A Survey on Objective Quality Assessment of Omnidirectional Images

Xiangjie Sui†, Shiqi Wang*, and Yuming Fang‡
† City University of Macau, Macau
E-mail: xjsui@cityu.edu.mo
* City University of Hong Kong, Hong Kong
E-mail: shiqwang@cityu.edu.hk
‡ Jiangxi University of Finance and Economics, China
E-mail: fa0001ng@e.ntu.edu.sg

***Abstract*—Omnidirectional (360°) image quality assessment (OIQA) has become crucial with the rising popularity of virtual reality (VR). Despite significant efforts to develop objective OIQA methods, there is a lack of comprehensive reviews of these methods, which hinders in-depth understanding and analysis. This paper presents an extensive survey of objective OIQA methods, classifying them into three principal types: 2D-plane-based, sphere-based, and viewport-based methods. We then analyze the limitations of current methods, highlighting challenges such as insufficient consideration of viewing conditions and diverse viewing behaviors. Moreover, this paper suggests potential future research directions, such as multi-modality approaches and personalized assessments. By systematically reviewing existing methodologies and proposing potential advancements, this paper aims to guide future research towards more accurate and efficient OIQA solutions.**

I. Introduction

The advent of virtual reality (VR) technology, particularly propelled by advancements in smart wearable devices such as the Apple Vision Pro, has significantly increased attention on 360° images. These images, created by capturing and stitching scenes from multiple fisheye lenses, form a complete $360^\circ \times 180^\circ$ view. When exploring a scene via head-mounted displays (HMDs), 360° images offer an immersive experience that closely mimics physical presence in the scene. Compared to traditional 2D images, 360° images provide enhanced realism, interactivity, and immersion, steering digital media toward higher quality and richer information content.

The processing pipeline for 360° images encompasses several stages, including image capture, stitching and projection, encoding, transmission, decoding, and rendering [1]. Throughout these stages, 360° images may suffer significant quality degradation due to various factors, making it challenging to maintain high-quality outputs. For instance, the visual quality of 360° images is often affected by the accuracy of image registration, fusion, and projection algorithms. Thus, researching effective methods to automatically provide quantitative quality indicators for 360° images is essential for optimizing the processing pipeline and ensuring high-quality outputs.

Omnidirectional image quality assessment (OIQA) can be broadly divided into subjective and objective methods. Subjective quality assessment involves human subjects who evaluate the visual quality in controlled experimental environments. In contrast, objective quality assessment aims to develop computational models that automatically infer visual quality by simulating the human visual system [2]–[8], viewing behaviors [9]–[12], natural statistical features [13], [14]. Despite significant efforts to develop objective OIQA methods, there is a lack of comprehensive reviews categorizing these methods and analyzing their limitations and future research directions [15]. This paper addresses this gap by reviewing objective OIQA methods from the perspective of common formats of 360° images: 2D projection planes, spheres, and viewports. Moreover, we discuss the limitations of current OIQA methods and suggest potential future research directions to guide future research towards more accurate and efficient OIQA solutions.

II. Objective Quality Assessment of 360° Images

360° images can be represented in several ways, primarily as 2D projection planes, spheres, and viewports (see Fig. 1). Typically, 360° images are stored in a 2D plane format achieved via equirectangular projection. During viewing, these images are decoded and reprojected onto a spherical surface for rendering and display. The viewport, representing the visual content observed by the viewer at a given moment, can be extracted using rectilinear projection. Based on these representations, current OIQA methods are categorized into 2D-plane-based, sphere-based, and viewport-based methods. This section provides a detailed introduction to these categories.

