

A Large-scale Compressed 360-Degree Spherical Image database: from Subjective Quality Evaluation to Objective Model Comparison

Wei Sun[†], Ke Gu[‡], Siwei Ma[#], Wenhan Zhu[†], Ning Liu[†] and Guangtao Zhai[†]

[†]Institute of Image Communication and Information Processing, Shanghai Jiao Tong University

[‡]Faculty of Information Technology, Beijing University of Technology

[#]School of Electronic Engineering and Computer Science, Peking University

Email: sunguwei@sjtu.edu.cn, guke.doctor@gmail.com, swma@pku.edu.cn, {zhuwenhan823, ningliu, zhaiguangtao}@sjtu.edu.cn

Abstract—360-degree images/videos have been dramatically increasing in recent years. But the high resolution makes it difficult to be transported, compressed and stored, and thus constrains the development of 360-degree images/videos. Therefore, it is important to study how popular coding technologies influence the quality of 360-degree images. In this paper, we present a study on subjective assessment of compressed 360-degree images and investigate whether existing objective image quality assessment (IQA) methods can effectively evaluate the quality of compressed 360-degree images. We first construct the largest compressed 360-degree image database (CVIQD2018) including 16 source images and 528 compressed ones with three prevailing coding technologies. Then, we implement 16 full reference (FR) IQA metrics, which include 10 traditional IQA metrics for 2D images and 3 PSNR-based metrics for 360-degree images, as well as 5 no reference (NR) IQA metrics and calculate the correlation between each above metric and subjective assessment in terms of three commonly used performance indices. The experiment results reveal structure information, visual saliency information and compensation for geometric distortion are crucial for evaluating the quality of compressed 360-degree images.

I. INTRODUCTION

360-degree images/videos, also called panoramic, omnidirectional or virtual reality (VR) images/videos, can provide immersive experience of real-world scenes in VR systems. With the rapid development of VR technologies in recent years, 360-degree images/videos have been widely applied in social media, live concert events or sport events, and VR movies. According to the report released by Huawei iLab [1], more than 90% of VR content is in the form of 360-degree videos. The explosive growth in 360-degree videos has attracted many researchers' interests. However, 360-degree image/video quality assessment, which can be the evaluation criterion for many image processing techniques such as image stitching, denoising and coding, has been rarely studied. What's more, the low quality video content will aggravate the untoward effect such as motion sickness in the VR system, which dramatically degrades the quality of experience (QoE) [2]. Therefore, it is important to study the image quality assessment (IQA) for 360-degree image/video.

Both objective and subjective methods are important for IQA on 360-degree images/videos. For objective IQA, there

are already some attempts to extend the traditional 2-dimensional (2D) IQA methods to 360-degree images. It is known that the spherical image should be mapped to a rectangular plane for easily storage and visualization. The equirectangular projection is widely used for VR content representation. When the users view the content in VR, the equirectangular image is remapped to the sphere domain. Hence, the current studies mainly focus on the geometric distortion occurring in the projection. For example, Yu et al. proposed the Spherical PSNR (S-PSNR) [3], which computes PSNR for the set of points uniformly distributed on a spherical surface instead of on the rectangular domain. Sun et al. [4] noticed that the geometric distortion degree of equirectangular image is proportional to the cosine power of the latitude of corresponding pixels, then proposed the Weighted Spherical PSNR (WS-PSNR), which the value of each pixel is multiplied by the cosine power of the latitude of the corresponding pixel. Zakharchenko et al. proposed Craster Parabolic Projection PSNR (CPP-PSNR) [5]. They remapped both the distorted and reference images to a Craster parabolic projection and computed the PSNR in that domain.

For subjective IQA, it still lacks a reliable 360-degree image database to evaluate the effect of the 360-degree IQA models as well as the effect of successful 2D IQA models. Upenik et al. built a small 360-degree image database including 4 reference images along with 96 distorted images [6]. But considering the number of 360-degree images is too small, it is urgent to provide a large and reliable database to offer the baseline for existing and following 360-degree IQA models.

