Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Degree in ICT for Internet and Multimedia

Visual distortions in 360° images and their impact on Quality of Experience

Supervisor:
Prof.ssa Federica Battisti

Candidate:
Mustapha Colley
2041480

Academic Year 2023-2024

Abstract

This thesis comprehensively explores visual distortions in 360° images and their consequential impacts on Quality of Experience (QoE). Leveraging insights from existing literature, a meticulously curated 360° Image Quality Dataset is introduced, facilitating nuanced analysis of distortion impacts on QoE. A detailed subjective evaluation involving 161 participants unravels the perceptual intricacies influenced by various distortions and image content. Employing Mean Opinion Scores (MOS) and ANOVA analysis, the study quantitatively assesses the perceptual impact of distortions across various types and intensity levels. The findings highlight the importance of customized image processing strategies to mitigate distortion effects. In addition, the performance of existing image quality metrics is evaluated in the context of 360-degree images, providing information on their suitability and limitations. Synthesizing key findings, this thesis advances understanding of image quality assessment methodologies for this growing medium, guiding the development of algorithms and optimization strategies to enhance user experience and satisfaction with visual content.

Index Terms—Omnidirectional image; 360°- image; Visual Distortions; Artifacts; perception; Annoyance; Dataset; Feature extraction; Visual attention; Regions of interest; Saliency; scene interpretation; Attention; Visual Perception

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, there has been a notable proliferation of immersive technologies in various aspects of our daily existence, covering entertainment, education, professional training, safety protocols, and security measures. The widespread adoption of immersive media can be attributed to its ability to generate a deep sense of physical presence within virtual environments, facilitating interactive engagement with the digital world. This capability enables the delivery of experiences that closely mimic those encountered in the tangible world, thus enhancing the efficacy and appeal of immersive technologies across diverse domains.

It should be noted that the technology required to capture and present immersive content has undergone substantial advancement, allowing users to enjoy immersive experiences at accessible price points. Furthermore, an expanding array of services are harnessing this technology to their advantage, exemplified by the proliferation of online panoramic images. These platforms enable users to seamlessly interact with the system, effortlessly manipulating their perspective by employing intuitive actions, such as clicking and dragging the cursor, to navigate in their desired viewing direction.

Among the array of technologies designed to facilitate immersive experiences, omnidirectional, or 360°, imaging stands out for its profound impact. Through the utilization of an omnidirectional acquisition configuration and a specialized rendering system, it becomes feasible to recreate virtual environments with remarkable fidelity. 360-degree imagery encapsulates the visual information that encompasses the recording point, giving viewers the freedom to observe the scene from any desired vantage point. Typically, multiple cameras or paired fisheye cameras serve as the primary acquisition devices in this set-up. Crucially, the cameras within the omnidirectional apparatus operate in synchrony, obviating the need for additional processing to stitch together video feeds.

The recorded data can be presented to users through a variety of devices, spanning from budget-friendly 2D displays and Head-Mounted Displays (HMDs) to more sophisticated and costly cave systems. Although cave systems offer an unparalleled level of immersion, rendering solutions such as HMDs still provide an acceptable Quality of Experience (QoE). Specifically, HMDs facilitate viewing of egocentric scenes, empowering users to dynamically adjust their perspective simply by moving their head or body. This intuitive interaction enhances the sense of immersion and engagement, even when utilizing more accessible rendering solutions.

The primary challenge facing the omnidirectional framework stems from its novelty, as the full scope of its intricacies and capabilities remain incompletely understood. One such challenge arises from the considerable size of omnidirectional images or videos compared to their 2D counterparts. Consequently, specialized processing and compression algorithms are imperative for efficient storage and transmission. Drawing from past experiences, such as the fluctuating trajectory of stereo entertainment systems and the limited integration of certain frameworks into everyday life, it becomes evident that assessing end-user appreciation and evaluating the impact of various processing steps on perceived quality are crucial endeavors. In essence, addressing the QoE factor is paramount in ensuring the successful adoption and utilization of omnidirectional technologies.

In this context, the principal contributions of this thesis encompass:

- *360° Image Quality Dataset:* Although there are numerous quality assessment systems and test datasets for 2D images and videos, the availability of such resources for 360° content is relatively scarce. Currently, some quality assessment metrics tailored to 360° images have been introduced and tested on specially curated test datasets. However, disparate methods for quality assessment are validated using varying testing data, complicating direct performance comparisons and the elucidation of methodological nuances. To address this gap, this thesis introduces a dataset comprising omnidirectional images annotated with the corresponding Mean Opinion Scores (MOS), facilitating benchmarking of image quality metrics and analysis of the impact of different distortions on QoE.

- *Subjective Evaluation of 360° Images:* There is a great deal of literature on the study of distortions that affect both 2D and 3D content throughout various phases of distribution. Consequently, the effects of such distortions on perceived quality have been thoroughly investigated. However, 360° media represent a relatively new frontier, and research on the distortions that may occur during its distribution is still an ongoing endeavor. In

[9], the authors present a comprehensive list of artifacts that can manifest during the distribution process and potentially influence the user's quality of experience. Building on this analysis, our study delves into the effects of various distortions and image content on the subjective quality of 360° images, contributing to a deeper understanding of the perceptual intricacies inherent to this emerging medium.

- ***Performance Evaluation of Existing Image Quality Metrics for 360° Images:*** An essential aspect that merits investigation is the assessment of image quality in omnidirectional images. A prevalent approach involves applying image quality metrics originally designed for 2D content to the realm of 360° images. In this contribution, we examine the efficacy of state-of-the-art metrics when applied to distortions typical of omnidirectional images. Through this analysis, the objective is to elucidate the suitability and limitations of existing metrics in the context of 360° imagery, thus facilitating the refinement and development of customized assessment methodologies for this evolving medium.

The thesis is structured as follows: In Section I (Introduction), an overview of the research topic and the study's objectives are provided, setting the stage for the subsequent sections. Section II (Literature Review) offers a comprehensive survey of existing work, focusing on studies related to visual distortions in 360° images. The subsections of this section dive into specific aspects such as datasets used to understand visual distortions and the evaluation of image quality metrics. Section III (Dataset Description) provides detailed information on the dataset designed for the study, including the methodology employed and the rationale behind the selection of specific data. In Section IV (Experimental results), the findings of the conducted experiments are presented. This section comprises three subsections: analysis of mean opinion scores for different types of distortion, in-depth examination of ANOVA results, and evaluation of the performance of existing image quality metrics. Finally, Section V (Conclusions) summarizes the key findings of the study and draws conclusions based on the results obtained, highlighting the implications of the research and suggesting potential avenues for future work.

Chapter 2

Literature Review

Immersive media, particularly 360-degree content, have presented novel challenges to visual distortions and their implications on QoE. A significant contribution to this field comes from *Azevedo et al.*[9], who offer an in-depth exploration of visual artifacts in 360° images.

2.1 Studies on Visual Distortions in 360° Images

The study of visual distortions in 360° images has received a lot of attention in the literature, indicating a growing interest in immersive media experiences. This section summarizes the main works that have studied the complexities of visual distortions in the context of panoramic imaging. Understanding the intricacies of these distortions is critical for improving the overall visual integrity of immersive media, as the demand for high-quality 360° material continues to increase. It presents an overview of research activities devoted to finding and characterizing visual distortions, revealing technological issues related to the capture and production of panoramic images. The literature has, for the first time, extensively reviewed the prevalent visual distortions altering signals in 360° as they traverse through various processing stages of the visual communication pipeline [9]. These distortions significantly impede the objectives of acquiring, transmitting, compressing and displaying high-quality 360° content. Therefore, it becomes crucial to understand these distortions and their potential impacts on the viewing and display quality of 360° content. In the work of *Chen et al.* [5], they constructed a more comprehensive database that includes both traditional picture distortions and VR-specific stitching distortions and offers six different types of distortion with varying degrees of severity. The six types of distortion are ***VP9 compression, H.265 compression, stitching distortion, downsampling distortion, Gaussian blur, and Gaussian noise***. In [11] they introduce two new distortions on top of Gaussian blur and Gaussian noise, which are ***JPEG compression***, and

JPEG2000 compression. Most of these distortions arise from either stitching or camera lens distortion. In practice, a mosaicking algorithm, known as stitching, produces the omnidirectional output signal. This algorithm combines the overlapped field-of-view signals of all dioptric sensors to create a wide-view panorama image [46].

2.1.1 Datasets and Understanding Visual Distortions

Recently, there has been a surge in interest in evaluating 360-degree image quality. Consequently, several efforts have been made to produce 360° image quality datasets. Table 2.1 provides further details on the referenced datasets, highlighting their key characteristics and contributions to the field.

The Compressed VR Image Quality dataset (CVIQ) [44] comprises 16 reference 360° images and 528 compressed images. Three encoding techniques were used, namely JPEG, H.264, and H.265. *Chen et al.* [5] introduced the 3D Virtual Reality (VR3D) dataset, which includes 450 distorted images derived from 15 pristine 3D VR images. These distorted images were created using six types of distortion with varying severity and were evaluated by 42 subjects, with MOS collected from 20 subjects. In [11], the authors proposed the Omnidirectional Database (OIQA), which contains 16 sources (SRC) and 320 hypothetical reference circuits (HRC) generated by compression, blurring, and noise effects. Scores from 20 subjects were also recorded in this dataset. It should be noted that these datasets were specifically developed for 360° Image Quality Assessment (IQA).

In this study, the main focus lies on the artifacts that can affect a 360° image throughout its distribution process. In particular, artifacts that manifest themselves in all four phases of the distribution pipeline(capture, encoding, transmission, and display). As such, they exert a substantial influence on the perceived visual quality of a 360° image.

While existing datasets such as CVIQ primarily address artifacts arising from compression, others like VR3D and OIQA explore a broader spectrum, encompassing compression, blur, and noise effects. However, the dataset expands on this by considering seven types of artifacts with varying levels of severity. This comprehensive approach allows the capture of a more nuanced understanding of the factors impacting the visual quality of 360° images across different stages of their distribution pipeline.

It has been demonstrated that image saliency can play a crucial role in assessing the objective quality of a 360° image. In recent years, there has been a growing interest in studying image saliency in the context of omnidirectional images. However, standard 2D saliency detection

models often fail to maintain the same performance for 360° images due to their unique characteristics. Consequently, dedicated datasets have been developed to address this challenge.

The 360-SOD dataset [25] comprises 500 equirectangular images extracted from five 360° video datasets, which include indoor and outdoor scenes. This dataset has been specifically designed to study the detection of salient objects in 360-degree images, with existing neural network-based models benchmarked against it. From the spatial distribution of salient objects in this dataset, it has been observed that typical 360° distortions significantly impact human visual attention.

Similarly, *Zhang et al.* [64] introduced the F-360iSOD dataset, which consists of 107 equirectangular images for the detection of salient objects based on fixation. The reference images in this dataset are a combination of images from Salient360! [36] and Stanford360 [42]. Eye gaze data are utilized to manually annotate salient objects, resulting in the identification of 1165 salient objects belonging to 72 categories across various aspects of real scenes. In this dataset, the performance of standard saliency estimation models is evaluated, highlighting the challenges posed by omnidirectional images.

Additionally, the ICME'17 Grand Challenge introduced Salient360! dataset [17], which comprises 85 reference images. During the test session, subjects viewed the images while their head movements and gazes were recorded. The corresponding saliency maps were generated based on the recorded head movements and fixation maps that were produced using the recorded eye gazes. This dataset provides valuable information on how viewers interact with 360° images.

Furthermore, *Sitzmann et al.* [42] captured gaze and head orientation data from 169 users while viewing 22 omnidirectional images. This dataset, called Stanford360, analyzes the viewing behaviors of VR users under various conditions, providing information on factors such as similarity in viewing behavior between users, fixation bias in VR, and the impact of scene content on viewing behavior.

Furthermore, *Upenik et al.* [47] proposed a tool to perform subjective evaluations of 360° images and videos, including a software application suitable for mobile devices and HMDs. This tool enables the viewing of omnidirectional images and videos with different projection formats, and the dataset consists of six Source Reference Circuits (SRCs) projected using both equirectangular and cube map projections and compressed using the JPEG encoder at varying compression rates.

In addition to the plethora of 360° image datasets tailored for IQA, there is also a significant body of literature dedicated to 360° video datasets. These datasets serve as invaluable resources

for evaluating video quality metrics in a range of scenarios and content types.

Some notable examples of 360° video datasets include 360-VHMD [8], VR-VQA48 [54], Sports-360 [20], IVQAD [12], PVS-HM [55], VQA-OV [24], VR-scene [56], 360-saliency [65], and Wild-360 [7]. These datasets cover diverse content genres and scenarios, allowing for comprehensive evaluations of video quality metrics. Researchers have benchmarked various quality metrics on these datasets to assess their performance in different contexts and under different types of distortions.

However, it is important to note that the work in this study primarily focuses on understanding the distortions encountered by 360° images throughout the distribution pipeline. Although these video datasets offer valuable insight into video quality assessment, the emphasis lies specifically on addressing the challenges posed by distortions in the context of 360° images.

Moving beyond technical considerations, it is essential to explore how visual distortions affect the Quality of Experience (QoE) of consumers who engage with 360-degree content. Studies by *Chen et al.* [5] and *Duan et al.* [11] explore the subjective dimensions of the viewer experience, evaluating distorted images within controlled virtual reality settings. These studies underscore the importance of developing VR quality models and benchmarking VR quality prediction algorithms using representative subjective VR quality databases.

Moreover, salient object detection (SOD) has emerged as a crucial aspect in understanding viewer attention in panoramic images. Datasets like the 360-SOD dataset[25], F-360iSOD dataset [64], and Salient360! dataset [36] provides valuable resources for studying how visual distortions impact human visual attention in 360° images.

The goal of image-based salient object detection (SOD) is to identify and classify items that draw human attention. This is a crucial first step in several tasks, including object identification [37], tracking [19], and image parsing [23]. *Jia Li et al.* [25] propose a 360 degree image-based SOD dataset of 500 high-resolution equirectangular pictures in their work. Selecting typical equirectangular pictures from five popular 360° video datasets, they manually label every item and location on these images using exact masks in a free-viewpoint manner. Through analysis of this information, they discover that the most notable features are tiny salient objects, large-scale complex scenes, and projection distortion.