A. 2D-Plane-Based Methods

In the 2D projection plane, 360° images can be processed similarly to traditional 2D images, which allows for the extension of 2D quality metrics to OIQA. However, different map projections introduce distinct problems. For instance, equirectangular projection causes significant shape distortions near the poles (see Fig. 1). Thus, the fundamental aim of 2D-plane-based methods is to address the non-uniform sampling

TABLE I
SUMMARY OF OIQA METHODS

Type	Model	Heuristic	Data-driven	Scanpath	Viewing conditions	Weighting allocation	Re-sampling
2D-Plane-Based	[1]CPP-PSNR [16]	✓					✓
	[2]WS-PSNR [17]	✓				✓	
	WS-SSIM [18]	✓				✓	
	DeepVR-IQA [19]		✓			✓	
	[3]SAP-Net [2]		✓				
	Liu23 [20]	✓					
Shpere-Based	[4]S-PSNR [21]	✓					✓
	S-SSIM [6]	✓				✓	✓
	Sendjasni23 [3]		✓			✓	✓
Viewport-Based	[5]MC360IQA [5]		✓				✓
	[6]VGCN [4]		✓				✓
	[7]Sui21 [9]	✓		✓	✓	✓	✓
	Zhou21 [8]		✓				✓
	MFILGN [13]	✓					✓
	MP-BOIQA [14]	✓					✓
	Fang22 [22]		✓		✓		✓
	Zhang22 [7]		✓				✓
	TVFormer [10]		✓	✓		✓	✓
	PW-360IQA [23]		✓	✓			✓
	[8]Assessor360 [11]		✓	✓	✓		✓
	[9]GSR-X [12]		✓	✓	✓		✓
	Liu24 [24]		✓	✓	✓		✓

Open source (retrieved in July 2024):

[1] https://github.com/Samsung/360tools [2] https://github.com/Rouen007/WS-PSNR [3] https://github.com/yanglixiaoshen/SAP-Net [4] https://github.com/Samsung/360tools
[5] https://github.com/sunwei925/MC360IQA [6] https://github.com/weizhou-geek/VGCN-PyTorch [7] https://github.com/xiangjieSui/img2video
[8] https://github.com/TianheWu/Assessor360 [9] https://github.com/xiangjieSui/GSR

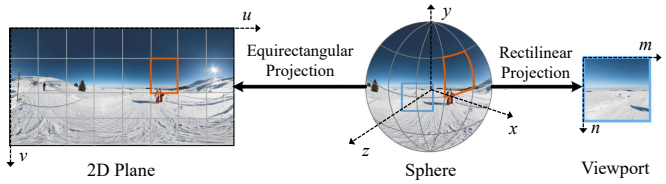

Fig. 1. An illustration of different projection of a 360° image.

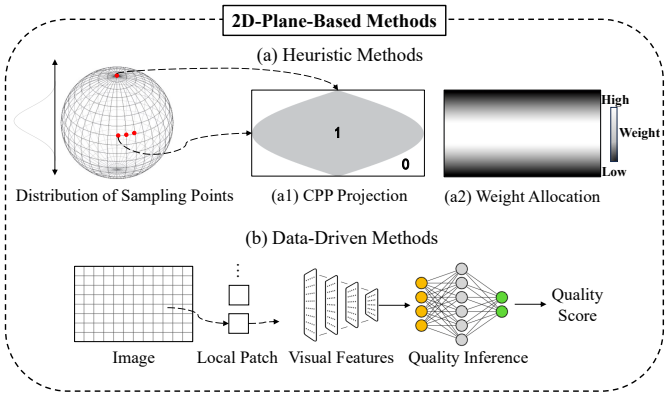

Fig. 2. Basic strategies of 2D-plane-based OIQA methods.

resulting from sphere-to-plane projection. These methods can be further divided into heuristic and data-driven approaches (see Fig. 2).