The study in this paper is an extension of our previous work in [7]. In this paper, we attempt to build a new 360-degree image database to promote the studies of IQA for VR images. Since the high resolution of 360-degree images/videos makes it hard for transport, compression and storage, we mainly focus on how coding technologies affect 360-degree image/video quality. First, we construct a compressed VR image quality database (CVIQD2018), which consists of 16 source 360-degree images and 528 corresponding compressed images derived from three popular coding technologies, JPEG, H.264/AVC and H.265/HEVC. The Single Stimulus (SS)



Fig. 1: The source 360-degree spherical images in CVIQA2018 database. (a) teaching building; (b) playground; (c) square; (d) lake; (e) sculpture; (f) street lamp; (g) gate; (h) bicycles; (i) road; (j) classroom; (k) multimedia room; (l) rally; (m) expressway; (n) town; (o) valley.

method is adopted for gathering subject ratings because observers can only see one 360-degree image in the head-mounted display. Then we compare 13 full reference (FR) IQA metrics, which include 10 state-of-art IQA metrics for 2D images and 3 PSNR-based metrics for 360-degree images, and 5 no reference (NF) IQA metrics in terms of subjective scores using this database. Results reveal that SSIM, IW-SSIM and VSI metrics achieve a good performance on 360-degree images. PSNR-based IQA metrics for 360-degree images are superior to PSNR metric.

The remainder of this paper is organized as follows. Section II introduces the subjective assessment methodology of 360-degree images, followed by data processing and analysis for the database. Section III compares and evaluates the object IQA metrics on the database in terms of the correlation between objective predictions and subjective scores. Section IV gives the concluding remarks.

II. SUBJECTIVE QUALITY ASSESSMENT

This section is used to build the CVIQD2018 database. First, the images in the database are described. Next, subjective evaluation is applied to collect the mean opinion scores (MOSs) from subjects. Finally, the MOSs are presented and discussed.

A. Compressed VR Image Quality Database

The database includes sixteen source images, where twelve images are shot by Insta360 4K Spherical VR Video Camera

and the other four images are extracted from the test video of the JVET. The source images contain diverse scenes such as landscapes, towns, objects and persons, as shown in Figure 1. All the source images have the same resolution of 4096×2048 .

Three coding technologies are deployed in the database. The first one is the Joint Photographic Experts Group (JPEG) [8], which is a commonly used method of lossy compression for digital images. Typically, JPEG can achieve 10:1 compression with little perceptible loss in image quality, which makes it one of the most commonly employed compressed formats for photographic images on the World Wide Web. The second and third coding technologies are H.264/AVC (Advanced Video Coding) [9] and H.265/HEVC (High Efficiency Video Coding) [10], which were developed for video compression. As compared with H.264/AVC, the H.265/HEVC can lead to more than 50% performance gains in most cases. According to this, these three coding technologies are introduced in this work to establish the VR image quality database.

To be more specific, we use the JPEG to compress each reference image with quality factors ranging from 50 to 0 with an interval of -5, and use the H.264/AVC and H.265/HEVC with factors from 30 to 50 with an interval of 2. On this basis, we generate 33 compressed images from each source 360-degree image. Overall, a database including 16 reference images and 528 compressed images is built.

B. Subjective Experiment Methodology

In the following, we present the general methodology and configuration of the subjective test.