Dataset	Year	Purpose	Type of collected data	No of reference images	No of images in dataset	Acquisition device	No of subjects	Dimension	Transfor. type	# Transformations	Transfor. types	Transfor. levels
CVIQ	2020	Image quality	Images, MOS	16	528	HTC Vive	20	4096 x 2048	Compression	3	JPEG, H.264, H.265	11 levels
VR3D	2020	Image quality	Images, MOS, and eye gaze	15	450	HTC Vive, Tobii eye tracker	42	4096 x 2048	Compression, distortion	6	VP9, H.265, Gaussian noise, Gaussian blur, stitching, downsampling	Quality factor for compression varied
OIQA	2018	Image quality	Images, MOS, and eye gaze	16	320	HTC Vive, aGlass eye tracker	20	11332 x 5666 to 13320 x 6660	Compression, distortion	4	JPEG, JPEG2000, Gaussian blur, Gaussian noise	5 levels
360-SOD	2020	Image Saliency	Saliency maps	6870	500	Key frames extracted from 360 video datasets	6	1024 x 1024	NA	NA	NA	NA
F-360iSOD	2020	Image Saliency	Saliency maps, annotated objects	107	107	Images from Salient360!, Stanford360	3	2048 x 1024	NA	NA	NA	NA
Salient360!	2017	Image Saliency	Images, Head movement, eye fixations, saliency maps	85	85	Oculus DK2, SMI eye tracker	40	5376 x 2688 to 18332 x 9166	NA	NA	NA	NA
Stanford360	2017	Image Saliency	Saliency maps, eye fixations	22	22	Oculus DK2, Pupil labs eye tracker	169	1920 x 1080	NA	NA	NA	NA

Table 2.1: List of the 360° image datasets available in literature and their corresponding descriptions.

2.1.2 Image quality metrics

In recent years, significant progress has been made in the development of objective metrics to evaluate image quality. This section presents a review of the 2D image quality metrics that have been applied to 360° images and of new metrics, specifically designed for omnidirectional media.

1. ***2D Image Quality Metrics:*** Quality assessment algorithms can be divided into Full Reference (FR), Reduced Reference (RR), and No Reference (NR) algorithms. The FR, NR, and RR models require the original reference image, some information about it, and only the distorted image, respectively. The list of FR and NR quality metrics that are compared in this work is provided in Table 2.2.

   The structural similarity index (SSIM) [50] computes the luminance, contrast, and structural similarity between the reference image and the distorted image. The multiscale structural similarity index (MS-SSIM) incorporates variations in viewing conditions. In the IW-SSIM metric [51], the SSIM is modified by introducing the weighted local information content pool. Similarly, the weighted mean squared error of the information content (IW-MSE) and the weighted PSNR of the information content (IW-PSNR) are also calculated in [51], incorporating information from the Laplacian pyramid transform domain. However, Visual Information Fidelity (VIF) [40]combines two parameters: information on the reference image and the amount of similar information that can be extracted from the distorted image.

   Phase congruence and low-level image features of gradient magnitude are combined to estimate the complete reference feature similarity metric (FSIM) [63]. However, the Information Fidelity Criterion (IFC) metric uses statistics from the natural world [41]. The Universal Image Quality Index (UQI) exploits image distortion as a combination of three factors: correlation loss, luminance, and contrast distortion [49]. Noise injection is another technique that is adopted in Noise Quality Measure (NQM) [10]. It is a human-visual system-inspired metric that depends on the viewing distance of the image, the variance of the luminance, and the contrast effect. Efforts have also been made to involve saliency and image content in the quality estimation task. *Zhang et al.* in [61] studied the influence of saliency on the evaluation of image quality. Distortions in an image affect visual attention, and thus estimating the saliency of the image can potentially help in predicting the quality score. The Visual Saliency-Based Index (VSI) metric first computes the local quality map of the distorted image using the saliency

map. Then, when estimating the quality score, a weighting function defined by visual saliency is used to show the influence of a local region. A content-based metric (CBM) proposed in [13] is based on fuzzy measures and the fuzzy integral quality assessment method [43]. The CBM merges the local information with the structural similarity of the image.

Blind no-reference quality metrics have recently been developed [5]. Blind image integrity using the DCT statistics metric (BLIINDS) [38] uses a machine learning-based approach that considers local discrete cosine transform coefficients. The advantage of BLIINDS is that it is not specific to any particular type of distortion. The referenceless image quality evaluation engine based on feature maps (FRIQUEE) [14] explores perceptually relevant statistics from natural scenes of distorted images in different color spaces and transform domains. The image integrity and iNtegrity assessment metric based on distortion identification (DIIVINE) [33] is based on natural scene statistics. The model first identifies the type of distortion, and, subsequently, a distortion-specific quality assessment is made.

However, unlike BLIINDS and FRIQUEE, DIIVINE depends on the type of distortion in the image. *Mittal et al.* [30] introduced a blind referenceless image spatial quality evaluator (BRISQUE) that uses scene statistics of locally normalized luminance coefficients to analyze image quality degradation.

In [38], a blind image quality index is proposed based on DCT statistics. The blind multiple pseudoreference image-based metric (BMPRI) [28] performs a quality assessment by aggravating distortions and subsequently computing similarity to the distorted images. Integrated Local Natural Image Quality Evaluator (ILNIQE) [62] uses a multivariate Gaussian model of natural image patches that integrates various image features derived from multiple cues. A no-reference perceptual quality assessment of JPEG compressed images (JPEGSNR) is presented in [52]. The blind image quality index (BIQI) [32] is a popular no-reference metric that uses a two-step framework based on statistics of the natural scene. Here, the first step is to classify the type of distortion and the second step is to employ a quality assessment technique based on the identified distortion type. The novel blind IQA method (NBIQA) [35] explores features in both the spatial and transform domains. *Gu et al.* [16] propose visual saliency-guided sampling and Gabor filtering (IQVG). The perception-based image quality assessor (PIQE) [34] estimates blockwise distortion in image patches, and the estimated quality score is inversely proportional to image distortion. The spatial-spectral-entropy-based quality index (SSEQ) [26] uses

Full Reference			No Reference		
Year	**Metric**	**Type**	**Year**	**Metric**	**Type**
2021	FGIQA	2D/360°	2023	PW-360IQA	360°
2021	BL	360°	2022	MFILGN	360°
2018	S-SSIM	360°	2019	MC360IQA	360°
2018	NCP-PSNR	360°	2020	VMAF	2D
2018	CP-PSNR	360°	2019	Voronoi	2D
2016	WS-PSNR	360°	2019	NBIQA	2D
2016	CPP-PSNR	360°	2018	BMPRI	2D
2015	S-PSNR	360°	2017	FRIQUEE	2D
2011	FSIM	2D	2016	PSS	2D
2010	IW-MSE	2D	2015	PIQE	2D
2010	IW-PSNR	2D	2015	ILNIQE	2D
2010	IW-SSIM	2D	2014	SSEQ	2D
2006	VIF	2D	2013	IQVG	2D
2005	IFC	2D	2012	NIQE	2D
2005	CBM	2D	2012	BRISQUE	2D
2004	SSIM	2D	2012	BLIINDS	2D
2004	PSNR	2D	2011	DIVINE	2D
2003	MS-SSIM	2D	2010	BIQI	2D
2002	UQI	2D	2010	DCT	2D
2000	NQM	2D	2002	JPEGSNR	2D

Table 2.2: List of full reference and no reference quality metrics and the corresponding image type.

both local spatial and spectral entropy features in distorted images. It is capable of assessing the quality of a distorted image in multiple categories of distortion. Similarly, a pseudo-structural similarity metric (PSS) is introduced in [27].

Another subsequent research in [31] introduces a completely blind natural image quality evaluator (NIQE) that does not learn from distortions. Instead, it takes advantage of the deviations from the statistical regularities in natural images. The quality assessment of single and multiple distorted images is performed using a six-step blind metric (SISBLIM) [[15]. This model estimates image quality based on individual distortions and also the joint effects of different distortions.

2. ***360° Image Quality Metrics:*** It can be seen that the above-mentioned quality metrics have been designed for assessing 2D images. However, the peculiarities of a 360° content require the assessment metric to be more flexible in understanding how and on what factors image quality depends. Especially during the capture, encoding, and transmission phase, several distortions can affect the 360° content, thus modifying the visual quality.

Assessing the effect of such 360° specific distortions on quality is a challenging problem. Interestingly, quality assessment metrics for 360° media are mostly developed for videos such as full reference metrics (e.g., SSSIM, NCP-PSNR, WS-PSNR) and the no-reference metric, M360IQA. S-PSNR [58] evaluates the quality between the different projections. Sample a set of points from the spherical surface and compute their PSNR to estimate the generalized quality of the images. However, S-PSNR uses the equirectangular projection, which accounts for distortions along the peripheral regions of the image. In this direction, *Zakharchenko et al.* [60] use the Craster parabolic projection (CPP) format and subsequently use a gamma-corrected pixel value to calculate PSNR. The weighted-to-spherically uniform quality metric (WS-PSNR) [59], [45]does not perform sampling of points in the spherical domain for the calculation of PSNR. Instead, the error in pixel values is computed, and different weights are assigned to the error pixels depending on the observed spherical area. *Xu et al.* [53] exploit the region of interest in videos to define two objective quality metrics, non-content-based perceptual PSNR (NCP-PSNR) and content-based perceptual PSNR (CP-PSNR). NCP-PSNR uses the viewing direction frequency close to the equator of each viewport. However, the viewing direction is usually driven by the content. In this regard, the CP-PSNR that uses content information has shown better performance than the NCP-PSNR.

Recently, a full reference structural similarity in the spherical domain (S-SSIM) was proposed in [6], which is related to the traditional SSIM measure for 2D images.

Sun et al. presented two variants of a non-reference quality metric (MC360IQA) [44]. In both variants, a ResNet34 architecture is used for extracting the features, and subsequently, hyper-ResNet34 merges the extracted features. The first variant uses the quality score of individual viewports, whereas the other variant uses the mean of quality scores for each viewport.

The PW-360IQA (Perceptually Weighted Multichannel CNN for 360 degree image quality assessment), introduced in [39], utilizes a multichannel CNN with a weight-sharing strategy to evaluate 360 degree image quality. By extracting visually significant viewports based on visual scan path predictions and integrating human visual system properties, PW-360IQA achieves robust performance on CVIQ and OIQA datasets while maintaining reduced computational complexity. This approach represents a novel fusion of machine learning and perceptual understanding, effectively capturing the importance of different image regions in the evaluation of 360-degree image quality.

On the contrary, MFILGN (Multifrequency Information and Local Global Naturalness for

No-Reference Omnidirectional Image Quality Assessment), proposed in [68], combines multifrequency information and local global naturalness measurements using wavelet decomposition and natural scene statistics. By leveraging insights from human visual perception and VR viewing processes, MFILGN offers an effective quality assessment method for 360-degree images. Its superior performance in no-reference quality assessment, demonstrated in publicly available OIQA databases, highlights its significance in considering both global and local image characteristics.

Furthermore, the compressed IQA metric presented in [66] addresses the need for a fine-grained evaluation of compressed images. Although the main full reference (FR) metrics excel at predicting quality at coarse-grained levels, the proposed FR-IQA metric focuses on structural and textural differences to offer improved QoE and guidance for compression algorithms. Validation in the fine-grained compression IQA (FGIQA) database showcases its superiority over standard FR-IQA metrics, with competitive performance in coarse-grained compression IQA databases.

Lastly, the reference image-based objective blur level (BL) metric proposed in [1] accurately assesses image quality degraded by motion blur. Unlike traditional metrics such as SSIM and PSNR, the BL metric utilizes point spread function/blur kernel analysis to provide more reliable quality assessments. Experimental results demonstrate its superior performance in describing perceived image quality, particularly in challenging low-light and low-texture conditions, thus offering promise for effective image-deblurring processes. In the following section, a detailed description of the proposed database is presented.

Chapter 3

Dataset Description

A. Source Reference Circuits (SRC)

The proposed dataset comprises 10 equirectangular Source Reference Circuits (SRCs), chosen from the pool of 85 images available in Salient360! dataset presented in the ICME Grand Challenge [36], [17]. The selection process mainly focused on two key features: Spatial Information (SI) and Colorfulness (C), which are commonly used criteria for SRC selection [5].

Spatial information serves as an indicator of edge energy and helps to estimate the complexity of an image [44], [57]. A higher SI value corresponds to a more intricate scene. On the other hand, colorfulness is measured by the intensity and diversity of colors present within an image [5], [18].

The 10 SRCs chosen for inclusion in this thesis are depicted in Figure3.1.

ImageName	Spatial Information	Colorfulness
Image 16.jpg	7.36	0.79
Image 18.jpg	7.45	0.19
Image 19.jpg	7.23	0.49
Image 2.jpg	7.79	0.78
Image 20.jpg	7.72	0.32
Image 24.jpg	7.60	0.51
Image 28.jpg	7.50	0.20
Image 4.jpg	7.34	0.36
Image 42.jpg	7.55	0.23
Image 47.jpg	7.65	0.39

Table 3.1: Spatial Information and Colorfulness for Each Reference Image.

B. Hypothetical Reference Circuits (HRC)

Creating a dataset for Image Quality Assessment (IQA) requires the definition of Hypothetical Reference Circuits (HRC) derived from the available Source Reference Circuits (SRC) [22]. Throughout the 360-degree media distribution pipeline, various distortions may occur during

the capture, encoding, transmission, and display phases [9]. Existing datasets such as CVIQ employ encoding techniques such as JPEG, H.264, and H.265 to generate HRC. On the contrary, data sets such as VR3D and OIQA utilize a combination of encoding techniques (e.g., JPEQ, H.265, VP9), noise, and blurring distortions. Currently, there is no standardized procedure for determining which encoding/rendering/distortion techniques to adopt when generating HRCs.

In this thesis, I have chosen three types of artifacts: distortion, blur, and noise, to generate

(a) Image 2   (b) Image 4   (c) Image 16   (d) Image 18

(e) Image 47   (f) Image 42   (g) Image 28

(h) Image 24   (i) Image 19   (j) Image 20

Figure 3.1: The list of images used as reference images in our dataset.

my hypothetical reference circuits (HRCs). The rationale for selecting these artifacts stems from their potential appearance during three of the four phases of the 360° media distribution pipeline[9], namely acquisition, encoding, and rendering. These artifacts represent common challenges encountered throughout the distribution process and are therefore relevant for evaluating image quality across various stages of the pipeline.

