

Copyright 2021 Society of Photo-Optical Instrumentation Engineers (SPIE). One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this publication for a fee or for commercial purposes, and modification of the contents of the publication are prohibited.

This is the accepted version of the paper available at <https://doi.org/10.1117/12.2597327>

A no-reference 3D virtual reality image quality assessment algorithm based on saliency statistics

Ajay Kumar Reddy Poreddy^a, Peter A. Kara^b, Balasubramanyam Appina^a, and Aniko Simon^c

^aIndian Institute of Information Technology, Design and Manufacturing Kancheepuram,
Chennai, India

^bBudapest University of Technology and Economics, Budapest, Hungary

^cSigma Technology, Budapest, Hungary

ABSTRACT

In this work, we propose a supervised no-reference (NR) Image Quality Assessment (IQA) model for the objective evaluation of the perceptual quality of 3D virtual reality (VR) images. To achieve such practical algorithm, we first study the scene statistics of saliency maps of the individual left and right views of VR images, and empirically model these statistics with Univariate Generalized Gaussian Distribution (UGGD). We compute the UGGD model parameters at multi-scale and multi-orient steerable subband decomposition, and introduce these features as distortion discriminables. This is followed by the computation of the entropy and normalized root mean square scores of each subband and then these values are utilized as weights to pool the individual view features. We apply the popular 2D supervised BRISQUE model on the individual views to estimate the overall spatial quality of VR images. As the last step of the algorithm, the predicted saliency score and the spatial BRISQUE score are pooled to derive the final quality score of VR images. The performance of the proposed model is evaluated on the popular LIVE 3D VR IQA dataset. The results indicate robust and competitive performance against the off-the-shelf 2D full-reference and NR supervised algorithms.

Keywords: VR image quality, objective quality assessment, supervised learning, saliency

1. INTRODUCTION

In recent years, the usage of virtual reality (VR) and its applications have increased significantly due to the launch of various end-user devices, such as HTC-Vive, Oculus Rift and Playstation. Eeden and Chow¹ states that the budget expenditure on VR devices will reach nearly 20 billion US dollars by 2022. With the rapid development of fast wireless networks and the availability of affordable VR devices, entertainment platforms like Netflix, Amazon Prime and YouTube now support 360°video viewing. Unlike 2D images, VR images are captured by using a 360°camera equipped with different lenses to cover the entire 360°of a scene. For example, Insta-360°Pro Titan is a specialized camera that can capture 2D and 3D images with eight 200°lenses and a resolution of up to 11K. Finally, the captured scenes are stitched together and stored in an equirectangular projection (ERP) format.

In case of VR, users can view the content in any direction with high resolution. However, the requirements to store these large contents with different formats and different viewing conditions pose significant challenges regarding acquisition, storage, transmission and visualization. Due to these challenges, it is necessary to properly understand the effects of the various types of distortions on the perceptual quality of VR contents.

Quality assessment (QA) is a procedure that can measure the severity of distortion. QA can be classified into subjective and objective methods. In case of subjective QA, human observers rate the quality based on perception. One may state that this procedure is hectic and cumbersome due to the involvement of test participants,

Further author information: (Send correspondence to Balasubramanyam Appina)

Ajay Kumar Reddy Poreddy: E-mail: edm20d012@iiitdm.ac.in

Peter A. Kara: E-mail: kara@hit.bme.hu

Balasubramanyam Appina: E-mail: appina@iiitdm.ac.in

Aniko Simon: E-mail: aniko.simon@sigmatechnology.se

yet these scores serve as the de facto benchmark for objective QA. Objective QA is an automatic procedure that computes the perceptual quality based on extracted features. It is classified into full-reference (FR), reduced-reference (RR) and no-reference (NR) metrics, based on the availability of the distortion-free reference content. Evidently, FR methods utilize the complete reference content, RR methods consider a few of its features and NR methods do not require any information at all.

In this paper, we propose a supervised NR QA algorithm for VR images. Our work is based on estimating the scene statistics of saliency regions and spatial Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) scores.

The rest of the paper is organized as follows: Section 2 provides a brief overview of the related VR image QA algorithms. Section 3 introduces the proposed method. Section 4 presents the evaluation results and discusses the obtained data. Section 5 concludes the paper and highlights potential future continuations of the work.