- **Method:** Several subjective testing methodologies for assessing image quality have been defined by ITU-R BT500-11 [11], including Single-Stimulus (SS), Double-Stimulus Impairment Scale (DSIS) and Paired Comparison (PC). Since the viewers only see a part of the 360-degree image that falls into the field of view (FoV) of the head mounted display (HMD), we adopt the SS method in our test.
- **Participants:** The study conducted by [12] suggests that at least 15 subjects are required in the subjective quality assessment for VR images. Here, **20 subjects** including 14 males and 6 females participated in the subjective test. Their ages range from 21 to 25. All participants have normal or corrected-to-normal vision.
- **Test Condition:** Unlike other subjective experiments conducted on the traditional displays, we do not need to consider the environment factors, e.g. viewing distance [13], ambient luminance [14], etc. The experiment was conducted in an empty room with no noise. The subjects sat on a swivel chair so they could turn their viewing direction freely.
- **Test Device:** We used the HTC VIVE as the HMD because of its excellent graphic display and high precision tracking ability. For easy operation, we designed an interaction system to automatically display the test images and collect the subjective quality scores using Unity3D software. The subjects used the controller to switch images and select the perceptual scores. Unity3D was run on a computer with 4.00GHz Intel Core i7 processor, 32GB main memory, and Nvidia GeForce GTX 1080 graphics.
- **Quality Rating:** The scales ranging from the lowest to highest perceptual quality are divided into 10 levels. The higher value means the better quality.

Before starting the experiment, the goal of this subjective test and instruction were introduced to each subject. The whole experiment involves two stages. The first stage is pilot experiment. Subjects previewed some example images which would not appear in the formal experiment so they would have an idea on how to provide their scores on the image quality. The second stage is formal experiment. 20 subjects participated in the test. They were asked to provide their perceptual opinions. The presentation order of the images was randomized for each subject. After the subjective experiment, we collected the scores of all the images rated by all the subjects and done further analysis.

C. Data Processing and Analysis

From the subjective test, we have collected all the subjects' scores. We follow the MOS calculation method as detailed in [11]. Let m_{ij} denote the raw subjective scores assigned by subject i to image j . First, the score m_{ij} needs to be converted to a Z-score Z_{ij} using

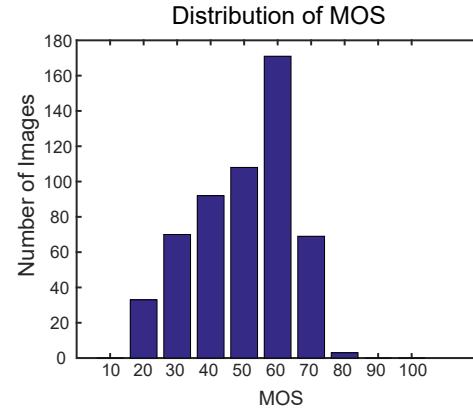


Fig. 2: Histogram of MOS in the CVIQD2018 database. The x axis represents the MOS and the y axis represents the number of images falling into the certain range of MOS. The range of each bin is from the value of bin minus 5 to the value of bin plus 5. For example, the bin of MOS 20 means the MOSs of images falling into this bin range from 15 to 25.

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} m_{ij}, \quad \sigma_i = \sqrt{\frac{1}{M_i - 1} \sum_{j=1}^{M_i} (m_{ij} - \mu_i)^2} \quad (1)$$

$$Z_{ij} = \frac{m_{ij} - \mu_i}{\sigma_i} \quad (2)$$

where N_i denotes the number of test image viewed by subject i .

After that, we discard scores from unreliable subjects by using the subject rejection procedure specified in the ITU-R BT500-11 [11]. Then Z-score Z_{ij} needs to be linearly rescaled to lie in the range of [0,100]:

$$Z'_{ij} = \frac{100(Z_{ij} + 3)}{6} \quad (3)$$

Finally, the MOS of the image j is calculated by averaging the Z'_{ij} from M_j subjects:

$$MOS_j = \frac{1}{M_j} \sum_{i=1}^{M_j} Z'_{ij} \quad (4)$$

In order to have a clear observation of those MOS values, we drew the histogram of the MOSs, which is illustrated in Figure 2. As seen, the MOSs are mainly centralized from scores "30" to "70" and the number of MOSs which are more than "80" is none. This means that the visual effect is still barely satisfied with those compressed 360 degree spherical images with a resolutions of 4K. That is, more advanced coding technologies and higher resolutions are required to improve the quality of experience (QoE) in the VR applications.