- *Distortion:* Photographic lenses commonly introduce optical aberrations during image acquisition, such as pincushion and barrel distortions. Pincushion distortion causes the magnification of the image to increase with distance from the optical axis, resulting in the inward folding of peripheral regions towards the center while leaving the central region unaffected. On the contrary, barrel distortion causes magnification to decrease with distance from the optical axis, causing the central region to protrude outward. Both artifacts create distinct spatial patterns in the acquired image and have a significant impact during the capturing phase of omnidirectional images.

- *Blur:* Blur effects are observed during both the image acquisition and the display phases.
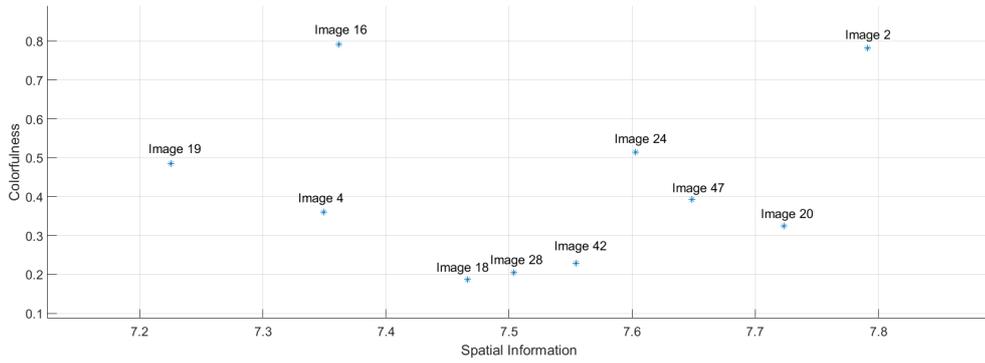
Figure 3.2: Scatter plot showing the relationship between Spatial Information and Colorfulness for the 10 reference images.

We investigate the effects of motion blur and Gaussian blur on the perceived quality of distorted 360° images. During image capture, discontinuities between different cameras can lead to motion blur in the captured scene. Similarly, when viewing omnidirectional images using a Head-Mounted Display (HMD), there is often a delay between the movement of the user's head and the response of the display, resulting in motion blurring that may cause motion sickness in end users[67]. Additionally, the coding phase involves projection and reprojection between spherical and planar representations, which can inherently cause blurring effects, particularly Gaussian blurring.

- *Noise:* The limitations of sensors in 360-degree multicamera rigs lead to noise during the capture phase. For this dataset, we considered three types of noise: salt and pepper, Gaussian, and Poisson noise, as mentioned in [2]. These types of noise introduce disturbances into the captured images, affecting their overall quality and fidelity.

The effect of each artifact is studied systematically by considering various intensity levels. With the exception of Poisson noise, all other artifacts are evaluated at three severity levels: low, average, and high. Table 3.2 provides a summary of the distortions considered in this document, along with the corresponding stage in the 360° content distribution pipeline where they occur. Furthermore, the table outlines the parameter settings for each distortion level to achieve the desired severity levels.

In Figure3.3, an example of hypothetical reference circuits (HRC) generated from a sample Source Reference Circuit (SRC) is presented, illustrating the impact of these distortions on image quality.

To collect Mean Opinion Scores (MOS) for each stimulus, a comprehensive subjective test is

Transformation Type	Content Distribution Stage		Parameters
	Capturing	Capturing, Display	
Barrel Distortion	3	-	1, 3, 5
Pincushion Distortion	3	3	0.3, 0.6, 0.9
Motion Blur	3	3	10, 15, 25
Gaussian Blur	-	3	4, 7, 10
Salt and Pepper Noise	3	-	1, 3, 5
Gaussian Noise	3	-	1, 2, 3
Poisson Noise	1	-	1

Table 3.2: Transformations and Parameters for Content Distribution Stages.

conducted. Details of the subjective test methodology and the MOS collection strategy are elaborated in the following section.

(a) Reference(IMG 2) (b) Barrel Low (c) Barrel Avg (d) Barrel High

(e) Pincushion Low (f) Pincushion Avg (g) Pincushion High (h) Motion Blur Low

(i) Motion Blur Avg (j) Motion Blur High (k) Gauss. Blur Low (l) Gauss. Blur Avg

(m) Gauss.Blur High (n) Salt&Pep. Low (o) Salt&Pep. Avg (p) Salt&Pep. High

(q) Gauss. N Low (r) Gauss. N Avg (s) Gauss. N High (t) Poisson Noise
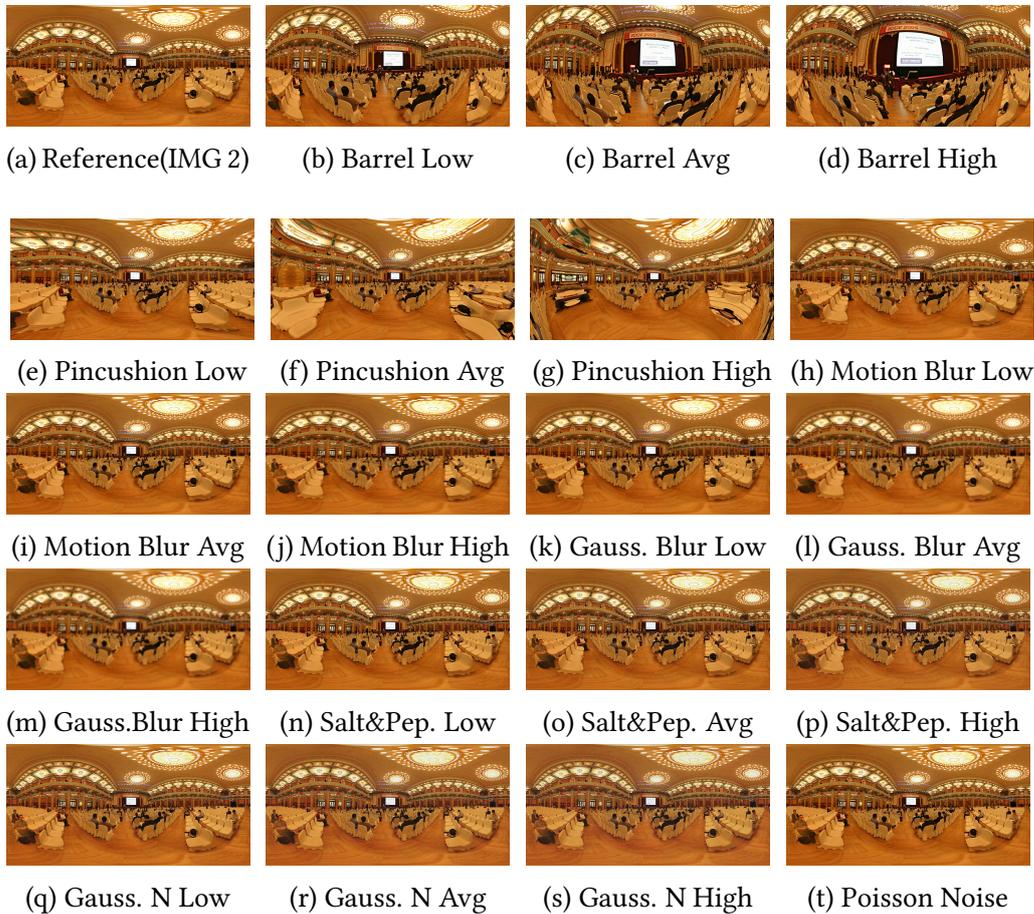
Figure 3.3: List of images depicting all the types of distortion (and their level of severity) used in this work. We used Image # 2 as a sample image for this example

3.1 Subjective Test

A. Subjective experiment methodology

According to the ITU-R BT500-11 recommendation [21], subjective testing methodologies typ-

18

ically encompass three basic types: single stimulus (SS), double stimulus (DS), and paired comparison (PC). Given that 360° images are viewed using a Head-Mounted Display (HMD), where subjects can only see a portion of the image known as the Field of View (FOV), we opted for the SS method in our test setup.

The test protocol was organized into distinct sections, including instructions, visual acuity test, color vision test, training session, rest period, and the main test session. Clear instructions were provided to the participants, which outline the purpose of the experiment and how to accurately score the stimuli. Subsequently, visual acuity and color vision tests were conducted to ensure the suitability of the participants for the study.

The participants then underwent a training session to familiarize themselves with the experimental environment and the task at hand. Following the training session, a two-minute break was provided to allow the participants to rest before beginning the actual test session. This structured approach ensured that the participants were adequately prepared and comfortable throughout the subjective evaluation process.

B. Subjects

A total of 161 subjects participated in the test, consisting of 117 men and 44 women, according to the estimates provided in [3]. The age range of the subjects ranged from 18 to 61 years, with an average age of 27 years. Participants with comorbid conditions such as vertigo, anxiety, and amblyopia were excluded from the test to ensure the reliability and validity of the results.

C. Experiment Duration

The total duration of the experiment for each subject was 15 minutes. Specifically, the instructions provided to the subjects prior to starting the test and the visual acuity and color tests were completed in less than 5 minutes. The training session lasted approximately 3 minutes, while the actual test session lasted approximately 6 minutes. This duration covers the time required to record the scores provided by the subjects.

D. Training

At the beginning of the experiment, subjects were briefed on the purpose of the test and detailed instructions were provided regarding the test procedure. The test consisted of two phases: training and testing.

During the training phase, subjects were assigned to view images using the HMD device. This phase aimed to familiarize participants with the HMD and the free viewing feature of the 360° content. Subjects were instructed to provide image quality scores during the training phase on a scale of 1 to 5, where 1 corresponds to poor quality and 5 to excellent. The purpose of this exercise was to familiarize participants with the rating system and the experimental setup. In

particular, scores collected during the training phase were not recorded for analysis.

It is important to note that the image used in the training phase was identical for all participants. Furthermore, the images used for training purposes were different from those used in the subsequent test phase to ensure unbiased evaluation during the testing session.

E. Apparatus and Environment

The image stimuli were presented using an HTC Vive Head-Mounted Display (HMD). The Witoo VR photo viewer application for HTC Vive facilitated the display and manipulation of images through a controller. Each test session comprised 20 images, 19 distorted, and one as a reference.

The experiment was conducted in a quiet environment to minimize external distractions. Subjects were seated on swivel chairs to facilitate free viewing of 360-degree content. Participants were allowed to wear glasses or contact lenses if necessary.

Test conditions such as viewing distance and luminance were not controlled, as the experiments were conducted using an HMD rather than traditional displays. Most of the subjects were unexperienced with the use of HMDs and had no previous experience in video quality assessment.

Throughout training and test sessions, a gray image always followed each test stimulus. This protocol, according to [21], aimed to mitigate any bias arising from factors such as color, contrast, and brightness. During viewing the gray image, subjects were instructed to rate the perceptual quality of the stimuli they had just viewed using a rating scale ranging from 1 to 5.

Chapter 4

Experimental Results

The analysis carried out in this thesis delves into various aspects of image quality evaluation, including MOS for distortion types, and ANOVA analysis. Each component offers valuable insight into the perceptual quality of images and sheds light on the influence of different distortions and characteristics on visual perception. Through a comprehensive exploration of these factors, the objective is to deepen the understanding of image quality metrics and guide the development of image processing algorithms and compression techniques.

The primary objective of these analyses is to comprehensively evaluate the image quality metrics and their implications for image processing and analysis. Through MOS and ANOVA results, the perceptual impact of distortions and image characteristics is explained, guiding the development of algorithms prioritizing high perceptual quality. In addition, the analysis aims to contribute to understanding image quality assessment methodologies and their diverse applications.

4.1   Analysis of Mean Opinion Scores for Distortion Types

The MOS for distortion types measures the perceived image quality of participants at different distortion levels. Contrasting the MOS of each distortion type with the original image (Figure 4.1) allows one to assess the extent of perceptual degradation induced by various distortions. Scrutinizing the variability in MOS and confidence intervals provides insight into the consistency and reliability of participants' judgments, crucial for interpreting perceptual differences. Moreover, MOS serves as a crucial metric for evaluating the effectiveness of image processing techniques and distortion reduction algorithms. By comparing MOS across different types of distortion and severity levels, we can assess the relative perceptual impact of each distortion and identify areas for improvement in image processing pipelines.

Statistical analysis was aimed at assessing the MOS associated with various image distortions. It involved loading the dataset from an Excel file and computing summary statistics, including mean, standard deviation, and 95% confidence interval for each type of distortion; see Tables 4.1 and 4.2.

After importing the data into R using the `readxl` package, summary statistics were calculated for each column in the dataset, excluding non-numeric identifiers. These statistics provided insight into the central tendency and variability of MOS for each type of distortion. Furthermore, the distortion types were grouped on the basis of their characteristics, facilitating the visualization of MOS trends across different categories. Individual plots were generated for each distortion group using the package `ggplot2`, displaying the mean MOS along with error bars representing the confidence interval 95%.

Arranging the plots into a grid layout enabled a comprehensive comparison of MOS across various distortion types, facilitating the identification of significant differences in perceived image quality and contributing to a deeper understanding of the impact of distortions on viewer perception.

Distortion Type	Distortion Level	Code
Pincushion	High	PH
Pincushion	Average	PA
Pincushion	Low	PL
Barrel	High	BH
Barrel	Average	BA
Barrel	Low	BL
Motion Blur	High	MBH
Motion Blur	Average	MBA
Motion Blur	Low	MBL
Gaussian Blur	High	GBH
Gaussian Blur	Average	GBA
Gaussian Blur	Low	GBL
Gaussian Noise	High	GNH
Gaussian Noise	Average	GNA
Gaussian Noise	Low	GNL
Salt and Pepper Noise	High	SPH
Salt and Pepper Noise	Average	SPA
Salt and Pepper Noise	Low	SPL
Poisson Noise	-	PN
No Distortion	-	ND

Table 4.1: Distortion types and levels with corresponding codes.

Original Image (OR):

- The original image serves as the baseline reference for comparing the perceptual qual-

ity of distorted images. It has a mean opinion score of 3.95, indicating a high level of perceived quality among the participants.

Impact of Distortion Types:

- **Pincushion Distortions:** Distortions characterized by a pincushion effect (PH, PA, PL) generally exhibit lower mean opinion scores compared to the original image. This suggests that these distortions introduce perceptual degradation, especially at higher levels of distortion.