1) Heuristic Methods: Heuristic methods typically employ weighting matrices to model the density of sampling points on the spherical domain. Specifically, given an omnidirectional image $\mathbf{I}^{h\times w}$, a 2D weighting matrix $\mathbf{W}^{h\times w}$ is used to compensate for non-uniform sampling in quality inference: $\hat{Q} = \sum_{(x,y)\in\mathbf{I}} \hat{q}(x,y)\mathbf{W}(x,y)$, where $\hat{q}(x,y)$ represents the predicted quality score of the image pixel located at (x,y), and $\hat{Q}$ is the predicted overall score of the omnidirectional image. CPP-PSNR [16] employed Craster Parabolic Projection (CPP) to guarantee uniform sampling density. Given an initial weighting matrix $\mathbf{W} = 0$, the CPP projection maps uniformly distributed sample points (ϕ,θ) on the sphere to the corresponding 2D plane coordinates (x,y) by $x = \sqrt{\frac{3}{\pi}}\theta(2\cos\frac{2\phi}{3} - 1)$ and $y = \sqrt{3\pi}\sin\frac{\phi}{3}$. Then, by setting $\mathbf{W}(x,y) = 1$, the point (x,y) becomes a valid sample for Peak Signal-to-Noise Ratio (PSNR) calculation. Fig. 2 (a1) shows the resulting weighting matrix, where the number of sampling points decreases from the center to the sides, alleviating the oversampling problem near the poles. Similarly, WS-PSNR [17] and WS-SSIM [18] assigned weights to all pixels of the 360° image. Taking the equirectangular projection as an example, the value of $\mathbf{W}(x,y)$ can be determined by the corresponding stretch

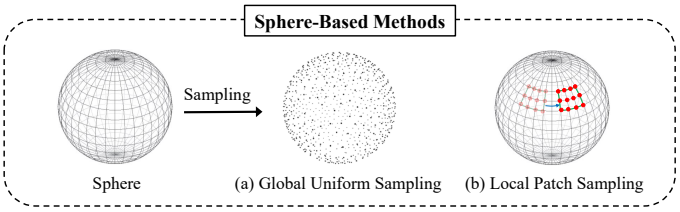

Fig. 3. Basic strategies of sphere-based OIQA methods.

ratio: $\cos\frac{y+0.5-\frac{h}{2}}{h}$, where h represents the height of the 360° image. Fig. 2 (a2) shows the resulting weighting matrix, where we can observe that the weight values decrease from the center to the sides. By considering that quality degradation leads to structural changes on the 2D projection plane, Liu *et al.* [20] proposed using histogram features to measure structural degradation and combining them with statistical and saliency features to compute the quality of 360° images. However, they did not consider the non-uniform sampling issue on the 2D projection plane.

2) Data-Driven Methods: Data-driven methods leverage deep learning to infer local image patch quality, which is then aggregated to predict the overall quality of the 360° image [2], [25]. More specifically, the 360° image is first divided into N image patches, which are then input into a pre-trained network to extract visual features $f_{1:N}$. Then, these features are fed into the quality inference module $\mathcal{Q}$ to obtain local image patch quality scores. Finally, a fusion strategy $\mathcal{W}$ is designed to integrate the local quality scores to obtain a global quality score. Overall, the goal of these methods is to maximize the prediction accuracy of the network, which could be expressed by: $\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha}} p(Q|\mathcal{W}(\mathcal{Q}(f_1), \mathcal{Q}(f_2), ..., \mathcal{Q}(f_N)); \boldsymbol{\alpha})$, where Q represents the ground-truth quality label, and $\boldsymbol{\alpha}$ denotes the learnable parameters of the network.

DeepVR-IQA [25] used ResNet-50 [26] as the visual feature extraction network and several fully connected layers to construct the quality inference module. Moreover, a fusion strategy was proposed to learn the weight of each image patch by encoding its coordinates. The final overall quality score of the 360° image was obtained by weighted averaging all local quality scores. SAPNet [2] included a self-supervised image enhancement module [27], which leveraged discrete wavelet transform features to enhance the quality of image patches. Then, the enhanced patches were regarded as "reference images" for quality inference. Finally, the overall quality of the 360° image was obtained by directly averaging the quality scores of all local image patches.

B. Sphere-Based Methods

Sphere-based methods calculate local quality estimates directly on the sphere domain, employing either global uniform sampling or local patch sampling strategies (see Fig. 3).