2. RELATED WORK

The generation of VR images involves several processing steps – such as sampling, quantization and post-acquisition – and each stage in the processing chain affects the perceptual quality of the source-generated content. Several FR and NR image QA (IQA) models²⁻⁹ are performed on ERP-projected views to compute the quality of VR images. However, these models may not offer sufficient performance due to the nonlinear characteristics of VR scenes. Yu *et al.*¹⁰ proposed an FR IQA model to evaluate the coding efficiency by considering the nonlinear characteristics of VR scenes. The model computes the motion trajectories of a viewport from the head-mounted display to estimate the sphere-to-plane image projection mappings. Furthermore, the viewport quality is represented using average Peak-Signal-to-Noise Ratio (PSNR), and thus, this algorithm is called Spherical PSNR (S-PSNR).

Sun *et al.*¹¹ proposed an FR IQA model by computing the pixel error values using the PSNR scores of VR images. The model estimates spherical area-based weights by using the Craster parabolic projection (CPP) model. The final quality of VR images is computed by calculating the product between PSNR values and spherical area weights. The algorithm is called the CPP-PSNR VR IQA model. Zakharchenko *et al.*¹² proposed an FR IQA model for VR images by computing the average PSNR scores of the CPP projection plane based on the field of view (FoV).

Xu *et al.*¹³ proposed multiple FR IQA models based on computing the content-based and the non-content-based PSNR scores. These models calculate a distance from the center to other regions to estimate the weights. Zhou *et al.*¹⁴ introduced a VR IQA model by computing the average structural similarity scores between test and reference images, in which pixels are weighted based on the distortion level in the corresponding projection area. The authors named this model the Weighted to Spherically-uniform SSIM (WS-SSIM).

Li *et al.*¹⁵ proposed FR IQA models based on computing the local index and global index scores between test and reference images. They compute the local index score from the attentive and stitching regions, and the global index score is measured by calculating the information loss during visual experience assessment. The work of Chen *et al.*¹⁶ presents a neural network architecture model to estimate the quality of VR images. The authors trained network architecture using eye and head movement weight maps to understand the human perceptual behavior while perceiving the VR scene. Kim *et al.*¹⁷ introduced an adversarial network learning approach to evaluate the quality of VR images. They used dataset images and corresponding difference mean opinion scores (DMOS) to train the generator and discriminator networks. They encoded the visual and positional features to estimate the final quality scores of VR images.

It is important to point out that none of the above VR IQA algorithms utilize scene statistics of the saliency of VR scenes. In this paper, we propose an NR IQA model for VR images based on performing the scene statistical analysis on saliency maps of VR images combined with a spatial estimate. Our algorithm is called **Saliency-Based Image Quality Estimator for VR scenes (SBIQE-VR)**, the mechanisms of which are presented in the following section.

3. PROPOSED METHOD

The proposed framework consists of five stages. In the first stage, we decompose six faces of left and right VR views by using the cube map projection (CMP) format. The second stage performs the computation of saliency maps of CMP decomposition, followed by steerable pyramid decomposition. The third stage involves the computation of quality-aware features. The fourth stage performs the saliency-based quality prediction and the computation of the spatial quality score. In the fifth and final stage, we pool spatial and saliency quality prediction scores to estimate the quality of VR images.

3.1 Cube map projection

Several works of the scientific literature^{18,19} perform human perceptual analysis on VR images to explore the FoV selection of observers. These works conclude that the FoV of an observer is highly present at the equator/front region. In order to achieve this, the ERP projection model is performed on the spherical VR image to convert it onto the 2D image plane. However, the ERP projection model is highly prone to geometric distortions and border artefacts. Hence, we perform CMP projection on ERP-projected images to overcome these distortions. We consider the CMP projection model based on the trade-off between accuracy and complexity. We generate six faces of each ERP image: front, back, top, bottom, left and right. Figure 1a shows the ERP projection map of the left view of the well-known ‘Cheese cake’ VR image. Figures 1b, 1c, 1d, 1e, 1f and 1g show the back, bottom, front, left, right and top faces of a CMP projection of the corresponding ERP-projected VR image, respectively. Finally, we generate six faces of each left and right views of stereoscopic VR images.



Figure 1: Illustration of the ERP projection of the left view of the ‘Cheese cake’ VR image and corresponding back face, bottom face, front face, left face, right face and top face of the CMP projection.

3.2 Saliency map generation

Yoshida *et al.*²⁰ and Veale *et al.*²¹ performed psycho-visual experiments on the Macaque visual cortex to explore the prominent region selectivity of a perceptual scene in the middle temporal (MT) visual area. They concluded that the neurons located in V1 are highly responsible for salient feature computation of perceptual scenes. Furthermore, they stated that the saliency map contains the variation of stimulus strength of prominent regions of images. We are motivated from these aforementioned psycho-visual studies, and thus, we generated a stimulus strength map by computing the saliency of a given image. Specifically, we performed the model of Achanta *et al.*²² to compute the saliency map. We chose this algorithm based on the accuracy in representing the wider perceptual strengths. Figures 2a, 2b, 2c, 2d, 2e and 2f show the saliency map of each face of the CMP projection of the left view of the ‘Cheese cake’ VR image.