TABLE I: Performance comparison on 16 FR IQA models. We highlight the best three performing models in each column.

Metrics	JPEG			AVC			HEVC			ALL		
	SRCC	PLCC	RMSE									
PSNR	0.7342	0.8643	8.5866	0.7572	0.7592	8.0448	0.7169	0.7215	8.3279	0.7320	0.7662	9.0397
WS-PSNR	0.7520	0.8772	8.1974	0.7690	0.7708	7.8743	0.7389	0.7428	8.0515	0.7467	0.7741	8.9066
CPP-PSNR	0.7604	0.8802	8.1019	0.7726	0.7748	7.8143	0.7430	0.7469	7.9974	0.7498	0.7755	8.8816
S-PSNR	0.7729	0.8886	7.8302	0.7815	0.7854	7.6506	0.7540	0.7578	7.8471	0.7574	0.7819	8.7695
SSIM	0.9334	0.9749	3.7986	0.9451	0.9457	4.0165	0.9220	0.9232	4.6219	0.8857	0.8972	6.2140
MS-SSIM	0.9140	0.9628	4.6101	0.8794	0.8805	5.8583	0.8604	0.8610	6.1165	0.8762	0.8875	6.4836
IW-SSIM	0.9337	0.9736	3.8998	0.9471	0.9485	3.9157	0.9315	0.9338	4.3044	0.8947	0.9010	6.1031
ADD-SSIM	0.9114	0.9705	4.1192	0.8568	0.8563	6.3829	0.8408	0.8409	6.5081	0.8637	0.8780	6.7347
VSNR	0.8169	0.9082	7.1457	0.7910	0.7930	7.5301	0.7814	0.7811	7.5094	0.7691	0.7830	8.7504
IGM	0.9106	0.9594	4.8172	0.8796	0.8797	5.8779	0.8568	0.8584	6.1700	0.8588	0.8704	6.9252
VSI	0.9315	0.9690	4.2166	0.9190	0.9224	4.7736	0.8821	0.8900	5.4831	0.8927	0.9138	5.7126
GMSD	0.9132	0.9648	4.4904	0.8661	0.8632	6.2413	0.8693	0.8677	5.9780	0.8451	0.8626	7.1169
PSIM	0.9089	0.9632	4.5877	0.8241	0.8241	7.0007	0.8182	0.8213	6.8605	0.8585	0.8841	6.5751

TABLE II: Performance comparison on 5 NR IQA models.

Metrics	JPEG			AVC			HEVC			ALL		
	SRCC	PLCC	RMSE									
BRISQUE	-0.8489	0.9091	7.1137	-0.7193	0.7294	8.4558	-0.7151	0.7104	8.4646	-0.7448	0.7641	9.0751
GMLF	-0.4484	0.7801	10.6822	-0.1748	0.4864	10.8000	-0.0232	0.1491	11.8923	-0.2246	0.6134	11.1101
NIQE	-0.8585	0.8525	8.9237	-0.8358	0.8467	6.5773	-0.8681	0.8649	6.0370	-0.5126	0.5329	11.9038
QAC	0.8680	0.9537	5.1324	0.8681	0.8681	6.1348	0.8764	0.8749	5.8249	0.8299	0.8681	6.9820
SISBLIM	-0.8433	0.9186	6.7479	-0.8122	0.8547	6.4159	-0.5041	0.5620	9.9474	-0.6554	0.7439	9.4014

III. COMPARISON OF OBJECTIVE QUALITY ASSESSMENT MODELS

In this section, both FR and NR IQA metrics are implemented to explore whether existing objective IQA models can effectively evaluate the quality of compressed 360-degree images. Among FR IQA metrics, there are 3 PSNR-based IQA metrics which are specifically designed for 360-degree images and 10 traditional IQA metrics which are commonly used in 2D natural images. There are 5 general-purpose NR IQA metrics including NSS-based metrics and learning-based metrics, which are popular for 2D natural images. Then, we compute the correlation between each quality metric and subjective assessment in terms of three commonly used performance indices. After that, we give the results and discussion.