- **Barrel Distortions:** Barrel distortion types (BH, BA, BL) also show a decrease in mean opinion scores relative to the original image. However, the Barrel Low (BL) demonstrates a slightly higher mean opinion score, indicating a less severe impact compared to other barrel distortions.

Variable	Mean	Standard Deviation
PH	2.57	± 1.27
PA	2.89	± 1.27
PL	3.18	± 1.19
BH	2.22	± 1.06
BA	2.39	± 1.19
BL	2.98	± 1.15
MBH	1.71	± 0.96
MBA	2.06	± 0.88
MBL	2.57	± 0.96
GBH	1.38	± 0.84
GBA	1.38	± 0.70
GBL	1.76	± 0.90
GNH	2.53	± 0.98
GNA	2.93	± 0.98
GNL	3.07	± 0.98
SPH	2.65	± 0.98
SPA	2.86	± 1.03
SPL	3.34	± 0.99
PN	3.60	± 0.98
OR	3.95	± 1.07

Table 4.2: Summary Statistics.

- **Motion Blur:** Distortions related to motion blur (MBH, MBA, MBL) are perceived to have a significant impact on image quality, as indicated by their lower mean opinion scores. Motion blur tends to reduce image clarity and detail, resulting in lower perceived quality.

- **Gaussian Blur:** Gaussian blur distortions (GBH, GBA, GBL) exhibit the lowest mean opinion scores among all types of distortion, indicating substantial degradation in image quality. These distortions often result in a loss of sharpness and detail.

- **Noise Distortions:** Distortions caused by noise (GNH, GNA, GNL, SPH, SPA, SPL, PN) show varying levels of impact on perceived quality. Gaussian Noise Low (GNL), Salt and Pepper Noise Low (SPL), and Poisson Noise (PN) have comparatively higher mean opinion scores, suggesting less perceptual degradation compared to other noise distortions.

Variability in Perceptual Quality:

- The standard deviations associated with mean opinion scores provide insights into the variability of participants' opinions. Higher standard deviations indicate a wider range of perceptions among participants for a given distortion type.

- The confidence intervals, represented by the upper and lower bounds, offer a range within which the true mean opinion score lies with a certain level of confidence. They help to assess the reliability of the mean opinion scores obtained from the study.
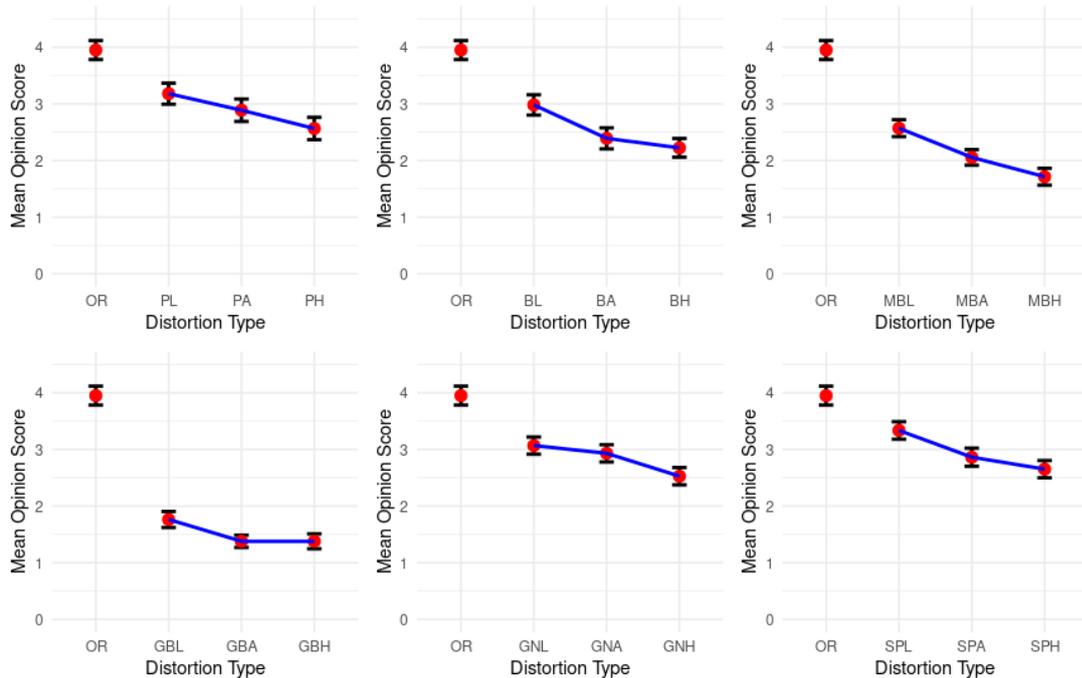
Figure 4.1: Mean Opinion Scores for Different Distortion Types.

In comparing Pincushion distortions (PH, PA, PL) with Gaussian blur distortions (GBH, GBA, GBL), we can recognize slight differences in their impact on perceived image quality and their adherence to the Just Noticeable Difference (JND) threshold.

Pincushion distortions manifest as geometric distortions resembling a pincushion effect, while Gaussian blur distortions primarily affect image sharpness and clarity. When analyzing the mean opinion scores (MOS) for both distortion types:

Pincushion Distortions:

- PH (Pincushion High): It shows a lower mean opinion score compared to the original image, indicating perceptual degradation. Being categorized as "High," PH introduces a significant level of distortion, likely exceeding the JND threshold, resulting in a noticeable degradation in perceived quality.

- PA (Pincushion Average): Similarly to PH, PA also shows a lower mean opinion score, suggesting perceptual degradation, although potentially less severe. Classified as "Average," PA introduces a moderate level of distortion, likely noticeable but possibly less pronounced than PH.

- PL (Pincushion Low): Although it still exhibits a lower mean opinion score than the original image, PL may have a relatively higher score compared to PH and PA, implying potentially less severe degradation. Classified as "Low," the PL introduces minimal distortion, possibly less noticeable to participants compared to higher levels of pincushion distortion.

(a) Pincushion High   (b) Pincushion Average   (c) Pincushion Low

Figure 4.2: Picushion Distortion

Gaussian Blur Distortions:

- GBH (Gaussian Blur High): It displays the lowest mean opinion score among all types of distortion, indicating substantial perceptual degradation. Classified as "High," GBH introduces a significant level of blur distortion, likely highly noticeable to participants, and surpassing the JND threshold.

- GBA (Gaussian Blur Average): Similar to GBH, GBA also shows a low mean opinion score, suggesting perceptual degradation, although potentially less severe. Classified as

"Average," GBA introduces a moderate level of blur distortion, noticeable but potentially less pronounced than GBH.

- GBL (Gaussian Blur Low): Although still showing a lower mean opinion score than the original image, GBL may have a relatively higher score compared to GBH and GBA, implying potentially less severe degradation. Classified as "Low," GBL introduces minimal blur distortion, possibly less noticeable compared to higher levels of Gaussian blur.

| (a) Gaussian Blur High | (b) Gaussian Blur Average | (c) Gaussian Blur Low |

Figure 4.3: Gaussian Blur Distortions

The analysis reveals the perceptual impact of various types of distortion on image quality relative to the original image. Understanding how different distortions affect perceived quality is essential for developing image processing algorithms and compression techniques that prioritize maintaining high perceptual quality. By addressing the identified perceptual challenges associated with specific distortion types, developers can improve the overall user experience and satisfaction with visual content across various applications and platforms.

Both Pincushion and Gaussian blur distortions lead to perceptual degradation compared to the original image, with higher mean opinion scores generally indicating less degradation. However, at higher levels of distortion, both PH and GBH exhibit pronounced degradation, likely surpassing the JND threshold. Pincushion distortions may have more noticeable effects at lower levels compared to Gaussian blur distortions, which exhibit significant degradation even at minimal levels. Integrating this understanding into the development of image processing algorithms can help prioritize maintaining image quality above the perceptual threshold, thus enhancing user experience and satisfaction with visual content.

4.2 In-depth Analysis of ANOVA Results

The ANOVA, or one-way analysis of variance, is a statistical test used to assess whether the means of three or more populations are equivalent, contrasting this notion with the proposition that not all the means of the population share this uniformity. For the ANOVA test, the

(a) Pincushion Distortion

	PL	PA	PH
PL	NA	0.03	0.00
PA	0.03	NA	0.02
PH	0.00	0.02	NA

(b) Barrel Distortion

	BL	BA	BH
BL	NA	0.00	0.00
BA	0.00	NA	0.18
BH	0.00	0.18	NA

(c) Motion Blur

	MBL	MBA	MBH
MBL	NA	0.00	0.00
MBA	0.00	NA	0.00
MBH	0.00	0.00	NA

(d) Gaussian Blur

	GBL	GBA	GBH
GBL	NA	0.00	0.00
GBA	0.00	NA	1.00
GBH	0.00	1.00	NA

(e) Gaussian Noise

	GNL	GNA	GNH
GNL	NA	0.21	0.00
GNA	0.21	NA	0.00
GNH	0.00	0.00	NA

(f) Salt and Pepper Noise

	SPL	SPA	SPH
SPL	NA	0.00	0.00
SPA	0.00	NA	0.06
SPH	0.00	0.06	NA

Table 4.3: ANOVA P-Value Differences Tables.

samples must maintain consistent variances, indicated by σ^2 [4]. This test is particularly relevant in the context of analyzing image quality metrics in different levels of distortion, where the objective is to determine if there are significant differences in the variables measured between various types of distortion.

Complementing the MOS assessment, ANOVA quantitatively evaluates the statistical significance of differences in variables measured between distortion levels. The P-values in Table 4.3 distinguish whether the observed differences are likely due to random chance or are attributable to distortion effects. Lower p-values underscore distortions that significantly impact perceived image quality, which warrants attention in image processing algorithms.

ANOVA is essential because it provides a formal statistical framework for the following.

- **Quantifying Differences:** ANOVA allows us to quantify the degree of variation in MOS between different distortion levels. By calculating the p-values, we can determine whether the observed differences are likely due to chance or if they represent true disparities in the perception of image quality.

- **Comparison of multiple groups:** With ANOVA, we can compare the means of multiple groups simultaneously, rather than making pairwise comparisons between each pair of distortion levels. This comprehensive approach helps identify general trends and patterns in perception of image quality.

- **Controlling Type I Error:** ANOVA helps control the risk of Type I error, which occurs

when we mistakenly conclude that there is a significant difference between groups when there is actually not much difference. Setting a significance level (such as $\alpha = 0.05$), ANOVA allows us to draw statistically valid conclusions about differences in image quality perception.

- **Interpreting Results:** The ANOVA test provides valuable information on which distortion types have a significant impact on image quality perception and which may be less influential. This information is essential to prioritize areas of focus in image processing algorithms and optimization strategies.

  In general, ANOVA is a powerful statistical tool that helps researchers rigorously analyze and interpret differences in image quality perception between various types of distortion, contributing to the advancement of image processing techniques and optimization algorithms.

The ANOVA statistical analysis provided above in Table 4.3 provides valuable information on the effects of various image distortions on the variables measured in different types of distortion. Each distortion type, such as Pincushion Distortion, Barrel Distortion, Motion Blur, Gaussian Blur, Gaussian Noise, Salt and Pepper Noise, Poisson Noise, and Original Image (serving as the baseline), is represented by three levels: Low, Average, and High. Here is an integrated analysis:

- **Pincushion Distortion (PH, PA, PL):**

  - Significant differences are observed between high, average, and low levels of pincushion distortion, indicating a notable impact on the measured variables. For example, the degree of pincushion distortion applied to the images significantly affects the image quality, as seen in the ANOVA results ($p < 0.05$).

  - Comparison of the PL, PA, and PH groups confirms these findings, revealing substantial variations in the measured variable at different levels of pincushion distortion.

- **Barrel Distortion (BH, BA, BL):**

  - Similarly, there are statistically significant differences between high, average, and low levels of barrel distortion, suggesting different impacts on the measured variables. These effects are evident in the ANOVA results ($p < 0.01$), highlighting the importance of barrel distortion correction in image processing.

- The BL, BA, and BH group analysis further emphasizes these distinctions, under-lining the need to consider different levels of barrel distortion when optimizing image quality.

- **Motion Blur (MBH, MBA, MBL):**

  - Significant distinctions are noted between high, average, and low levels of motion blur, indicating varying influences on the measured variables. The results of the ANOVA ($p < 0.001$) underscore the importance of motion blur reduction techniques in enhancing image clarity.

  - The MBL, MBA, and MBH group analysis supports these observations, demonstrating substantial variations in the measured variable across different levels of motion blur.

- **Gaussian Blur (GBH, GBA, GBL):**

  - The analysis reveals statistically significant differences between high, average and low levels of Gaussian blur, suggesting varying impacts on the measured variables. The ANOVA results ($p < 0.001$) highlight the importance of Gaussian blur reduction in preserving the details of the image.

  - Comparison of the GBL, GBA, and GBH groups further elucidates these differences, emphasizing the need for careful adjustment of Gaussian blur levels for optimal image quality.

- **Gaussian Noise (GNH, GNA, GNL):**

  - Significant differences are detected between high, average, and low levels of Gaussian noise, emphasizing the differential effects on the measured variables. The results of the ANOVA ($p < 0.001$) underscore the importance of controlling Gaussian noise for accurate image analysis.

  - The analysis of the GNL, GNA, and GNH groups reinforces these findings, highlighting the distinct impacts of different levels of Gaussian noise on the measured variable.

- **Salt and Pepper Noise (SPH, SPA, SPL):**

– Similarly to other distortions, statistically significant differences are observed between high, average, and low levels of salt and pepper noise, indicating varying impacts on the measured variables. The ANOVA results ($p < 0.001$) suggest marked disparities in the measured variable between different levels of salt and pepper noise.

– The comparison of the SPL, SPA and SPH groups provides further insight into these differences, highlighting the need to mitigate the noise from salt and pepper for accurate image analysis.

- **Poisson Noise (PN):**

  – Although not included in the pairwise comparison tables, Poisson noise likely introduces unique effects on the measured variables, necessitating further investigation to understand its impact fully.

- **Original Image (OR):**

  – Serving as a baseline, the original image condition provides a reference point to assess the impact of distortions on the measured variables. Although not subjected to specific distortion levels, the original image is crucial to understanding the magnitude of distortion effects.