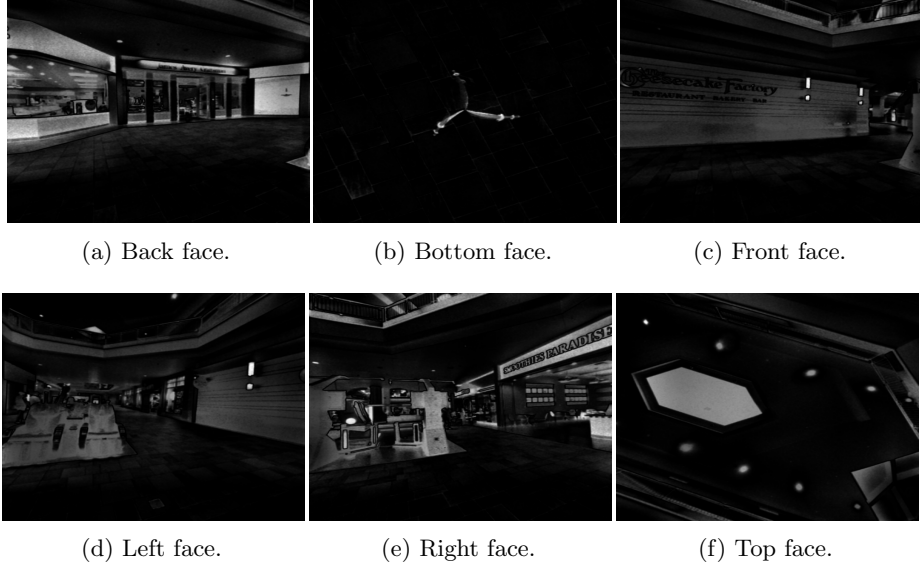


Figure 2: Illustration of saliency map of each face of the CMP projection of the left view of the ‘Cheese cake’ VR image.

In our work, we perform multiple-scale and multiple-orient steerable pyramid decomposition²³ on the saliency map of each face of VR images. The bandpass filter response in a steerable subband decomposition can mimic the properties of the human visual system (HVS). We are motivated by the earlier work of Appina *et al.*²⁴ to consider three scales (1, 2 and 3) and six orientation levels (0° , 30° , 60° , 90° , 120° and 150°) of steerable pyramid decomposition.

3.3 Quality-aware features

Consider a random vector ($\mathbf{x} \in \mathbb{R}^N$) and the corresponding UGGD distribution with zero mean is given by

$$f_X(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} \exp(-(\frac{|x|}{\beta})^\alpha), \quad (1)$$

where α and β are fitting model coefficients and $\Gamma(\cdot)$ is a gamma function. We compute the UGGD model parameters at 3 scales and 6 orientations of steerable pyramid decomposition.

Figures 3a and 3b show the saliency maps of the front face of reference-quality version and the Gaussian-distorted version of the left view of the stereoscopic ‘Cheese cake’ VR image, respectively. The saliency maps clearly show the variation with respect to the type of distortion. Figure 3c shows the log histograms of the saliency maps of different Gaussian-distorted versions of the corresponding reference image. It is clear that the shape and structure of the histograms vary with the type and level of test stimulus strength. These findings motivated us to use UGGD distribution to model the subband coefficients. The aforementioned histograms are computed at the first scale and at 0° orientation of the steerable pyramid decomposition.

Figure 3d shows the variation of the UGGD features (α and β) of the reference-quality and the Gaussian-distorted versions of the saliency map of the front face of the ‘Cheese cake’ VR image. From the plot, it is apparent that UGGD features accurately capture the perceptual quality variation of VR images. These observations motivated us to use α and β features as quality discriminate parameters. The features shown in the plot are computed at the first scale and six orientations of the steerable pyramid decomposition.