A. Full Reference IQA metrics

FR IQA models for 2D natural images have been comprehensively developed over the past decades. Many successful FR IQA metrics have been proposed to automatically predict the visual quality via a variety of strategies. The main difference existing in 2D images and 360-degree images is geometric distortion occurring in the projection. But both reference and distorted 360-degree images exist the geometric distortion and the current IQA metrics can effectively evaluate the similarity of the two images, it is deserved to investigate

how the salient IQA metrics for 2D images perform in the 360-degree images database. Here, we implement 10 FR successful IQA models which achieve good and reliable performance in popular 2D IQA databases. These IQA models are Peak signal-to-noise ratio (PSNR), Structural Similarity (SSIM) [15], Multi-SSIM (MS-SSIM) [16], Information Content Weight SSIM (IW-SSIM) [17], Analysis of Distortion Distribution for Pooling in SSIM (ADD-SSIM) [18], Visual Signal-to-Noise ratio (VSNR) [19], Internal Generative Mechanism (IGM) [20], Visual saliency-based index (VSI) [21], Gradient Magnitude Similarity deviation (GMSD) [22], and Perceptual Similarity (PSIM) [23], respectively.

Meanwhile, several IQA models for 360-degree images based on PSNR have been proposed. PSNR is a simple and widely used fidelity measure due to its simplicity and mathematical convenience. These PSNR-based algorithms extend PSNR to 360-degree images via compensating the sampling disequilibrium caused by geometric distortion. Spherical PSNR (S-PSNR) [3] computes PSNR for the set of points uniformly distributed on a spherical surface, where corresponding pixels from a reference and a distorted image are reprojected to this set. Weighted Spherical PSNR (WS-PSNR) [4] computes PSNR of the reference and distorted images when multiplying the weight to each pixel. The weight of each pixel is calculated by cosine power of latitude of the

corresponding pixel. Craster Parabolic Projection PSNR (CPP-PSNR) [5] computes the PSNR of images on the Craster parabolic projection. Therefore, both the reference image and the distorted image should be remapped to a Craster parabolic projection before.

B. No reference IQA metrics

Although FR IQA metrics have achieved remarkable performance over the decades, the requirement for a corresponding non-distorted reference image makes these metrics infeasible in practical applications, since it is hard, even impossible in most cases to obtain an ideal reference image. In contrast, NR IQA, which takes only the distorted image to be assessed as input, is more realistic and receives substantial attention in recent years. Therefore, the NR IQA is also very important for 360-degree images and it is worth exploring the performance of current NR IQA metrics in the 360-degree images. Five general-purpose NR IQA metrics are compared in this paper, which are Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [24], Natural Image Quality Evaluator (NIQE) [25], Gradient Magnitude and Laplacian Features (GMLF) [26], Quality-Aware Clustering (QAC) [27] and Six-step Blind Metric (SISBLIM) [28], respectively. BRISQUE and NIQE is based on natural scene statistics (NSS). GMLF metric utilizes the joint statistics of the gradient magnitude (GM) map and the Laplacian of Gaussian (LOG) response. QAC metric learns a set of centroids on different quality levels described by FR-IQA models. The learned centroids are then used as codebook to infer the quality of a given image. SISBLIM combines the single quality prediction of six emerging distortion types and joint effects of different distortion sources.

C. Performance of the IQAs

Before calculating performance of IQAs we list above, we map the predictions of the IQA metrics to MOSs through a five parameter logistic function for nonlinear removal:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(x\beta_3)}} \right) + \beta_4 x + \beta_5 \quad (5)$$

The three statistical indices are applied for the consistency performance comparison with predicted scores obtained from objective IQA metrics and subjective MOSs. They are respectively Spearman rank correlation coefficient (SRCC), Pearson's linear Correlation Coefficient (PLCC), Root mean square error (RMSE). The three indices have different meanings and demonstrate the prediction performance from different aspects. To specify, SRCC indicates the prediction monotonicity of the quality metric, PLCC reflects the prediction accuracy, and RMSE points out the prediction consistency. An excellent IQA metric is expected to obtain the value of SRCC and PLCC close to 1, yet the value near 0 for RMSE. We list performance of 13 FR IQA models in the Table I and 5 NR IQA models in the Table II on the single and overall compression distortion. The best three performance models are highlighted in each column in Table I.