1) Global Uniform Sampling Methods: Global uniform sampling methods, such as S-PSNR [21], extracted a large number (*i.e.*, $655,362$) of predefined uniformly distributed sample points on the sphere. The corresponding 2D plane coordinates were computed to retrieve pixel values from distorted and reference images. Then, PSNR score was calculated to obtain the overall quality score of the 360° image.

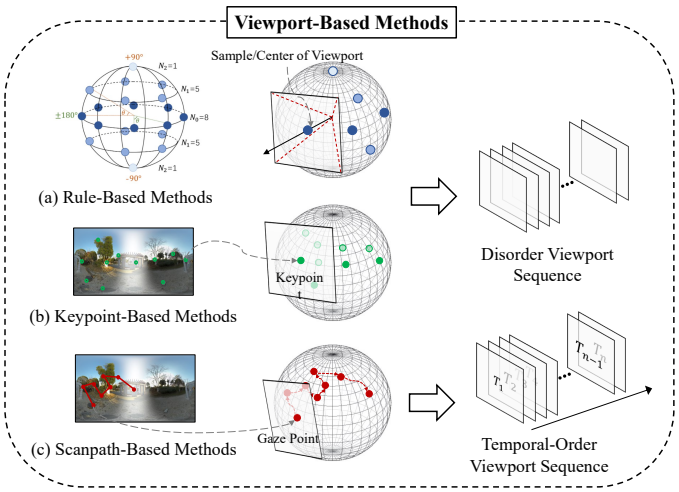

Fig. 4. Basic strategies of viewport-based OIQA methods.

2) Local Patch Sampling Methods: By considering that image patches contain structural information for quality inference, local patch sampling methods aggregate quality scores of local image patches (on sphere) to infer the overall quality of the 360° image. For instance, S-SSIM [6] calculated the Structural Similarity Index Measure (SSIM) [28] scores of image patches with a circularly symmetric Gaussian weighting function $g = \left\{g_i | \sum_{i=1}^{11\times11} g_i = 1\right\}$. These scores were weighted pooling based on coordinate stretch ratios to obtain an overall quality score. Sendjasni *et. al* [3] employed Convolutional Neural Networks (CNNs) to learn the quality-aware features of local image patches on the sphere and adaptively assigned weights to different local image patches based on the visual saliency map of the 360° image, giving higher weights to the salient regions in quality fusion.

C. Viewport-Based Methods

Given the limited field of view (FoV) in human perception, focusing only on the quality of the entire images might not be consistent with human perception. To address this problem, viewport-based methods assess the perceptual quality of 360° images by combining quality scores of viewport images (see Fig. 4). These methods are categorized into rule-based, keypoint-based, and scanpath-based approaches based on their viewport sampling strategies. Rule-based and keypoint-based methods commonly focus on learning spatial quality features of viewport images. While scanpath-based methods aim to model the spatio-temporal quality of 360° images by simulating the dynamic viewing experience of users.

1) Rule-Based Methods: Rule-based methods design sampling strategies based on characteristics of 360° images. For example, Fang22 [22] and MC360IQA [5] sampled viewports at equal intervals, with the number of samples decreasing

from the equator to the poles to maintain a relatively uniform sampling density (see Fig. 4 (a)). Additionally, MC360IQA rotated the 360° image around the y-axis at equal intervals and repeated the sampling process to augment training and testing data. In quality inference, both Fang22 [22] and MC360IQA [5] used ResNet [26] to extract pre-trained features of viewport images, and employed fully connected layers to regress these features to quality scores. Notably, Fang22 [22] encoded the coordinates of the viewport center as learnable weights to adaptively fuse the quality scores of different viewports. Several studies [8], [13], [14] used cubemap projection to obtain six minimally overlapping viewport of the 360° image. Zhou *et al.* [8] designed a multi-task learning-based OIQA quality inference network with quality score prediction as the main task and distortion type classification as an auxiliary task. MFILGN [13] and MP-BIQA [14] extracted statistical features of viewport images, and used support vector regression[13] or random forest [14] to map the statistical features to quality scores.

