: An ultra high definition video dataset for perceptual quality study,” *Journal of Visual Communication and Image Representation*, vol. 82, p. 103374, Jan. 2022.

[19] H.-W. Chang, X.-D. Bi, C.-Y. Du, C.-W. Mao, and M.-H. Wang, “Image Quality Evaluation Based on Gradient, Visual Saliency, and Color Information,” *International Journal of Digital Multimedia Broadcasting*, vol. 2022, pp. 1–10, Mar. 2022.

[20] Z. Ni, L. Ma, H. Zeng, C. Cai, and K.-K. Ma, “Gradient Direction for Screen Content Image Quality Assessment,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1394–1398, Oct. 2016.

[21] S. Mehravar *et al.*, “Varying weighted spatial quality assessment for high resolution satellite image pan-sharpening,” *International Journal of Image and Data Fusion*, vol. 13, no. 1, pp. 44–70, Jan. 2022.

[22] Q. Jiang *et al.*, “Single Image Super-Resolution Quality Assessment: A Real-World Dataset, Subjective Studies, and an Objective Metric,” *IEEE Trans. on Image Process.*, vol. 31, pp. 2279–2294, 2022.

[23] K. C. Noland, M. Pindoria, and A. Cotton, “Modelling brightness perception for high dynamic range television,” in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, pp. 1–6, May 2017.

[24] C. Blakemore and F. W. Campbell, “On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images,” *The Journal of Physiology*, vol. 203, no. 1, pp. 237–260, Jul. 1969.

[25] R. Rassool, “VMAF reproducibility: Validating a perceptual practical video quality metric,” in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Cagliari, Italy, pp. 1–2, Jun. 2017.

[26] R. Sotelo, J. Joskowicz, R. Puddu, D. D. Giusto and M. Anedda, "Extension to 4K Resolution of a Parametric Model for Perceptual Video Quality Estimation," *2018 IEEE Broadcast Symposium (BTS)*, pp. 1-5, 2018.

[27] T. Tian, H. Wang, L. Zuo, C. . -C. J. Kuo and S. Kwong, "Just Noticeable Difference Level Prediction for Perceptual Image Compression," in *IEEE Transactions on Broadcasting*, vol. 66, no. 3, pp. 690-700, Sept. 2020.

[28] Y. Shishikui, "Quality of 8K Ultra-High-Definition Television Viewing Experience in Practical Viewing Conditions," *in IEEE Transactions on Broadcasting*, vol. 68, no. 1, pp. 2-12, March 2022.

[29] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

[30] 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*, Pacific Grove, CA, USA, pp. 1398–1402, 2003.

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