We first observe the performance of the PSNR and PSNR-based IQA models for 360-degree images. From Table I,

the performance of all PSNR-based IQA metrics has been improved when comparing with PSNR, which shows the effect of compensating the geometric distortion in the 360-degree image. WS-PSNR and CPP-PSNR are both approximate compensation for geometric distortion. But S-PSNR is directly calculated on the spherical surface, which means the geometric distortion can be ignored. Therefore, the performance of S-PSNR is the best among these PSNR-based IQAs. However, PSNR performs not well on content-independent distortions and lots of studies reveal that PSNR does not agree with experience of human vision system, which causes all the PSNR-based IQA metrics to be inferior to the traditional successful IQA models for 2D natural images.

Then we compare the performance of 10 traditional FR IQA models for 2D natural images. From Table I, we find that SSIM, IW-SSIM and VSI metrics achieve the best three performance in terms of these traditional FR IQA models. More specially, SSIM and IW-SSIM perform better than other methods on each single distortion type, but the performance drops off on the overall database. That is because the SSIM metric detects structural change which is the main distortion existing in each compression technology. But for different compression technologies, the laws in structural distortion are not same and this reduces the whole performance of SSIM-based metrics. VSI metric achieves the superior performance on the overall performance but is slightly worse than SSIM and IW-SSIM metrics on each compression technology, which indicates the features extracted by VSI metric are less affected by the compression types than SSIM-based metrics. It inspires us that the visual saliency feature, which is extracted by VSI metric, is a good image quality feature for 360-degree images. What's more, the saliency model used in VSI metric is not specially designed for 360-degree images. Since we lack the saliency model for 360-degree images now, we hope that the saliency feature for 360-degree images can achieve the better performance.

Lastly, we compare the performance of 5 NR IQA models for 2D natural images. Although these NR IQA metrics can achieve competitive performance on several prevalent IQA databases consisting of 2D natural images, the values of PLCC and SRCC are saliently inferior than the performance of FR IQA metrics on the CVIQD2018 database. The reason may be that the FR IQA metrics take both the reference and distorted 360-degree image as inputs and may offset the effect of geometric distortion to some extent, but the NR IQA metrics only take the distorted images as input and are severely affected by geometric distortion. More specially, BRISQUE and NIQE are both the NSS-based NR IQA metrics. It is doubtful whether natural scene statistics for a 360-degree image obey a specific distribution and this issue deserves to explore in our future studies. QAC needs to partition the distorted images into overlapped patches. But these patches are heterogeneous distortion for 360-degree images because the degree of geometric distortion of the two poles is much greater than that of the equator, which makes QAC metric less effective. GMLF utilizes the joint statistics of GM map

and LOG response and SISBLIM considers multiple common distortions, which may be also less effective on 360-degree images.

From the above discussion, we can make the conclusion that some traditional FR IQA models for 2D natural images are still robust when applied to 360-degree images. But as we see, there is still much space to improve the performance for evaluating image quality of 360-degree images. First, the structure and visual saliency information are particular important for IQA models of 360-degree images. Then the compensation for geometric distortion improves the performance. Lastly, the characteristic of the 360-degree image should still be considered. For example, not the whole image can be viewed at once and users can only see the content in the viewport. This is extremely different from 2D images. It may be solved by adding the weight representing the viewing frequency to each pixel. Some researchers have tried this approach [3], [12] and improved the performance. But it is difficult to extend to other metrics when lacking a reliable saliency algorithm for 360-degree images. Therefore, it deserves deeper explorations for designing better objective IQA models of 360-degree images, especially for NR IQA models.