2) Keypoint-Based Methods: Keypoint-based methods commonly extract keypoints on the 2D projection plane, and then extract viewport centered on these points for quality inference, as illustrated in Fig. 4 (b). VGCN [4] used the Speeded Up Robust Features (SURF) algorithm [29] to extract 20 key points for viewport extraction. A graph convolutional network was constructed to learn the spatial relationships between viewports to adaptively fuse the quality scores of different viewport images. Similarly, Zhang *et al.* [7] used the Oriented FAST and Rotated BRIEF (ORB) algorithm [30] to extract keypoints for viewport extraction. By considering the potential prediction bias cased by relying on local viewports only, the studies [4], [7] measured the global quality of 360° images by measuring the quality of 2D projection image [4] or the quality of a set of viewport extracted along hypothetical scanpaths [7]. The final quality score of a 360° image was calculated by combining the local and global quality scores [4], [7].

3) Scanpath-Based Methods: Scanpath-based OIQA methods aim to predict the perceived quality 360° images by learning the spatio-temporal quality features of the viewport sequences extracted along scanpaths. The studies [9], [24] proposed OIQA methods based on human scanpath, the quality of viewport sequences were measured by mature 2D metrics [9] or the sequence model [24]. TVFormer [10] consisted of a scanpath prediction network and a quality inference network. The scanpath prediction network was based on the ViT [31] architecture, which included a memory unit modeling the memory mechanism of human. The quality inference network included the global and local branches, where the global branch learned quality features from the 2D projection plane, and the local branch learned spatio-temporal quality features of viewport sequences. Assessor360 [11] included a scanpath prediction strategy based on the entropy of viewport images. This was inspired by that human visual attention tends to focus on scenes with higher information content. In quality inference, a CNN-based module was proposed to learn the multi-scale spatial features of viewport images. Then, these features were fed into a temporal modeling module to adaptively fuse the spatial quality-aware features. The study [12] conducted a unique generative scanpath representation (GSR), which aggregates gaze-focused patches of different hypothesis users at each moment [32]. This representation provided a comprehensive global overview of dynamic perceptual experiences of multi-hypothesis users. For quality inference, the video backbone [33] was employed to learn the spatio-temporal features of GSR sequences for quality inference.

III. Limitations and Future Directions

A. Limitations

Despite notable advancements in the field of OIQA, several critical limitations persist. These limitations hinder the comprehensive application of OIQA methods, necessitating further research to address these challenges.

- **Overlook of viewing conditions**. Current OIQA methods commonly neglect the impact of varying viewing conditions, which significantly influence user scanpath patterns and perceptual quality [9], [34]. Although several studies [9], [11], [12], [22], [24] have considered factors such as the starting point of viewing and exploration time, other critical aspects remain under explored. For instance, the resolution, FoV, and display constraints of HMDs can significantly affect the perceptual quality of 360° images [9]. Therefore, a more comprehensive consideration of these viewing conditions is essential for developing robust OIQA models.
- **Neglect of diverse viewing behaviors**. Most OIQA methods rely on a single, fixed viewport sequence to predict the perceptual quality of 360° images [4], [5], [7], [8], [10], [13], [14]. Such a deterministic approach fails to account for the probabilistic nature of human viewing behaviors, which exhibit considerable variability and randomness. As a result, these methods may introduce prediction bias and fail to accurately reflect the perceptual quality experienced by users. Incorporating models that simulate diverse and dynamic viewing behaviors is crucial for enhancing the reliability of OIQA methods.
- **High computational complexity**. The high computational complexity of current OIQA methods poses significant challenges for real-time applications. One of the reasons is that the process of extracting viewports is often cumbersome and time-consuming. For example, methods such as MC360IQA [5] require the generation of a substantial number of viewports (*e.g.*, $1,080$) for a single 360° image before performing quality inference. Similarly, scanpath-based methods [9], [24] demand extensive viewport extraction. Streamlining these processes is vital for the practical deployment of OIQA methods in real-time scenarios.
- **Lack of large-scale datasets**. There is a notable scarcity of large-scale annotated datasets for training and evaluating OIQA models. To our best knowledge, the largest

available OIQA dataset [22] comprises only 258 reference images and 1032 distorted images. Insufficient training and testing data might limit the generalization ability of these methods. For example, the performance of OIQA models, especially those based on natural statistical features, is heavily dependent on the diversity and scale of the training data. Expanding the availability of large-scale and diverse datasets is critical for advancing the field.