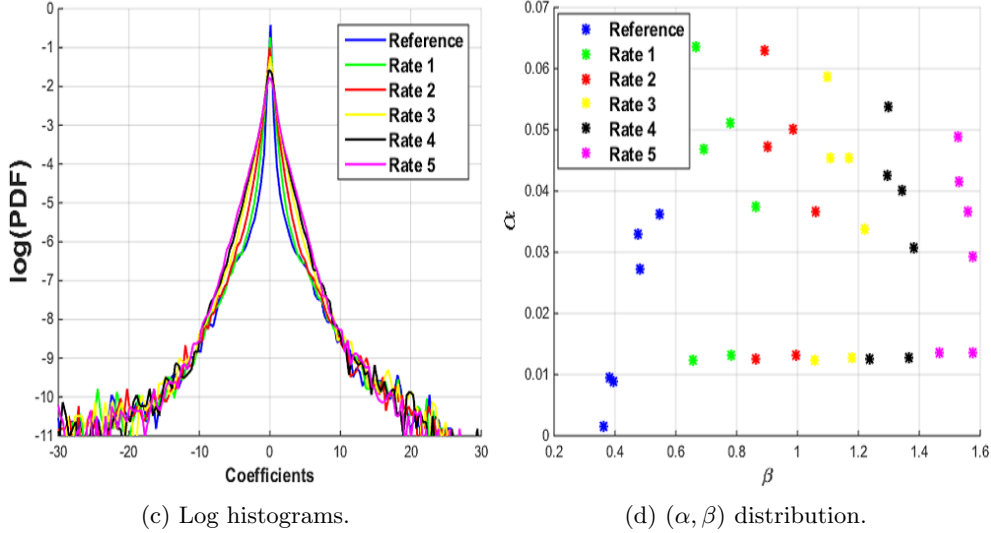
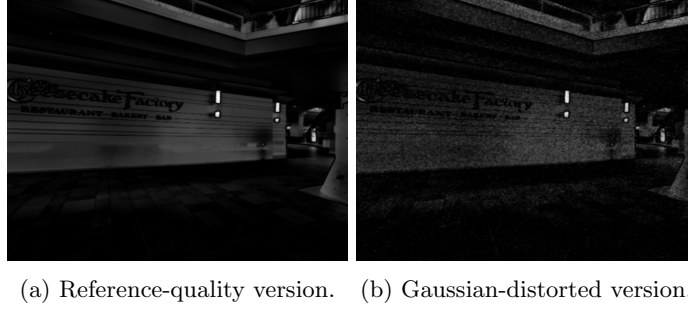


Figure 3: Illustration of the saliency map of the front face of the reference-quality and the Gaussian-distorted versions of the left view of the stereoscopic ‘Cheese cake’ VR image. Visualization of the variations of log histograms and UGGD features (α and β) for the reference-quality and the Gaussian-distorted versions.

3.4 Prediction of saliency based quality and spatial quality

As stated earlier, we perform steerable pyramid decomposition at three scales and six orientations on the saliency map of each face of the left and right views of stereoscopic VR images. We compute UGGD model parameters at each subband of the individual views of a VR image and they are denoted as

$$f_{s\alpha\beta}^l = [\alpha_{si}^l; \beta_{si}^l], \quad (2)$$

$$f_{s\alpha\beta}^r = [\alpha_{si}^r; \beta_{si}^r], \quad (3)$$

where i is the subband level ($1 \leq i \leq 18$), s represents saliency map of each face, l and r indicate left and right views, respectively, and $f_{s\alpha\beta}^l$ and $f_{s\alpha\beta}^r$ are each face-level feature sets of the aforementioned left and right views, respectively.

We compute the entropy and the normalized root mean square values of each subband to pool the UGGD features of each saliency map of individual views of VR scenes.

$$f_{\alpha\beta}^l = \sum_{s=1}^6 f_{s\alpha\beta}^l \times ent_{sji}^l \times rms_{sji}^l, \quad (4)$$

$$f_{\alpha\beta}^r = \sum_{s=1}^6 f_{s\alpha\beta}^r \times ent_{sji}^r \times rms_{sji}^r, \quad (5)$$

where *ent* represents the entropy score and it measures the average amount of information present in the subband. *rms* indicates the normalized root mean square value and it presents the band strength of a subband. We are motivated by previous works^{25,26} to consider these scores as pooling weights in the proposed algorithm. $f_{\alpha\beta}^l$ and $f_{\alpha\beta}^r$ indicate the final feature vector set of the left and right views.