In summary, ANOVA analysis provides valuable insight into the significance of distortion intensity across different types in influencing the measured variables. These findings are essential to refine image processing techniques and optimize image quality in various applications. By integrating MOS and ANOVA, this analysis advances our understanding of image quality assessment and informs the development of more effective image processing techniques. Taking into account both subjective and objective metrics facilitates the development of algorithms that prioritize high perceptual quality, thus improving the user experience and satisfaction with visual content in various applications and platforms.

4.3 Existing Metrics Performance

The performance of existing IQA metrics in the data set is evaluated using the evaluation technique recommended by VQEG [48]–[29]. This technique calculates the correlation between the predicted scores by the objective metrics and the Mean Opinion Scores (MOS) provided by the

subjects. The three metrics used in this evaluation are the Spearman correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC), and the root mean square error (RMSE). SRCC measures the strength and direction of the association between the predicted scores generated by the IQA metrics and the MOS provided by human subjects. It assesses how well the predicted scores rank image quality compared to human judgments. A higher SRCC indicates better alignment between the predicted and human-assigned scores.

PLCC, similar to SRCC, measures the correlation between the predicted scores and MOS. However, it specifically evaluates the linear relationship between the two sets of scores. A higher PLCC value signifies a stronger linear relationship between predicted and human scores.

RMSE measures the average magnitude of the differences between predicted and actual scores. It indicates how accurately the IQA metrics predict the MOS values. A lower RMSE indicates a better accuracy of the IQA metrics in predicting human judgments.

Several state-of-the-art quality metrics (both full reference and no reference) are tested on the dataset. The full reference metrics include UQI, PSNR, SSIM, MS-SSIM, IFC, VIF, NQM, CBM, IW-MSE, IW-PSNR, IW-SSIM, and FSIM. No-reference metrics include BIQI, BLIINDS, IQVG, NIQE, DIVINE, BRISQUE, PIQE, BMPRI, FRIQUEE, JPEGSNR, DCT, NBIQA, PSS, SSEQ, and ILNIQE.

In general, these metrics collectively provide a comprehensive evaluation of the performance of IQA metrics by assessing correlation, linearity, and prediction accuracy, thus facilitating the selection of appropriate IQA metrics for specific applications.

Table 4.4 below presents the performance of various full reference IQA metrics in the proposed data set. The metrics evaluated include Universal Image Quality Index (UQI), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Multiscale Structural Similarity Index (MS-SSIM), Image Fidelity Criterion (IFC), Visual Information Fidelity (VIF), Complexity-Based Metric (CBM), Noise Quality Measure (NQM), Improved PSNR (IW-PSNR), Improved SSIM (IW-SSIM), Improved Mean Squared Error (IW-MSE) and Feature Similarity Index (FSIM). Observations on the table are captured as follows after applying the various metrics.

- **SRCC:** The highest correlation coefficient is observed for VIF (0.43), indicating a relatively strong alignment with human judgments. On the contrary, the lowest correlation coefficient for IW-MSE is observed (-0.34), indicating a negative correlation and poor alignment with human judgments.

- **PLCC:** VIF demonstrates the highest linear correlation (0.48) with human judgments, indicating a strong linear relationship, while IW-MSE exhibits the lowest linear correla-

Metric	UQI	PSNR	SSIM	MS-SSIM	IFC	VIF	CBM	NQM	IW-PSNR	IW-SSIM	IW-MSE	FSIM
SRCC	0.30	0.39	0.22	0.04	0.40	0.43	0.22	0.22	0.34	0.18	-0.34	0.36
PLCC	0.26	0.25	0.15	0.03	0.46	0.48	0.21	0.09	0.28	0.06	0.03	0.04
RMSE	1.31	1.26	1.54	0.93	1.08	1.21	1.33	1.42	1.26	2.12	1.62	2.08

Table 4.4: State-of-the-art Full reference metrics performance on the proposed dataset.

tion (0.03), suggesting a weak linear relationship with human judgments.

- **RMSE:** MS-SSIM has the lowest prediction error (0.93), indicating higher precision to predict image quality, while IW-SSIM has the highest prediction error (2.12), suggesting lower precision to predict image quality.

Metrics such as VIF and CBM consistently perform well across all evaluation criteria, showing higher correlation coefficients, stronger linear relationships, and lower prediction errors. These metrics are promising for accurately assessing image quality. In contrast, metrics such as IW-MSE and IW-SSIM exhibit poorer performance, with lower correlation coefficients, weaker linear relationships, and higher prediction errors, making them less suitable for precise image quality assessment. The selection of an appropriate IQA metric depends on specific application requirements, with metrics such as VIF and CBM, known for their robust performance across various evaluation criteria, being preferred in applications where accurate image quality assessment is crucial.

Table 4.5 presents the performance of various no-reference IQA metrics on the proposed dataset. The metrics evaluated include blind image quality index (BIQI), blind/referenceless image spatial quality assessor (BLIINDS), image quality visual grader (IQVG), natural image quality assessor (NIQE), deep image quality assessor (DIVINE), blind/referenceless image spatial quality assessor (BRISQUE), perceptual image quality assessor (PIQE), blockiness measure for perceived ringing image (BMPRI), reliable quality evaluation (FRIQUEE), JPEG signal-to-noise ratio (JPEGSNR), discrete cosine transform (DCT), natural blind image quality assess (NBIQA), pixel-based sharpness score (PSS), spatial structural entropy quality assessor (SSEQ), and Illumination and naturalness invariant image quality assessor (ILNIQE). Observations on the table are captured as follows after applying the various metrics.
SRCC ranges from -0.70 (NIQE) to 0.64 (FRIQUEE), indicating varying degrees of alignment with human judgments in grading image quality, since a higher SRCC indicates better agreement with human perception. Similarly, PLCC ranges from -0.74 (NIQE) to 0.72 (FRIQUEE), suggesting varying levels of linear relationship with human judgments; Metrics with higher

Metric	BIQI	BLIINDS	IQVG	NIQE	DIVINE	BRISQUE	PIQE	BMPRI	FRIQUEE	JPEGSNR	DCT	NBIQA	PSS	SSEQ	ILNIQE
SRCC	-0.66	-0.52	0.43	-0.70	-0.62	-0.61	-0.55	-0.59	0.64	0.47	-0.27	-0.48	-0.40	-0.39	-0.52
PLCC	-0.69	-0.54	0.47	-0.74	-0.63	-0.67	-0.65	-0.69	0.72	0.57	-0.23	-0.56	-0.53	-0.40	-0.57
RMSE	1.64	1.67	1.39	1.69	1.50	1.58	1.59	1.73	1.30	1.85	1.53	1.61	1.36	1.26	1.63

Table 4.5: State-of-the-art No reference metrics performance on the proposed dataset.

PLCC demonstrate stronger alignment. RMSE ranges from 1.26 (SSEQ) to 1.85 (JPEGSNR), indicating varying prediction errors. As highlighted above, lower RMSE values indicate better predictive accuracy.

Metrics such as FRIQUEE demonstrate relatively higher correlation coefficients (SRCC: 0.64, PLCC: 0.72), indicating better alignment and stronger linear relationship with human judgments. On the contrary, metrics such as NIQE and DIVINE exhibit lower correlation coefficients and higher prediction errors, suggesting poorer performance in predicting human judgments. Thus, the choice of IQA metric depends on the specific application requirements, considering factors such as correlation, linearity, and prediction accuracy.

Table 4.6 presents the performance of various full-reference IQA metrics for different types of distortions used in the dataset. The metrics evaluated include UQI, PSNR, SSIM, MS-SSIM, IFC, VIF, FI, NQM, IW_PSNR, IW_SSIM, IW_MSE, and FSIM.

Different types of distortion exhibit varying levels of correlation coefficients (SRCC and PLCC) and prediction errors (RMSE) across IQA metrics; for example, distortions like Motion Blur and Gaussian Blur generally have higher correlation coefficients and lower prediction errors compared to distortions like Salt and Pepper, and Poisson Noise. Performance metrics are often classified into high, average, and low severity levels for each type of distortion. In many cases, the severity of the distortion directly affects the performance of the IQA metrics. For example, higher levels of distortion severity tend to result in lower correlation coefficients and higher prediction errors. Metrics such as VIF and NQM consistently demonstrate higher correlation coefficients and lower prediction errors in multiple types of distortion. These metrics are particularly effective at assessing image quality across a wide range of types of distortion and severity levels. Some distortions may exhibit inconsistencies in performance across different IQA metrics. For example, while one metric may perform well for a particular distortion type, another metric may show poor performance. The data highlight the importance of comprehensive evaluation and selection of IQA metrics based on the specific characteristics of the distortion and the requirements of the application. Different distortions may require different sets of metrics for an accurate assessment, and a one-size-fits-all approach may not be suitable. Distortions like Gaussian Noise and Poisson Noise pose unique challenges due to their

Distortion	Metric	UQI	PSNR	SSIM	MS-SSIM	IFC	VIF	FI	NQM	IW PSNR	IW SSIM	IW MSE	FSIM
Pincushion	SRCC	-0.08	0.20	0.00	0.22	0.31	0.14	0.20	0.12	-0.21	0.06	0.21	0.00
	PLCC	-0.01	0.13	0.04	0.26	0.37	0.23	0.19	0.16	-0.22	0.08	0.19	0.02
	RMSE	1.62	1.19	1.02	0.73	1.83	1.81	0.91	1.40	1.56	1.06	0.87	0.85
Barrel	SRCC	-0.29	0.46	0.03	0.31	-0.08	-0.16	0.05	0.12	-0.09	-0.12	0.09	0.14
	PLCC	-0.08	0.52	-0.06	0.47	-0.05	-0.11	0.02	0.13	0.02	-0.10	-0.02	0.13
	RMSE	1.53	1.24	1.17	0.69	1.59	1.58	1.08	1.44	1.39	1.51	1.47	1.13
Motion Blur	SRCC	0.82	0.67	0.66	0.49	0.89	0.92	0.63	0.53	0.88	0.52	-0.88	0.90
	PLCC	0.79	0.69	0.62	0.44	0.82	0.91	0.62	0.50	0.86	0.51	-0.63	0.71
	RMSE	1.73	1.56	2.35	1.06	0.72	1.13	2.10	1.88	1.50	2.82	1.20	2.82
Gaussian Blur	SRCC	0.78	0.52	0.58	0.66	0.80	0.79	0.60	0.50	0.68	0.72	-0.68	0.79
	PLCC	0.78	0.49	0.49	0.59	0.90	0.89	0.53	0.45	0.79	0.67	-0.51	0.61
	RMSE	1.34	1.74	2.28	1.57	0.44	0.63	1.96	2.05	1.53	2.91	0.60	3.07
Gaussian Noise	SRCC	0.70	0.62	0.77	-0.03	0.76	0.70	0.74	0.33	0.60	0.40	-0.60	0.75
	PLCC	0.68	0.57	0.73	0.00	0.70	0.70	0.69	0.34	0.54	0.51	-0.55	0.73
	RMSE	0.66	0.87	0.82	0.66	0.54	0.70	0.55	0.79	0.67	2.16	1.86	2.10
Salt and Pepper	SRCC	0.66	0.61	0.69	0.18	0.76	0.67	0.66	0.59	0.64	0.31	-0.64	0.67
	PLCC	0.60	0.61	0.67	0.18	0.62	0.66	0.66	0.59	0.62	0.36	-0.61	0.64
	RMSE	0.85	1.01	1.11	0.49	0.55	1.00	0.84	0.93	0.82	2.09	1.98	2.03
Poisson Noise	SRCC	-0.26	0.58	0.05	0.39	-0.43	0.48	0.04	0.24	0.78	0.43	-0.78	0.13
	PLCC	-0.36	0.61	0.31	0.33	-0.32	0.38	0.30	0.29	0.75	0.45	-0.79	0.23
	RMSE	0.91	0.81	1.15	0.69	0.81	1.08	0.95	0.74	0.71	1.46	2.61	1.44

Table 4.6: State-of-the-art Full reference metrics performance for each type of distortions used in the dataset.

stochastic nature. Metrics may perform differently for these distortions compared to deterministic distortions such as Blur, and Salt and Pepper.

Table 4.7 presents the performance of various no-reference IQA metrics for each type of distortion used in the dataset. Distortion types include Pincushion, Barrel, Motion Blur, Gaussian Blur, Gaussian Noise, Salt and Pepper, and Poisson Noise. The table evaluates IQA metrics such as BIQI, BLIINDS, IQVG, NIQE, and others as seen in the table.

SRCC ranges from negative to positive values, indicating the strength and direction of association between predicted IQA scores and human-assigned scores for each distortion type. Negative SRCC values suggest a reverse correlation, while positive values indicate a positive correlation.

PLCC varies between negative and positive values, representing the linear relationship between predicted and human-assigned scores for each type of distortion.

RMSE represents the average magnitude of the prediction errors for each distortion type and IQA metric combination.