B. Promising Directions

In addition to addressing the above limitations, several promising directions can be pursued:

- **Multi-modality OIQA methods**. Developing multi-modality OIQA methods that integrate various modalities, such as visual, text (*e.g.*, information of viewing conditions), auditory, and haptic feedback, can provide a more holistic evaluation of 360° image quality. By leveraging complementary information from different modalities, these methods can enhance the robustness and accuracy of quality assessments, offering a deeper insight of the user experience.
- **Personalized OIQA methods**. Personalized OIQA methods that account for individual user preferences and viewing behaviors can significantly improve the relevance and applicability of quality metrics. By tailoring the evaluation process to reflect the subjective quality experiences of different users, these methods can provide more accurate and user-specific assessments. This personalization can be achieved through adaptive models that learn from user interactions and feedback.
- **Quality assessment of generated 360° images**. The increasing use of AI techniques to generate 360° images presents unique challenges for quality assessment. Research in this direction might focus on developing specialized criteria and algorithms to evaluate the perceptual quality of synthetic content. These methods could consider multiple dimensions of quality (*e.g.*, authenticity and fidelity), ensuring that AI-generated 360° images meet high visual standards and provide satisfactory user experiences.

IV. Conclusion

In this survey, we have explored the current landscape of objective OIQA methods, categorizing them into 2D-plane-based, sphere-based, and viewport-based approaches. Despite significant advancements in the field, challenges such as high computational complexity, limited consideration of diverse viewing behaviors and viewing conditions, and the scarcity of large-scale datasets still persist. In addition to addressing these challenges, we point out several future research directions. By highlighting these limitations and potential research directions, we aim to guide future studies towards more precise, efficient, and comprehensive OIQA solutions. Ultimately, the evolution of OIQA methods will play a crucial role in advancing VR technology and improving the quality of experience in immersive environments.

References

[1] Y. Ye, E. Alshina, and J. Boyce, "JVET-G1003: Algorithm description of projection format conversion and video quality metrics in 360lib version 4," Joint Video Exploration Team, Turin, Italy, Rep. JVET-G1003, Tech. Rep., 2017.

[2] L. Yang, M. Xu, X. Deng, and B. Feng, "Spatial attention-based non-reference perceptual quality prediction network for omnidirectional images," in *IEEE International Conference on Multimedia and Expo*, 2021, pp. 1–6.

[3] A. Sendjasni and M.-C. Larabi, "Attention-aware patch-based CNN for blind 360-degree image quality assessment," *Sensors*, vol. 23, no. 21, 2023.

[4] J. Xu, W. Zhou, and Z. Chen, "Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 1724–1737, 2021.

[5] W. Sun, X. Min, G. Zhai, K. Gu, H. Duan, and S. Ma, "MC360IQA: A multi-channel CNN for blind 360-degree image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 64–77, 2020.

[6] S. Chen, Y. Zhang, Y. Li, Z. Chen, and Z. Wang, "Spherical structural similarity index for objective omnidirectional video quality assessment," in *IEEE International Conference on Multimedia and Expo*, 2018, pp. 1–6.

[7] C. Zhang and S. Liu, "No-reference omnidirectional image quality assessment based on joint network," in *ACM International Conference on Multimedia*, 2022, pp. 943–951.

[8] Y. Zhou, Y. Sun, L. Li, K. Gu, and Y. Fang, "Omnidirectional image quality assessment by distortion discrimination assisted multi-stream network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1767–1777, 2022.

[9] X. Sui, K. Ma, Y. Yao, and Y. Fang, "Perceptual quality assessment of omnidirectional images as moving camera videos," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 8, pp. 3022–3034, 2022.