IV. CONCLUSION

This paper has comprehensively investigated an emerging quality assessment problem of compressed 360-degree spherical images in VR display systems. We built the largest compressed VR image quality database (CVIJD2018), including 16 sources images and 528 compressed ones under three coding technologies, i.e. JPEG, H.264/AVC and H.265/HEVC. Moreover, we compare 13 FR IQA models including 3 PSNR-based models for 360-degree images as well as 5 NR IQA models. The IW-SSIM and VSI achieve high consistency with the subjective ratings. The results also show that structure information, visual saliency information and compensation for geometric distortion are crucial factors when designing objective IQA models for 360-degree images.

V. ACKNOWLEDGEMENT

This work was supported by the National Science Foundation of China (61521062, 61527804) and Science and Technology Commission of Shanghai Municipality (15DZ0500200).

REFERENCES

- [1] HUAWEI iLab, "VR Data Report," <http://www-file.huawei.com/-/media/CORPORATE/PDF/ilab/vr-ar-cn.pdf>, accessed Jan 29, 2018.
- [2] J.-W. Lin, H. B.-L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness, "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment," in *Virtual Reality, 2002. Proceedings. IEEE*. IEEE, 2002, pp. 164–171.
- [3] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 31–36.
- [4] S. Yule, A. Lu, and Y. Lu, "Ws-psnr for 360 video objective quality evaluation," *MPEG Joint Video Exploration Team*, vol. 116, 2016.
- [5] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," in *Optics and Photonics for Information Processing X*, vol. 9970. International Society for Optics and Photonics, 2016, p. 99700C.
- [6] E. Upenik, M. Rerabek, and T. Ebrahimi, "On the performance of objective metrics for omnidirectional visual content," in *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on*. IEEE, 2017, pp. 1–6.
- [7] W. Sun, K. Gu, G. Zhai, S. Ma, W. Lin, and P. L. Callet, "Cviqd: Subjective quality evaluation of compressed virtual reality images," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3450–3454, 2017.
- [8] G. K. Wallace, "The jpeg still picture compression standard," *IEEE transactions on consumer electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
- [9] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [10] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [11] I. Recommendation, "500-11,methodology for the subjective assessment of the quality of television pictures, recommendation itu-r bt. 500-11," *ITU Telecom. Standardization Sector of ITU*, vol. 7, 2002.
- [12] M. Xu, C. Li, Z. Wang, and Z. Chen, "Visual quality assessment of panoramic video," *arXiv preprint arXiv:1709.06342*, 2017.
- [13] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Transactions on Broadcasting*, vol. 61, no. 3, pp. 520–531, 2015.
- [14] W. Sun, G. Zhai, X. Min, Y. Liu, S. Ma, J. Liu, J. Zhou, and X. Liu, "Dynamic backlight scaling considering ambient luminance for mobile energy saving," in *Multimedia and Expo (ICME), 2017 IEEE International Conference on*. IEEE, 2017, pp. 25–30.
- [15] 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.
- [16] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2. Ieee, 2003, pp. 1398–1402.
- [17] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [18] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Transactions on Broadcasting*, vol. 62, no. 2, pp. 446–456, 2016.
- [19] D. M. Chandler and S. S. Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE transactions on image processing*, vol. 16, no. 9, pp. 2284–2298, 2007.
- [20] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 43–54, 2013.
- [21] L. Zhang, Y. Shen, and H. Li, "Vsi: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014.
- [22] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans Image Process*, vol. 23, no. 2, pp. 684–695, 2014.
- [23] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 3903–3912, 2017.
- [24] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 4695–4708, 2012.
- [25] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, pp. 209–212, 2013.
- [26] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, pp. 4850–4862, 2014.
- [27] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 995–1002, 2013.
- [28] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Transactions on Broadcasting*, vol. 60, pp. 555–567, 2014.