Each distortion type exhibits unique characteristics, leading to varying performance of IQA metrics across different distortions. Motion Blur distortion poses significant challenges for many IQA metrics, as indicated by low SRCC and PLCC values and high RMSE values. This suggests that accurately assessing image quality under motion blur remains a complex task for no-reference IQA metrics. On the other hand, distortions like Gaussian Noise and Salt and Pepper show relatively higher correlation coefficients (SRCC and PLCC) for certain IQA

Distortion	Metric	BIQI	BLIINDS	IQVG	NIQE	DIVINE	BRISQUE	PIQE	BMPRI	FRIQUEE	JPEGSNR	DCT	NBIQA	PSS	SSEQ	ILNIQE
Pincushion	SRCC	-0.02	-0.01	-0.04	0.21	0.35	0.12	0.31	0.06	0.12	-0.06	-0.08	0.29	-0.35	0.23	-0.15
	PLCC	0.04	0.00	0.02	0.18	0.33	0.13	0.35	0.09	0.10	0.01	-0.02	0.28	-0.36	0.32	-0.23
	RMSE	1.39	0.77	1.43	1.46	0.75	0.72	0.82	1.24	1.53	1.61	1.09	0.56	1.12	0.93	1.63
Barrel	SRCC	-0.06	-0.21	-0.30	-0.09	-0.11	-0.26	0.28	0.11	0.23	0.12	0.15	-0.24	0.06	0.00	0.00
	PLCC	-0.10	-0.20	-0.47	-0.20	-0.25	-0.33	0.24	0.10	0.21	0.15	0.23	0.08	-0.13	0.09	-0.05
	RMSE	1.13	0.94	2.03	1.26	0.82	0.92	0.68	1.00	1.79	1.89	0.94	0.64	0.89	0.92	1.39
Motion Blur	SRCC	-0.56	-0.29	-0.11	-0.82	-0.65	-0.62	-0.63	-0.44	0.78	0.06	-0.31	-0.71	-0.48	-0.14	-0.75
	PLCC	-0.57	-0.33	-0.20	-0.81	-0.63	-0.63	-0.66	-0.50	0.79	0.09	-0.26	-0.72	-0.47	-0.20	-0.73
	RMSE	0.73	1.68	1.30	0.90	1.57	1.47	1.34	0.91	0.73	2.39	1.85	1.84	1.01	0.81	0.98
Gaussian Blur	SRCC	-0.76	-0.37	0.52	-0.76	-0.57	-0.66	-0.73	-0.63	0.47	0.52	-0.50	-0.83	-0.81	0.01	-0.82
	PLCC	-0.78	-0.67	0.70	-0.85	-0.60	-0.84	-0.82	-0.72	0.76	0.54	-0.47	-0.89	-0.80	-0.09	-0.68
	RMSE	2.42	2.92	0.73	2.46	2.69	3.04	2.75	2.69	0.61	1.70	2.67	2.93	2.52	1.75	1.53
Gaussian Noise	SRCC	-0.70	-0.15	0.31	-0.81	-0.73	-0.47	-0.37	-0.31	0.43	-0.25	0.26	0.37	-0.19	-0.14	-0.68
	PLCC	-0.69	-0.11	0.04	-0.78	-0.71	-0.47	-0.39	-0.40	0.41	-0.30	0.17	0.40	-0.12	-0.17	-0.72
	RMSE	1.35	1.45	1.37	1.39	1.07	0.96	1.42	1.55	1.52	1.94	1.08	1.35	0.60	1.18	1.54
Salt and Pepper	SRCC	-0.42	0.46	0.30	-0.74	-0.38	0.13	-0.03	0.00	0.19	-0.39	0.21	0.50	-0.09	0.05	-0.47
	PLCC	-0.47	0.41	0.03	-0.70	-0.43	0.16	-0.13	-0.12	0.13	-0.41	0.20	0.49	-0.04	0.02	-0.51
	RMSE	1.54	1.48	1.47	1.57	1.04	1.04	1.48	1.72	1.46	1.91	1.08	1.19	0.68	1.19	1.70
Poisson Noise	SRCC	0.02	0.77	0.02	0.28	-0.12	0.05	0.54	0.13	-0.37	0.19	0.52	0.61	-0.42	0.67	0.31
	PLCC	0.07	0.67	-0.26	0.12	-0.15	-0.04	0.55	0.18	-0.40	0.17	0.35	0.57	-0.32	0.65	0.45
	RMSE	2.27	1.55	1.00	2.25	1.46	1.49	1.76	2.28	0.92	1.17	1.14	1.21	1.35	1.59	2.35

Table 4.7: State-of-the-art No reference metrics performance for each type of distortions used in the dataset.

metrics, indicating better alignment with human judgments under these types of noise. Some IQA metrics, such as BRISQUE and PSS, demonstrate relatively consistent performance across multiple distortion types, with moderate to high correlation coefficients and lower prediction errors. This suggests that these metrics may be more robust and reliable in a range of distortion scenarios. Other metrics, such as BLIINDS and ILNIQE, exhibit more variability in performance in different types of distortion, with fluctuating correlation coefficients and RMSE values. This indicates that the effectiveness of these metrics may be more dependent on the specific characteristics of the distortion being evaluated. Evaluating IQA metrics across various types of distortion highlights the complexity of image quality assessment in real-world scenarios.

Table 4.8 presents the performance of various full reference IQA metrics categorized by three levels of distortions. High, Average, and Low. The metrics evaluated include UQI, PSNR, SSIM, MS-SSIM, IFC, VIF, FI, NQM, IW_PSNR, IW_SSIM, IW_MSE, and FSIM.
The performance of full-reference IQA metrics varies between different levels of distortion, indicating the sensitivity of these metrics to the severity of image distortions. Metrics such as FI and VIF demonstrate relatively higher correlation coefficients across all distortion levels, suggesting better alignment and stronger linear relationships with human judgments. Metrics such as MS-SSIM and IW PSNR exhibit higher prediction errors across all distortion levels, indicating poorer performance in predicting human judgments. The choice of the IQA metric depends on the specific application requirements and the severity of image distortions, considering factors such as correlation, linearity, and prediction accuracy.

Distortion Level	Metric	UQI	PSNR	SSIM	MS-SSIM	IFC	VIF	FI	NQM	IW PSNR	IW SSIM	IW MSE	FSIM
High	SRCC	0.07	0.09	-0.17	-0.22	0.23	0.22	-0.17	-0.16	0.10	0.10	-0.10	0.16
	PLCC	0.10	-0.06	-0.18	-0.22	0.37	0.34	-0.16	-0.23	0.04	-0.06	0.19	-0.11
	RMSE	1.19	1.27	1.51	1.01	0.88	1.00	1.29	1.42	1.12	2.23	1.27	2.23
Average	SRCC	0.02	0.15	-0.15	-0.07	0.15	0.21	-0.16	-0.13	0.06	0.09	-0.06	0.08
	PLCC	0.02	0.01	-0.13	-0.03	0.22	0.26	-0.11	-0.18	0.00	-0.03	0.16	-0.11
	RMSE	1.36	1.31	1.60	0.99	1.10	1.20	1.38	1.50	1.29	2.19	1.41	2.20
Low	SRCC	-0.02	0.13	-0.12	-0.02	0.11	0.16	-0.13	-0.11	0.03	-0.23	-0.03	0.04
	PLCC	-0.09	-0.06	-0.19	-0.06	0.11	0.12	-0.16	-0.24	-0.12	-0.27	0.26	-0.24
	RMSE	1.49	1.34	1.64	0.80	1.29	1.44	1.44	1.54	1.46	2.18	1.56	2.05

Table 4.8: State-of-the-art Full reference metrics performance for each of the three levels of distortions used in the dataset. Here, each level (high, average and low) is averaged over all types of distortion.

Distortion Level	Metric	BIQI	BLIINDS	IQVG	NIQE	DIVINE	BRISQUE	PIQE	BMPRI	FRIQUEE	JPEGSNR	DCT	NBIQA	PSS	SSEQ	ILNIQE
High	SRCC	-0.58	-0.69	0.48	-0.74	-0.69	-0.69	-0.67	-0.64	0.77	0.44	-0.48	-0.66	-0.46	-0.52	-0.68
	PLCC	-0.63	-0.70	0.50	-0.76	-0.69	-0.73	-0.73	-0.68	0.81	0.55	-0.44	-0.72	-0.58	-0.46	-0.67
	RMSE	1.52	1.78	1.53	1.57	1.59	1.69	1.62	1.60	1.54	2.13	1.76	1.85	1.50	1.16	1.36
Average	SRCC	-0.59	-0.65	0.42	-0.67	-0.62	-0.63	-0.62	-0.59	0.71	0.46	-0.33	-0.63	-0.37	-0.46	-0.48
	PLCC	-0.68	-0.65	0.49	-0.76	-0.65	-0.72	-0.72	-0.70	0.77	0.60	-0.30	-0.67	-0.51	-0.44	-0.61
	RMSE	1.47	1.71	1.48	1.56	1.48	1.58	1.61	1.63	1.37	1.95	1.60	1.68	1.30	1.18	1.41
Low	SRCC	-0.68	-0.53	0.33	-0.72	-0.61	-0.58	-0.51	-0.47	0.70	0.38	-0.30	-0.49	-0.30	-0.43	-0.44
	PLCC	-0.80	-0.60	0.44	-0.84	-0.69	-0.71	-0.67	-0.74	0.80	0.61	-0.29	-0.62	-0.42	-0.47	-0.55
	RMSE	1.57	1.57	1.30	1.62	1.39	1.44	1.47	1.68	1.09	1.68	1.37	1.32	1.09	1.19	1.67

Table 4.9: State-of-the-art No reference metrics performance for each of the three levels of distortions used in the dataset.

The data provided in the table 4.9 presents a comprehensive evaluation of various types of image distortion in different levels and metrics. Each distortion type is assessed at three distinct levels: High, Average, and Low. This multilevel evaluation allows for a nuanced understanding of how the severity of distortion influences image quality metrics. It should be noted that certain distortions may exhibit varying effects on image quality depending on their intensity levels. For instance, Motion Blur may have a more pronounced impact on image quality at higher levels compared to lower levels.

A wide array of image quality metrics are employed to assess distortion effects. These metrics include established measures such as PSNR , and MS-SSIM , as well as more specialized metrics such as VIF and NQM. Using multiple metrics, the evaluation provides a comprehensive view of image quality from different perspectives, considering factors such as luminance fidelity, structural similarity, and perceptual relevance.

The SRCC and PLCC are computed for each metric and distortion level. These correlation coefficients offer insights into the relationship between metric scores and human perception of image quality. Positive correlations indicate that higher metric scores align with better perceived image quality, while negative correlations suggest the opposite. Analyzing these correlations helps to understand the effectiveness of each metric in capturing perceptual image quality across different distortion scenarios.

Tables 4.10 and 4.11 provided present the performance of the state-of-the-art full reference and no-reference image quality metrics for different types and levels of distortion, respectively. The analysis covers various types of image distortion, including Pincushion, Barrel, Motion Blur, Gaussian Blur, Gaussian Noise, Salt and Pepper, and Poisson Noise; for each distortion type, the metrics are evaluated at three levels: High, Average, and Low. These levels represent different degrees of severity or intensities of distortion, allowing a comprehensive assessment of the performance of the metric in a variety of scenarios. The evaluation includes several performance metrics, such as SRCC, PLCC, and RMSE. SRCC and PLCC indicate the correlation between metric scores and human perception of image quality, with higher values suggesting a better alignment with human judgment. RMSE provides information on the precision of metric predictions compared to ground-truth quality scores.

These distortions exhibit mixed performance across different metrics and levels. For example, in the case of pincushion distortion at high levels, most metrics show negative correlations (SRCC and PLCC), indicating a weak association between metric scores and perceived quality. However, at low levels, some metrics achieve relatively higher correlations, suggesting better performance in quantifying quality degradation. Metrics generally perform poorly in capturing quality degradation caused by motion blur, especially at high levels where correlations are consistently negative. This indicates a significant challenge for existing no-reference metrics in accurately assessing image quality in the presence of motion blur. These distortions pose challenges to metrics, especially at higher levels, where correlations tend to be negative. However, some metrics show better performance at lower levels, with higher correlations and lower RMSE values. The performance of the metrics varies between distortion levels, with some metrics demonstrating better correlations at low or average levels compared to high levels. Poisson noise, in particular, seems to be challenging for metrics, as indicated by relatively low correlations and higher RMSE values. This analysis highlights the strengths and limitations of existing full reference and no-reference image quality metrics in assessing different types and levels of distortion.