[10] L. Yang, M. Xu, T. Liu, L. Huo, and X. Gao, "TVFormer: Trajectory-guided visual quality assessment on 360° images with transformers," in *ACM International Conference on Multimedia*, 2022, pp. 799–808.

[11] T. Wu, S. Shi, H. Cai, *et al.*, "Assessor360: Multi-sequence network for blind omnidirectional image quality assessment," in *Advances in Neural Information Processing Systems*, 2023.

[12] X. Sui, H. Zhu, X. Liu, Y. Fang, S. Wang, and Z. Wang, "Perceptual quality assessment of 360° images based on generative scanpath representation," *CoRR*, vol. abs/2309.03472, 2023. [Online]. Available: https://arxiv.org/abs/2309.03472.

[13] W. Zhou, J. Xu, Q. Jiang, and Z. Chen, "No-reference quality assessment for 360-degree images by analysis of multifrequency information and local-global naturalness," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1778–1791, 2022.

[14] H. Jiang, G. Jiang, M. Yu, T. Luo, and H. Xu, "Multi-angle projection based blind omnidirectional image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4211–4223, 2022.

[15] M. Xu, C. Li, S. Zhang, and P. L. Callet, "State-of-the-art in 360° video/image processing: Perception, assessment and compression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 5–26, 2020, ISSN: 1941-0484.

[16] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," in *Optics and Photonics for Information Processing X*, International Society for Optics and Photonics, vol. 9970, SPIE, 2016, pp. 57–65.

[17] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1408–1412, 2017, ISSN: 1558-2361.

[18] Y. Zhou, M. Yu, H. Ma, H. Shao, and G. Jiang, "Weighted-to-spherically-uniform SSIM objective quality evaluation for panoramic video," in *IEEE International Conference on Signal Processing*, 2018, pp. 54–57.

[19] H. G. Kim, H.-T. Lim, and Y. M. Ro, "Deep virtual reality image quality assessment with human perception guider for omnidirectional image," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 917–928, 2020.

[20] Y. Liu, X. Yin, Y. Wang, Z. Yin, and Z. Zheng, "HVS-based perception-driven no-reference omnidirectional image quality assessment," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023.

[21] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *IEEE International Symposium on Mixed and Augmented Reality*, 2015, pp. 31–36.

[22] Y. Fang, L. Huang, J. Yan, X. Liu, and Y. Liu, "Perceptual quality assessment of omnidirectional images," in *AAAI Conference on Artificial Intelligence*, vol. 36, 2022, pp. 580–588.

[23] A. Sendjasni and M.-C. Larabi, "PW-360IQA: Perceptually-weighted multichannel CNN for blind 360-degree image quality assessment," *Sensors*, vol. 23, no. 9, 2023.

[24] X. Liu, J. Yan, Z. Wan, Y. Fang, and H. Liu, "Blind quality assessment of panoramic images based on multiple viewport sequences," in *IEEE International Symposium on Circuits and Systems*, 2024, pp. 1–5.

[25] H. G. Kim, H. Lim, and Y. M. Ro, "Deep virtual reality image quality assessment with human perception guider for omnidirectional image," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 917–928, 2019, ISSN: 1558-2205.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[27] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *European Conference on Computer Vision*, 2018.

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

[29] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *European Conference on Computer Vision*, 2006, pp. 404–417.

[30] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: An efficient alternative to SIFT or SURF," *International Conference on Computer Vision*, pp. 2564–2571, 2011.

[31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.

[32] X. Sui, Y. Fang, H. Zhu, S. Wang, and Z. Wang, "ScanDMM: A deep markov model of scanpath prediction for 360° images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6989–6999.

[33] B. Ni, H. Peng, M. Chen, *et al.*, "Expanding language-image pretrained models for general video recognition," in *European Conference of Computer Vision*, 2022, pp. 1–18.

[34] V. Sitzmann, A. Serrano, A. Pavel, *et al.*, "Saliency in VR: How do people explore virtual environments?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1633–1642, 2018.