We compute $f_{\alpha\beta}^l$ and $f_{\alpha\beta}^r$ feature vectors of reference-quality VR images and create an individual feature vector set of reference-quality images of a dataset.

$$f_{\alpha\beta}^{l^q} = [f_{\alpha}^{l^1}, f_{\alpha}^{l^2}, \dots, f_{\alpha}^{l^q}]; 1 \leq q \leq Q, \quad (6)$$

$$f_{\alpha\beta}^{r^q} = [f_{\alpha}^{r^1}, f_{\alpha}^{r^2}, \dots, f_{\alpha}^{r^q}]; 1 \leq q \leq Q, \quad (7)$$

where q represents the reference-quality image sequence and Q indicates the total number of such VR images in the dataset. $f_{\alpha\beta}^{l^Q}$ and $f_{\alpha\beta}^{r^Q}$ are the final feature vector sets of the left and right views of the VR image dataset. We perform Multivariate Gaussian (MVG) distribution on the individual feature vector sets and estimate the mean $(\mu_{\alpha\beta}^{l^Q}, \mu_{\alpha\beta}^{r^Q})$ and covariance $(\Sigma_{\alpha\beta}^{l^Q}, \Sigma_{\alpha\beta}^{r^Q})$ of the corresponding feature vector sets of $f_{\alpha\beta}^{l^Q}$ and $f_{\alpha\beta}^{r^Q}$.

Similarly to the reference-quality VR image set, we compute UGGD model parameters $(f_{\alpha\beta}^{l^d}, f_{\alpha\beta}^{r^d}; d$ indicates a distorted image) of individual left and right views of a stereoscopic VR test image. Furthermore, we estimate the mean $(\mu_{\alpha\beta}^{l^d}, \mu_{\alpha\beta}^{r^d})$ and covariance $(\Sigma_{\alpha\beta}^{l^d}, \Sigma_{\alpha\beta}^{r^d})$ of the corresponding left and right feature vector sets of the test image.

We compute Mahalanobis distance measure between the estimated mean and covariance of the individual views of the reference and test features to measure the quality score of the left and right views.

$$T_l = \sqrt{(\mu_{\alpha\beta}^{l^Q} - \mu_{\alpha\beta}^{l^d})^T \left(\frac{\Sigma_{\alpha\beta}^{l^Q} + \Sigma_{\alpha\beta}^{l^d}}{2} \right)^{-1} (\mu_{\alpha\beta}^{l^Q} - \mu_{\alpha\beta}^{l^d})}, \quad (8)$$

$$T_r = \sqrt{(\mu_{\alpha\beta}^{r^Q} - \mu_{\alpha\beta}^{r^d})^T \left(\frac{\Sigma_{\alpha\beta}^{r^Q} + \Sigma_{\alpha\beta}^{r^d}}{2} \right)^{-1} (\mu_{\alpha\beta}^{r^Q} - \mu_{\alpha\beta}^{r^d})}, \quad (9)$$

where T_l and T_r represent the quality prediction of the left and right views based on saliency features.

Finally, we compute the average between T_l and T_r scores to estimate the overall saliency-based quality prediction of a stereoscopic VR test image.

$$T = \frac{T_l + T_r}{2}, \quad (10)$$

where T represents the overall saliency-based quality prediction score.

We estimate the spatial quality score by evaluating the BRISQUE model⁷ on ERP projections of the left and right views.

$$B = \frac{\text{BRISQUE}(l) + \text{BRISQUE}(r)}{2}, \quad (11)$$

where B represents the overall spatial quality score of a VR image.

3.5 Overall quality score:

We compute the product between the overall saliency quality score (T) and the spatial quality score (B) to derive the overall quality score of a VR image.

$$\text{SBIQE-VR} = T \times B, \quad (12)$$

where SBIQE-VR represents overall quality prediction of a VR image.

4. RESULTS AND DISCUSSION

The performance of the proposed SBIQE-VR model was tested on the LIVE 3D VR image dataset.²⁷ The dataset consists of 15 reference-quality and 450 distorted images. The test stimuli are a combination of Gaussian blur, Gaussian noise, downsampling, stitching, VP9 loss and H.265 compression. There are five levels of distortion strength applied to the left and right views to create the dataset. The DMOS scores are published as the quality indicators of the dataset.