Distortion	Level	Metric	UQI	PSNR	SSIM	MS-SSIM	IFC	VIF	FI	NQM	IW_PSNR	IW_SSIM	IW_MSE	FSIM
Pincushion	High	SRCC	-0.62	0.09	-0.60	-0.19	-0.56	-0.70	-0.47	-0.58	-0.32	-0.34	0.32	-0.49
		PLCC	-0.62	-0.12	-0.51	-0.22	-0.53	-0.65	-0.37	-0.48	-0.43	-0.35	0.33	-0.55
		RMSE	1.43	1.03	1.00	0.69	1.56	1.55	0.88	1.23	1.31	1.08	0.97	0.88
Pincushion	Average	SRCC	-0.43	0.19	-0.13	0.21	-0.19	-0.31	0.09	-0.35	-0.37	-0.24	0.37	-0.21
		PLCC	-0.49	0.00	-0.18	0.23	-0.25	-0.42	0.01	-0.40	-0.31	-0.24	0.25	-0.32
		RMSE	1.71	1.27	1.12	0.88	1.89	1.87	1.00	1.48	1.63	1.01	0.89	0.96
Pincushion	Low	SRCC	-0.23	0.10	0.04	0.50	0.48	0.12	0.13	-0.23	-0.31	-0.17	0.31	-0.10
		PLCC	0.07	-0.10	0.20	0.51	0.23	0.09	0.34	-0.24	-0.26	-0.13	0.31	-0.11
		RMSE	1.76	1.29	1.04	0.60	2.08	2.04	0.95	1.52	1.78	1.15	0.83	0.72
Barrel	High	SRCC	-0.53	0.30	-0.06	0.54	0.02	-0.02	0.05	-0.09	-0.29	-0.29	0.29	0.12
		PLCC	-0.22	0.44	0.13	0.57	0.19	0.19	0.24	-0.28	-0.02	-0.07	0.01	0.16
		RMSE	1.19	0.95	0.84	0.37	1.22	1.21	0.75	1.12	1.03	1.18	1.65	0.85
Barrel	Average	SRCC	-0.38	0.33	-0.03	0.51	0.07	-0.01	0.12	-0.14	-0.19	-0.17	0.19	0.09
		PLCC	-0.23	0.46	-0.06	0.48	0.11	0.06	0.03	-0.22	-0.03	-0.10	0.02	0.02
		RMSE	1.40	1.13	1.06	0.63	1.44	1.42	0.97	1.32	1.25	1.40	1.58	1.05
Barrel	Low	SRCC	-0.77	0.31	-0.60	0.52	-0.60	-0.67	-0.45	-0.37	-0.19	-0.57	0.19	-0.30
		PLCC	-0.37	0.39	-0.59	0.65	-0.56	-0.53	-0.49	-0.27	-0.10	-0.57	0.07	-0.13
		RMSE	1.92	1.55	1.50	0.94	2.01	2.00	1.42	1.79	1.78	1.86	1.14	1.42
Motion Blur	High	SRCC	0.72	0.22	0.23	0.33	0.79	0.66	0.26	0.16	0.61	0.57	-0.61	0.61
		PLCC	0.78	0.25	0.21	0.36	0.95	0.85	0.26	-0.01	0.71	0.40	-0.44	0.51
		RMSE	1.73	1.72	2.49	1.41	0.48	0.87	2.21	2.02	1.41	3.06	0.84	3.09
Motion Blur	Average	SRCC	0.52	0.68	0.44	0.47	0.66	0.90	0.32	0.25	0.84	0.52	-0.84	0.79
		PLCC	0.72	0.74	0.64	0.50	0.80	0.95	0.61	0.15	0.91	0.57	-0.76	0.81
		RMSE	1.83	1.59	2.43	1.02	0.79	1.21	2.17	1.95	1.58	2.88	1.07	2.89
Motion Blur	Low	SRCC	0.14	0.72	0.30	0.55	0.21	0.70	0.14	0.35	0.88	0.57	-0.88	0.85
		PLCC	0.45	0.77	0.68	0.57	0.47	0.88	0.55	0.22	0.86	0.60	-0.90	0.92
		RMSE	1.62	1.34	2.11	0.56	0.83	1.26	1.91	1.67	1.49	2.48	1.58	2.44
Gaussian Blur	High	SRCC	0.77	0.15	0.39	0.23	0.81	0.65	0.41	0.42	0.18	0.62	-0.18	0.67
		PLCC	0.72	0.02	0.12	0.25	0.74	0.72	0.24	0.20	0.45	0.63	-0.40	0.53
		RMSE	0.95	1.66	2.04	1.66	0.25	0.28	1.74	1.88	1.23	2.66	0.39	2.83
Gaussian Blur	Average	SRCC	0.76	0.53	0.50	0.33	0.95	0.92	0.49	0.53	0.85	0.70	-0.85	0.88
		PLCC	0.84	0.56	0.47	0.31	0.91	0.97	0.49	0.42	0.91	0.82	-0.73	0.85
		RMSE	1.31	1.79	2.30	1.66	0.41	0.52	1.98	2.10	1.51	2.85	0.44	3.17
Gaussian Blur	Low	SRCC	0.56	0.32	0.24	0.38	0.60	0.58	0.35	0.27	0.55	0.52	-0.55	0.81
		PLCC	0.79	0.42	0.47	0.42	0.92	0.92	0.49	0.20	0.81	0.44	-0.64	0.74
		RMSE	1.62	1.75	2.44	1.36	0.59	0.89	2.12	2.12	1.76	3.20	0.86	3.14
Gaussian Noise	High	SRCC	0.61	0.36	0.43	0.38	0.67	0.26	0.30	-0.07	0.39	0.52	-0.39	0.43
		PLCC	0.65	0.25	0.53	0.37	0.67	0.46	0.50	-0.09	0.15	0.64	-0.22	0.59
		RMSE	0.71	0.97	0.83	0.70	0.48	0.75	0.61	0.88	0.79	2.33	1.55	2.35
Gaussian Noise	Average	SRCC	0.73	0.63	0.71	-0.06	0.80	0.67	0.66	-0.12	0.52	0.52	-0.52	0.67
		PLCC	0.71	0.57	0.78	-0.06	0.73	0.72	0.73	-0.02	0.53	0.53	-0.57	0.75
		RMSE	0.54	0.74	0.65	0.82	0.60	0.54	0.40	0.69	0.54	2.15	1.93	2.00
Gaussian Noise	Low	SRCC	0.50	0.44	0.56	0.22	0.59	0.59	0.52	-0.04	0.55	0.55	-0.55	0.59
		PLCC	0.55	0.57	0.77	0.25	0.61	0.74	0.69	-0.11	0.56	0.58	-0.62	0.76
		RMSE	0.70	0.88	0.96	0.37	0.54	0.79	0.62	0.79	0.65	2.00	2.06	1.92
Salt and Pepper	High	SRCC	0.66	0.59	0.66	0.06	0.83	0.59	0.61	-0.03	0.60	0.52	-0.60	0.54
		PLCC	0.77	0.53	0.72	0.14	0.80	0.68	0.67	0.05	0.45	0.55	-0.51	0.70
		RMSE	0.78	0.99	0.95	0.53	0.39	0.86	0.66	0.92	0.78	2.30	1.65	2.26
Salt and Pepper	Average	SRCC	0.44	-0.04	0.09	0.18	0.37	-0.09	-0.03	-0.02	0.20	0.58	-0.20	0.03
		PLCC	0.25	0.06	0.18	0.18	0.26	0.12	0.14	0.25	0.07	0.58	-0.16	0.13
		RMSE	0.93	1.06	1.16	0.44	0.66	1.05	0.89	0.95	0.86	2.18	1.89	2.09
Salt and Pepper	Low	SRCC	0.36	0.45	0.52	0.25	0.44	0.60	0.39	0.19	0.71	0.57	-0.71	0.52
		PLCC	0.35	0.35	0.57	0.27	0.44	0.57	0.57	0.15	0.47	0.58	-0.58	0.52
		RMSE	0.84	0.97	1.19	0.49	0.57	1.08	0.93	0.90	0.80	1.73	2.34	1.69
Poisson Noise	-	SRCC	-0.26	0.58	0.05	0.30	-0.43	0.48	0.04	0.24	0.78	0.65	-0.78	0.13
		PLCC	-0.36	0.61	0.31	0.35	-0.32	0.38	0.30	0.29	0.75	0.69	-0.79	0.23
		RMSE	0.91	0.81	1.15	0.69	0.81	1.08	0.95	0.74	0.71	1.46	2.61	1.44

Table 4.10: State-of-the-art Full reference metrics performance for each of the three levels specific to the distortion types.

Distortion Level	Metric	BIQI	BLIINDS	IQVG	NIQE	DIVINE	BRISQUE	PIQE	BMPRI	FRIQUEE	JPEGSNR	DCT	NBIQA	PSS	SSEQ	ILNIQE
Pincushion High	SRCC	-0.25	-0.37	-0.03	-0.22	-0.04	-0.14	0.08	-0.03	0.53	-0.07	-0.27	0.02	-0.32	-0.13	-0.27
	PLCC	-0.21	-0.30	0.15	-0.22	0.03	-0.21	0.09	-0.05	0.62	-0.03	-0.20	0.12	-0.33	0.04	-0.28
	RMSE	1.11	0.76	1.55	1.25	0.76	0.80	0.64	0.96	1.83	1.82	1.20	0.49	0.82	0.73	1.28
Pincushion Average	SRCC	-0.20	-0.18	0.03	0.13	0.30	-0.20	0.20	0.03	0.24	0.15	0.13	0.16	-0.25	0.19	-0.09
	PLCC	-0.13	-0.19	-0.06	0.10	0.24	-0.17	0.28	0.14	0.20	0.12	0.09	0.18	-0.30	0.34	-0.17
	RMSE	1.46	0.90	1.42	1.46	0.77	0.73	0.90	1.25	1.49	1.56	1.09	0.66	1.12	0.96	1.68
Pincushion Low	SRCC	0.24	0.34	-0.63	0.35	0.34	0.33	0.52	0.62	-0.09	-0.20	-0.27	0.44	-0.40	0.61	-0.24
	PLCC	0.29	0.46	-0.61	0.21	0.17	0.33	0.52	0.65	0.04	-0.26	-0.25	0.43	-0.53	0.59	-0.23
	RMSE	1.65	0.66	1.34	1.72	0.96	0.79	0.95	1.55	1.14	1.35	0.96	0.59	1.40	1.14	1.94
Barrel High	SRCC	0.13	-0.09	-0.35	0.22	0.08	0.10	0.53	0.45	0.31	-0.16	0.37	0.55	-0.54	0.19	0.29
	PLCC	0.16	0.01	-0.35	0.19	-0.08	-0.10	0.48	0.41	0.08	-0.09	0.44	0.36	-0.32	0.26	0.25
	RMSE	0.70	0.79	2.20	0.80	0.62	0.65	0.32	0.55	1.99	2.10	0.99	0.56	0.59	0.55	1.01
Barrel Average	SRCC	0.21	0.03	-0.49	0.28	0.07	0.02	0.59	0.54	0.08	-0.26	0.20	0.61	-0.49	0.19	0.25
	PLCC	0.16	-0.01	-0.57	0.08	-0.11	-0.26	0.53	0.48	0.02	-0.05	0.33	0.45	-0.32	0.16	0.11
	RMSE	0.94	0.89	2.17	1.07	0.74	0.86	0.50	0.77	1.94	2.02	0.96	0.55	0.78	0.79	1.23
Barrel Low	SRCC	-0.37	-0.57	-0.22	-0.49	-0.38	-0.54	0.05	0.08	0.48	0.14	-0.04	-0.27	-0.02	-0.48	-0.38
	PLCC	-0.41	-0.54	-0.57	-0.60	-0.53	-0.49	0.01	0.02	0.39	0.21	0.01	-0.19	-0.07	-0.24	-0.37
	RMSE	1.56	1.13	1.67	1.72	1.03	1.19	1.02	1.46	1.38	1.51	0.85	0.79	1.19	1.26	1.81
Motion Blur High	SRCC	-0.04	0.04	-0.64	-0.56	-0.49	-0.44	-0.10	-0.09	0.27	-0.13	-0.27	-0.45	-0.68	0.19	-0.89
	PLCC	-0.36	-0.39	-0.63	-0.88	-0.85	-0.79	-0.69	-0.42	0.80	0.15	-0.37	-0.88	-0.24	0.04	-0.88
	RMSE	0.81	2.20	1.94	1.17	2.01	1.94	1.85	0.75	2.74	2.29	2.47	1.43	1.13	0.75	2.17
Motion Blur Average	SRCC	-0.38	0.02	-0.66	-0.59	-0.65	-0.31	-0.30	-0.20	0.43	-0.08	-0.28	-0.20	-0.14	-0.15	-0.52
	PLCC	-0.62	-0.10	-0.81	-0.73	-0.72	-0.51	-0.59	-0.33	0.72	-0.08	-0.26	-0.59	-0.19	-0.12	-0.57
	RMSE	0.50	1.59	1.35	0.66	1.52	1.37	1.21	0.69	0.76	2.42	1.81	1.73	0.73	0.65	0.77
Motion Blur Low	SRCC	-0.60	-0.02	-0.36	-0.52	-0.26	-0.55	-0.56	-0.39	0.79	-0.01	-0.24	-0.03	-0.36	-0.14	0.24
	PLCC	-0.61	-0.05	-0.56	-0.60	-0.53	-0.59	-0.44	-0.27	0.78	-0.11	-0.14	-0.36	-0.35	-0.16	0.17
	RMSE	0.82	1.05	1.06	0.81	1.02	0.92	0.72	0.72	1.94	1.94	1.32	1.03	0.71	0.51	1.31
Gaussian Blur High	SRCC	-0.41	-0.64	0.04	-0.33	-0.47	-0.35	-0.18	0.20	-0.16	0.48	-0.70	-0.68	-0.71	0.54	-0.83
	PLCC	-0.48	-0.77	0.31	-0.54	-0.40	-0.44	-0.41	-0.05	0.30	0.46	-0.66	-0.69	-0.85	0.47	-0.83
	RMSE	3.03	3.19	0.76	2.97	2.84	3.35	3.10	3.09	0.58	1.46	2.90	3.29	3.13	1.97	2.17
Gaussian Blur Average	SRCC	-0.83	-0.21	0.49	-0.92	-0.67	-0.60	-0.88	-0.88	0.42	0.30	-0.38	-0.95	-0.93	-0.10	-0.89
	PLCC	-0.92	-0.58	0.77	-0.98	-0.56	-0.89	-0.94	-0.80	0.82	0.39	-0.41	-0.98	-0.95	-0.04	-0.85
	RMSE	2.47	3.00	0.74	2.62	2.82	3.17	2.95	2.84	0.54	1.72	2.77	3.08	2.67	1.91	1.38
Gaussian Blur Low	SRCC	-0.82	-0.12	0.44	-0.78	-0.37	-0.87	-0.84	-0.79	0.66	0.42	-0.32	-0.81	-0.87	-0.15	-0.56
	PLCC	-0.83	-0.59	0.83	-0.94	-0.65	-0.91	-0.85	-0.86	0.82	0.61	-0.38	-0.94	-0.80	-0.31	-0.69
	RMSE	1.61	2.54	0.70	1.66	2.39	2.54	2.11	2.12	0.71	1.88	2.30	2.37	1.45	1.34	0.76
Gaussian Noise High	SRCC	-0.55	0.04	0.09	-0.61	-0.33	-0.20	-0.01	0.03	0.53	-0.35	0.24	0.05	-0.35	-0.32	-0.60
	PLCC	-0.58	-0.19	-0.01	-0.70	-0.48	-0.24	-0.14	-0.29	0.65	-0.48	0.08	0.09	-0.16	-0.38	-0.71
	RMSE	1.04	1.08	1.36	1.03	1.07	0.79	0.89	1.16	1.80	2.25	1.10	1.23	0.68	1.06	1.20
Gaussian Noise Average	SRCC	-0.72	-0.19	0.28	-0.89	-0.66	-0.30	-0.33	-0.36	0.65	-0.27	0.09	0.25	-0.13	-0.18	-0.71
	PLCC	-0.66	0.04	-0.09	-0.84	-0.75	-0.36	-0.38	-0.36	0.42	-0.30	0.21	0.40	-0.20	-0.13	-0.70
	RMSE	1.32	1.50	1.31	1.43	1.00	0.89	1.43	1.61	1.43	1.85	1.06	1.51	0.55	1.23	1.60
Gaussian Noise Low	SRCC	-0.59	0.06	0.12	-0.71	-0.58	-0.28	-0.03	0.32	0.48	-0.37	0.16	0.39	-0.10	-0.08	-0.67
	PLCC	-0.60	0.36	-0.28	-0.68	-0.73	-0.39	-0.14	-0.15	0.32	-0.18	0.17	0.55	-0.15	-0.08	-0.56
	RMSE	1.62	1.71	1.43	1.65	1.12	1.16	1.78	1.82	1.29	1.68	1.08	1.30	0.54	1.24	1.77
Salt and Pepper High	SRCC	-0.54	0.10	0.31	-0.92	-0.71	0.20	-0.31	-0.31	0.48	-0.72	-0.03	0.19	0.70	-0.03	-0.71
	PLCC	-0.50	0.17	-0.14	-0.95	-0.73	0.09	-0.33	-0.30	0.48	-0.72	0.04	0.25	0.61	-0.20	-0.77
	RMSE	1.07	1.30	1.51	1.18	0.94	0.72	1.18	1.36	1.70	2.18	1.10	1.21	0.30	0.99	1.33
Salt and Pepper Average	SRCC	0.21	0.47	0.05	-0.60	0.24	0.08	-0.08	0.47	-0.03	-0.32	0.18	-0.25	-0.32	-0.07	-0.22
	PLCC	-0.11	0.28	0.07	-0.49	-0.02	-0.08	-0.10	0.54	-0.15	-0.27	0.18	-0.34	-0.32	-0.05	-0.35
	RMSE	1.37	1.04	1.48	1.30	1.01	0.82	1.17	1.30	1.84	2.11	1.15	1.15	0.60	0.90	1.36
Salt and Pepper Low	SRCC	0.36	0.54	-0.43	-0.50	0.32	0.32	-0.17	0.57	-0.46	0.38	0.42	-0.10	-0.10	-0.14	0.23
	PLCC	0.24	0.42	-0.38	-0.45	0.16	0.10	-0.06	0.49	-0.36	0.27	0.25	-0.03	-0.14	-0.11	0.14
	RMSE	1.24	0.91	1.61	1.46	0.99	0.73	1.11	1.19	1.76	1.90	1.01	1.04	0.60	0.84	1.34

Table 4.11: State-of-the-art No reference metrics performance for each of the three levels specific to the distortion types.