Table 1: Performance evaluation of the proposed algorithm in terms of LCC scores on the LIVE 3D VR image dataset.²⁷

Algorithm	Gaussian Blur	Gaussian Noise	Downsampling	Stitched	VP9	H.265
PSNR ²⁸	0.8707	0.9178	0.8448	0.7297	0.5034	0.7478
WS-PSNR ¹¹	0.8726	0.9205	0.8535	0.7801	0.6526	0.8401
SSIM ²	0.8606	0.9360	0.8343	0.4810	0.7227	0.7433
MS-SSIM ³	0.9076	0.9329	0.8865	0.6945	0.8288	0.9169
WS-SSIM ¹⁴	0.8761	0.9416	0.8630	0.5675	0.7358	0.7876
FSIM ²⁹	0.9488	0.9379	0.9096	0.8046	0.8875	0.9375
VSI ³⁰	0.9465	0.9311	0.9021	0.7216	0.8634	0.9404
GMSD ³¹	0.9409	0.9289	0.9088	0.7653	0.8622	0.9335
MDSI ³²	0.9505	0.9369	0.9221	0.7983	0.8777	0.9339
NIQE ⁸	0.9479	0.9099	0.9192	0.5093	0.4774	0.6499
B ⁷	0.8912	0.9120	0.2413	0.3562	0.4756	0.5445
T	0.5284	0.2690	0.2804	0.1752	0.5067	0.3681
SBIQE-VR	0.9036	0.9237	0.5271	0.4355	0.5499	0.5480

Table 2: Performance evaluation of the proposed algorithm in terms of SROCC scores on the LIVE 3D VR image dataset.²⁷

Algorithm	Gaussian Blur	Gaussian Noise	Downsampling	Stitched	VP9	H.265
PSNR ²⁸	0.7964	0.8964	0.8179	0.7929	0.5036	0.7750
WS-PSNR ¹¹	0.8071	0.8893	0.8321	0.8000	0.6393	0.8571
SSIM ²	0.7750	0.9107	0.8286	0.5071	0.7607	0.7821
MS-SSIM ³	0.8536	0.9143	0.8000	0.7214	0.8107	0.9250
WS-SSIM ¹⁴	0.7964	0.9143	0.8571	0.5783	0.7500	0.8179
FSIM ²⁹	0.9179	0.9143	0.7893	0.8179	0.8821	0.9357
VSI ³⁰	0.9143	0.9143	0.7964	0.7964	0.8571	0.9429
GMSD ³¹	0.9000	0.9071	0.8214	0.7812	0.8429	0.9321
MDSI ³²	0.9179	0.9179	0.7929	0.8143	0.8679	0.9429
NIQE ⁸	0.9321	0.8929	0.8821	0.1107	0.5411	0.6786
B ⁷	0.8291	0.9171	0.2037	0.3275	0.4431	0.5407
T	0.4902	0.2124	0.2777	0.1355	0.5012	0.3927
SBIQE-VR	0.8221	0.9162	0.5599	0.4239	0.5299	0.5539

We compute Linear Correlation Coefficient (LCC), Spearman Rank Order Linear Correlation Coefficient (SROCC) and Root Mean Square Error (RMSE) statistics to indicate the performance of the proposed algorithm. LCC represents the linear relationship between predicted scores and DMOS. SROCC measures the monotonic relationship between the components and RMSE measures an error between estimates and DMOS scores. These statistics are reported after performing a 4-parameter non-linear logistic fit.³³

$$f(\zeta) = \frac{\tau_1 - \tau_2}{1 + \exp\left(\frac{\zeta - \tau_3}{|\tau_4|}\right)} + \tau_2, \quad (13)$$

where ζ denotes the raw objective score, and τ_1, τ_2, τ_3 and τ_4 are the free parameters selected to provide the best fit of the predicted scores to the DMOS values.

Table 3: Performance evaluation of the proposed algorithm in terms of RMSE scores on the LIVE 3D VR image dataset.²⁷

Algorithm	Gaussian Blur	Gaussian Noise	Downsampling	Stitched	VP9	H.265
PSNR ²⁸	7.079	4.1236	7.7767	5.9731	8.0565	8.9386
WS-PSNR ¹¹	6.7828	4.2306	7.6926	5.4808	7.5422	7.4981
SSIM ²	6.4503	3.7293	7.9927	7.8591	6.9617	9.1335
MS-SSIM ³	5.5060	3.9153	6.6031	6.4280	5.4026	5.5419
WS-SSIM ¹⁴	6.1632	3.6075	7.4096	7.5257	6.8337	8.2770
FSIM ²⁹	4.1814	3.5923	5.7645	5.2133	4.6247	5.0247
VSI ³⁰	4.3320	3.7436	6.0238	6.0504	5.1298	4.7776
GMSD ³¹	4.5461	3.8439	6.1944	5.6549	5.1752	4.9873
MDSI ³²	4.3169	3.5571	5.6266	5.3215	4.7613	4.7147
NIQE ⁸	4.3405	4.5707	6.0460	7.9863	8.1521	10.3686
B ⁷	5.2258	3.7399	12.914	7.1784	7.4863	9.6573
T	11.672	8.9890	12.7739	7.6658	7.3369	10.705
SBIQE-VR	4.9347	3.4937	11.0260	7.6529	8.5025	9.5929

Tables 1, 2 and 3 show the performance evaluation of the proposed algorithm on the LIVE 3D VR IQA image dataset. We compare the proposed algorithm performance to well-known 2D FR and NR IQA models. PSNR, WS-PSNR, SSIM, MS-SSIM, WS-SSIM, VSI, GMSD and MDSI are 2D FR IQA models, while NIQE is an NR IQA model. These models were applied to the ERP maps of the individual views to estimate the quality of VR images. From the results, it is clear that the proposed NR IQA model shows consistent and competitive performance against the state-of-the-art 2D FR and NR algorithms on all distortion types of the LIVE 3D VR image dataset.

5. CONCLUSION

In this paper, we proposed a supervised NR VR IQA algorithm based on modeling the scene statistics of saliency map of VR scenes. We estimate the UGGD model parameters of saliency map subband coefficients and show that these features are distortion discriminable. We use the supervised BRISQUE model to estimate the spatial quality of VR images. Finally, the saliency-based quality and spatial quality were pooled to estimate the overall quality of VR images. In the future, we plan to extend these ideas to an unsupervised quality assessment algorithm.

ACKNOWLEDGMENTS

The scientific efforts leading to the results reported in this paper were supported in part by the Department of Science and Technology – Science and Engineering Research Board, Government of India under Grant SRG/2020/000336. The work was also supported by the National Research Development and Innovation Fund based on the charter of bolster issued by the National Research Development and Innovation Office under the auspices of the Ministry for Innovation and Technology, Hungary.

REFERENCES

- [1] van Eeden, E. and Chow, W., “Perspectives from the Global Entertainment & Media Outlook 2018–2022. [Online]. Available: <https://www.pwc.com/gx/en/entertainment-media/outlook/perspectives-from-the-global-entertainment-and-media-outlook-2018-2022.pdf>,” (2018).
- [2] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).
- [3] Wang, Z., Simoncelli, E. P., and Bovik, A. C., “Multiscale Structural Similarity for Image Quality Assessment,” in [*The Thirty-Seventh Asilomar Conference on Signals, Systems Computers*], **2**, 1398–1402 (2003).
- [4] Sheikh, H. and Bovik, A., “Image Information and Visual Quality,” *IEEE Transactions on Image Processing* **15**(2), 430–444 (2006).

- [5] Chandler, D. M. and Hemami, S. S., “VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images,” *IEEE Transactions on Image Processing* **16**(9), 2284–2298 (2007).
- [6] Ye, P. and Doermann, D., “No-Reference Image Quality Assessment Using Visual Codebooks,” *IEEE Transactions on Image Processing* **21**(7), 3129–3138 (2012).
- [7] Mittal, A., Moorthy, A. K., and Bovik, A. C., “No-Reference Image Quality Assessment in the Spatial Domain,” *IEEE Transactions on Image Processing* **21**(12), 4695–4708 (2012).
- [8] Mittal, A., Soundararajan, R., and Bovik, A., “Making a “Completely Blind” Image Quality Analyzer,” *IEEE Signal Processing Letters* **20**, 209–212 (03 2013).
- [9] Ye, P., Kumar, J., Kang, L., and Doermann, D., “Unsupervised Feature Learning Framework for No-Reference Image Quality Assessment,” in [*Conference on Computer Vision and Pattern Recognition*], 1098–1105, IEEE (2012).
- [10] Yu, M., Lakshman, H., and Girod, B., “A Framework to Evaluate Omnidirectional Video Coding Schemes,” in [*2015 IEEE International Symposium on Mixed and Augmented Reality*], 31–36 (2015).
- [11] Sun, Y., Lu, A., and Yu, L., “Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video,” *IEEE Signal Processing Letters* **24**(9), 1408–1412 (2017).
- [12] Zakharchenko, V., Choi, K. P., and Park, J. H., “Quality Metric for Spherical Panoramic Video,” in [*Optics and Photonics for Information Processing X*], **9970**, 57 – 65, SPIE (2016).
- [13] Xu, M., Li, C., Chen, Z., Wang, Z., and Guan, Z., “Assessing Visual Quality of Omnidirectional Videos,” *IEEE Transactions on Circuits and Systems for Video Technology* **29**(12), 3516–3530 (2019).
- [14] Sun, Y., Lu, A., and Yu, L., “Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video,” *IEEE Signal Processing Letters* **24**(9), 1408–1412 (2017).
- [15] Li, J., Zhao, Y., Ye, W., Yu, K., and Ge, S., “Attentive Deep Stitching and Quality Assessment for 360° Omnidirectional Images,” *IEEE Journal of Selected Topics in Signal Processing* **14**(1), 209–221 (2020).
- [16] Li, C., Xu, M., Du, X., and Wang, Z., “Bridge the Gap Between VQA and Human Behavior on Omnidirectional Video: A Large-Scale Dataset and a Deep Learning Model,” in [*Proceedings of the 26th ACM International Conference on Multimedia*], 932–940, Association for Computing Machinery (2018).
- [17] Kim, H. G., Lim, H.-T., and Ro, Y. M., “Deep Virtual Reality Image Quality assessment With Human Perception Guider for Omnidirectional Image,” *IEEE Transactions on Circuits and Systems for Video Technology* **30**(4), 917–928 (2020).
- [18] Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., and Wetzstein, G., “Saliency in VR: How Do People Explore Virtual Environments ?,” *IEEE Transactions on Visualization and Computer Graphics* **24**(4), 1633–1642 (2018).
- [19] Xu, M., Song, Y., Wang, J., Qiao, M., Huo, L., and Wang, Z., “Predicting Head Movement in Panoramic Video: A Deep Reinforcement Learning Approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(11), 2693–2708 (2019).
- [20] Yoshida, M., Itti, L., Berg, D. J., Ikeda, T., Kato, R., Takaura, K., White, B. J., Munoz, D. P., and Isa, T., “Residual Attention Guidance in Blindsight Monkeys Matching Complex Natural Scenes,” *Current Biology* **22**(15), 1429–1434 (2012).
- [21] Veale, R., Hafed, Z. M., and Yoshida, M., “How is Visual Saliency Computed in the Brain? Insights from Behaviour, Neurobiology and Modelling,” *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**(1714) (2017).
- [22] Achanta, R. and Süsstrunk, S., “Saliency Detection using Maximum Symmetric Surround,” in [*International Conference on Image Processing*], 2653–2656, IEEE (2010).
- [23] Simoncelli, E. P. and Freeman, W. T., “The Steerable Pyramid: A Flexible Architecture for Multi-scale Derivative Computation,” in [*Proceedings., International Conference on Image Processing*], **3**, 444–447 vol.3 (1995).
- [24] Appina, B., Khan, S., and Channappayya, S. S., “No-Reference Stereoscopic Image Quality Assessment Using Natural Scene Statistics,” *Signal Processing: Image Communication* **43**, 1–14 (2016).
- [25] Khan Md, S., Appina, B., and Channappayya, S. S., “Full-Reference Stereo Image Quality Assessment Using Natural Stereo Scene statistics,” *IEEE Signal Processing Letters* **22**(11), 1985–1989 (2015).

- [26] Appina, B., “A ‘Complete Blind’ No-Reference Stereoscopic Image Quality Assessment Algorithm,” in [*International Conference on Signal Processing and Communications (SPCOM)*], 1–5 (2020).
- [27] Chen, M., Jin, Y., Goodall, T., Yu, X., and Bovik, A. C., “LIVE 3D VR IQA Database. [Online]. Available: <http://live.ece.utexas.edu/research/VR3D/index.html>,” (2019).
- [28] Damera-Venkata, N., Kite, T. D., Geisler, W. S., Evans, B. L., and Bovik, A. C., “Image quality assessment based on a degradation model,” *IEEE Transactions on Image Processing* **9**(4), 636–650 (2000).
- [29] Zhang, L., Zhang, L., Mou, X., and Zhang, D., “FSIM: A Feature Similarity Index for Image Quality Assessment,” *IEEE Transactions on Image Processing* **20**(8), 2378–2386 (2011).
- [30] Zhang, L., Shen, Y., and Li, H., “VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment,” *IEEE Transactions on Image Processing* **23**(10), 4270–4281 (2014).
- [31] Zhang, B., Sander, P. V., and Bermak, A., “Gradient Magnitude Similarity Deviation on Multiple Scales for Color Image Quality Assessment,” in [*International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 1253–1257, IEEE (2017).
- [32] Ziaei Nafchi, H., Shahkolaei, A., Hedjam, R., and Cheriet, M., “Mean Deviation Similarity Index: Efficient and Reliable Full-Reference Image Quality Evaluator,” *IEEE Access* **4**, 5579–5590 (2016).
- [33] “VQEG. Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II. [Online]. Available: <http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx;accesseddate:19/05/2021>,” (2003).