Chapter 5

Conclusions

In conclusion, this thesis advances the understanding of visual distortions in 360° images and their profound implications for QoE. Through a meticulous exploration of distortion impacts on perceived image quality, facilitated by a curated 360° Image Quality Dataset and comprehensive subjective evaluation, valuable insights have been gained. MOS and ANOVA analysis have quantitatively assessed the perceptual impact of distortions, highlighting the need for custom image processing strategies to improve QoE. Furthermore, the evaluation of existing image quality metrics has shed light on their performance in the context of 360° imagery, guiding the selection of appropriate metrics for specific applications. By synthesizing key findings and providing insight into the refinement of assessment methodologies for this emerging medium, it is with great belief that this thesis will contribute to the groundwork and state-of-the-art for future research efforts aimed at improving user experience and satisfaction with visual content across various platforms.

Bibliography

[1] Mohammad Abdullah-Al-Mamun, Vivek Tyagi, and Hong Zhao. A new full-reference image quality metric for motion blur profile characterization. *IEEE Access*, 9:156361–156371, 2021. `doi:10.1109/ACCESS.2021.3130177`.

[2] Atanas Boev, Danilo Hollosi, and Atanas Gotchev. Classification of stereoscopic artefacts. *Tech. Rep. 216503, MOBILE3DTV Project*, 01 2008.

[3] Kjell Brunnstrom and Marcus Barkowsky. Statistical quality of experience analysis: on planning the sample size and statistical significance testing. *Journal of Electronic Imaging*, 27(5):1–11, 2018. [Online]. URL: `https://doi.org/10.1117/1.JEI.27.5.053013`.

[4] Leslie Chandrakantha. Learning anova concepts using simulation. In *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*, pages 1–5, 2014. `doi:10.1109/ASEEZone1.2014.6820644`.

[5] Meixu Chen, Yize Jin, Todd Goodall, Xiangxu Yu, and Alan Conrad Bovik. Study of 3d virtual reality picture quality. *IEEE Journal of Selected Topics in Signal Processing*, 14(1):89–102, 2020. `doi:10.1109/JSTSP.2019.2956408`.

[6] Sijia Chen, Yingxue Zhang, Yiming Li, Zhenzhong Chen, and Zhou Wang. Spherical structural similarity index for objective omnidirectional video quality assessment. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2018. `doi:10.1109/ICME.2018.8486584`.

[7] Hsien-Tzu Cheng, Chun-Hung Chao, Jin-Dong Dong, Hao-Kai Wen, Tyng-Luh Liu, and Min Sun. Cube padding for weakly-supervised saliency prediction in 360° videos. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1420–1429, 2018. `doi:10.1109/CVPR.2018.00154`.

[8] Xavier Corbillon, Francesca De Simone, and Gwendal Simon. 360-degree video head movement dataset. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, MMSys'17, page 199–204, New York, NY, USA, 2017. Association for Computing Machinery. `doi:10.1145/3083187.3083215`.

[9] R. G. d. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli, and P. Frossard. Visual distortions in 360° videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(8):2524–2537, Aug 2020. `doi:10.1109/TCSVT.2019.2927344`.

[10] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, and A.C. Bovik. Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, 9(4):636–650, 2000. `doi:10.1109/83.841940`.

[11] Huiyu Duan, Guangtao Zhai, Xiongkuo Min, Yucheng Zhu, Yi Fang, and Xiaokang Yang. Perceptual quality assessment of omnidirectional images. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2018. `doi:10.1109/ISCAS.2018.8351786`.

[12] Huiyu Duan, Guangtao Zhai, Xiaokang Yang, Duo Li, and Wenhan Zhu. Ivqad 2017: An immersive video quality assessment database. In *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 1–5, 2017. `doi:10.1109/IWSSIP.2017.7965610`.

[13] X. Gao, T. Wang, and J. Li. A content-based image quality metric. In *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pages 231–240. Springer, 2005.

[14] D. Ghadiyaram and A. C. Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. *Journal of Vision*, 17(1):32–32, 2017.

[15] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE Transactions on Broadcasting*, 60(3):555–567, 2014. `doi:10.1109/TBC.2014.2344471`.

[16] Zhongyi Gu, Lin Zhang, and Hongyu Li. Learning a blind image quality index based on visual saliency guided sampling and gabor filtering. In *2013 IEEE International Conference on Image Processing*, pages 186–190, 2013. `doi:10.1109/ICIP.2013.6738039`.

[17] Jesús Gutiérrez, Erwan David, Yashas Rai, and Patrick Le Callet. Toolbox and dataset for the development of saliency and scanpath models for omnidirectional/360° still images. *Signal Processing: Image Communication*, 69:35–42, 2018. Salient360: Visual attention modeling for 360° Images. URL: https://www.sciencedirect.com/science/article/pii/S0923596518304594, doi:10.1016/j.image.2018.05.003.

[18] David Hasler and Sabine E. Suesstrunk. Measuring colorfulness in natural images. In Bernice E. Rogowitz and Thrasyvoulos N. Pappas, editors, *Human Vision and Electronic Imaging VIII*, volume 5007, pages 87 – 95. International Society for Optics and Photonics, SPIE, 2003. doi:10.1117/12.477378.

[19] Seunghoon Hong, Tackgeun You, Suha Kwak, and Bohyung Han. Online tracking by learning discriminative saliency map with convolutional neural network. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 597–606, Lille, France, 07–09 Jul 2015. PMLR. URL: https://proceedings.mlr.press/v37/hong15.html.

[20] Hou-Ning Hu, Yen-Chen Lin, Ming-Yu Liu, Hsien-Tzu Cheng, Yung-Ju Chang, and Min Sun. Deep 360 pilot: Learning a deep agent for piloting through 360° sports videos. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1396–1405, 2017. doi:10.1109/CVPR.2017.153.

[21] ITU-R. Methodology for the subjective assessment of the quality of television pictures. Technical report, International Telecommunication Union, 2002.

[22] ITU-T Study Group 12. J series: Cable networks and transmission of television, sound programme and other multimedia signals. Approval J.140-J.149: Measurement of the quality of service - Part 1, International Telecommunication Union, 2004. Approval date: 2004-03-15. Provisional name: J.fullref. Approval process: AAP. Status: In force. Maintenance responsibility: ITU-T Study Group 12. Further details: Patent statement(s). URL: https://handle.itu.int/11.1002/1000/7197.

[23] Baisheng Lai and Xiaojin Gong. Saliency guided dictionary learning for weakly-supervised image parsing. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3630–3639, 2016. doi:10.1109/CVPR.2016.395.

[24] Chen Li, Mai Xu, Xinzhe Du, and Zulin Wang. 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*, MM '18, page 932–940, New York, NY, USA, 2018. Association for Computing Machinery. `doi:10.1145/3240508.3240581`.

[25] Jia Li, Jinming Su, Changqun Xia, and Yonghong Tian. Distortion-adaptive salient object detection in 360° omnidirectional images. *IEEE Journal of Selected Topics in Signal Processing*, 14(1):38–48, 2020. `doi:10.1109/JSTSP.2019.2957982`.

[26] L. Liu, B. Liu, H. Huang, and A. C. Bovik. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, 29(8):856–863, 2014.

[27] Xiongkuo Min, Guangtao Zhai, Ke Gu, Yuming Fang, Xiaokang Yang, Xiaolin Wu, Jiantao Zhou, and Xianming Liu. Blind quality assessment of compressed images via pseudo structural similarity. In *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2016. `doi:10.1109/ICME.2016.7552955`.

[28] Xiongkuo Min, Guangtao Zhai, Ke Gu, Yutao Liu, and Xiaokang Yang. Blind image quality estimation via distortion aggravation. *IEEE Transactions on Broadcasting*, 64(2):508–517, 2018. `doi:10.1109/TBC.2018.2816783`.

[29] Xiongkuo Min, Guangtao Zhai, Ke Gu, Yucheng Zhu, Jiantao Zhou, Guodong Guo, Xiaokang Yang, Xinping Guan, and Wenjun Zhang. Quality evaluation of image dehazing methods using synthetic hazy images. *IEEE Transactions on Multimedia*, 21(9):2319–2333, 2019. `doi:10.1109/TMM.2019.2902097`.

[30] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. `doi:10.1109/TIP.2012.2214050`.

[31] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. `doi:10.1109/LSP.2012.2227726`.

[32] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, 2010. `doi:10.1109/LSP.2010.2043888`.

[33] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20(12):3350–3364, 2011. `doi:10.1109/TIP.2011.2147325`.

[34] Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, Sumohana S. Channappayya, and Swarup S. Medasani. Blind image quality evaluation using perception based features. In *2015 Twenty First National Conference on Communications (NCC)*, pages 1–6, 2015. `doi:10.1109/NCC.2015.7084843`.

[35] Fu-Zhao Ou, Yuan-Gen Wang, and Guopu Zhu. A novel blind image quality assessment method based on refined natural scene statistics. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1004–1008, 2019. `doi:10.1109/ICIP.2019.8803047`.

[36] Yashas Rai, Jesús Gutiérrez, and Patrick Le Callet. A dataset of head and eye movements for 360 degree images. New York, NY, USA, 2017. Association for Computing Machinery. `doi:10.1145/3083187.3083218`.

[37] Zhixiang Ren, Shenghua Gao, Liang-Tien Chia, and Ivor Wai-Hung Tsang. Region-based saliency detection and its application in object recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(5):769–779, 2014. `doi:10.1109/TCSVT.2013.2280096`.

[38] Michele A. Saad, Alan C. Bovik, and Christophe Charrier. A dct statistics-based blind image quality index. *IEEE Signal Processing Letters*, 17(6):583–586, 2010. `doi:10.1109/LSP.2010.2045550`.

[39] Abderrezzaq Sendjasni and Mohamed-Chaker Larabi. Pw-360iqa: Perceptually-weighted multichannel cnn for blind 360-degree image quality assessment. *Sensors*, 23(9), 2023. URL: `https://www.mdpi.com/1424-8220/23/9/4242`, `doi:10.3390/s23094242`.

[40] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006. `doi:10.1109/TIP.2005.859378`.

[41] H.R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005. `doi:10.1109/TIP.2005.859389`.

[42] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. Saliency in vr: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642, 2018. doi:10.1109/TVCG.2018.2793599.

[43] M. Sugeno. Fuzzy measures and fuzzy integrals—a survey. In *Readings in Fuzzy Sets for Intelligent Systems*, pages 251–257. Elsevier, 1993.

[44] Wei Sun, Xiongkuo Min, Guangtao Zhai, Ke Gu, Huiyu Duan, and Siwei Ma. Mc360iqa: A multi-channel cnn for blind 360-degree image quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 14(1):64–77, 2020. doi:10.1109/JSTSP.2019.2955024.

[45] Yule Sun, Ang Lu, and Lu Yu. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE Signal Processing Letters*, 24(9):1408–1412, 2017. doi:10.1109/LSP.2017.2720693.

[46] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2007. URL: http://dx.doi.org/10.1561/0600000009, doi:10.1561/0600000009.

[47] Evgeniy Upenik, Martin Řeřábek, and Touradj Ebrahimi. Testbed for subjective evaluation of omnidirectional visual content. In *2016 Picture Coding Symposium (PCS)*, pages 1–5, 2016. doi:10.1109/PCS.2016.7906378.

[48] V. Q. E. Group et al. Final report from the video quality experts group on the validation of objective models of video quality assessment. In *VQEG meeting*, Ottawa, Canada, March 2000.

[49] Zhou Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002. doi:10.1109/97.995823.

[50] Zhou 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*, 13(4):600–612, 2004. doi:10.1109/TIP.2003.819861.

[51] Zhou Wang and Qiang Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, 2011. doi:10.1109/TIP.2010.2092435.

[52] Zhou Wang, H.R. Sheikh, and A.C. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Proceedings. International Conference on Image Processing*, volume 1, pages I–I, 2002. `doi:10.1109/ICIP.2002.1038064`.

[53] Mai Xu, Chen Li, Zhenzhong Chen, Zulin Wang, and Zhenyu Guan. Assessing visual quality of omnidirectional videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(12):3516–3530, 2019. `doi:10.1109/TCSVT.2018.2886277`.

[54] Mai Xu, Chen Li, Yufan Liu, Xin Deng, and Jiaxin Lu. A subjective visual quality assessment method of panoramic videos. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, pages 517–522, 2017. `doi:10.1109/ICME.2017.8019351`.

[55] Mai Xu, Yuhang Song, Jianyi Wang, Minglang Qiao, Liangyu Huo, and Zulin Wang. Predicting head movement in panoramic video: A deep reinforcement learning approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(11):2693–2708, 2019. `doi:10.1109/TPAMI.2018.2858783`.

[56] Yanyu Xu, Yanbing Dong, Junru Wu, Zhengzhong Sun, Zhiru Shi, Jingyi Yu, and Shenghua Gao. Gaze prediction in dynamic 360° immersive videos. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5333–5342, 2018. `doi:10.1109/CVPR.2018.00559`.

[57] Honghai Yu and Stefan Winkler. Image complexity and spatial information. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 12–17, 2013. `doi:10.1109/QoMEX.2013.6603194`.

[58] Matt Yu, Haricharan Lakshman, and Bernd Girod. A framework to evaluate omnidirectional video coding schemes. In *2015 IEEE International Symposium on Mixed and Augmented Reality*, pages 31–36, 2015. `doi:10.1109/ISMAR.2015.12`.

[59] S. Yule, A. Lu, and Y. Lu. Ws-psnr for 360 video objective quality evaluation. *MPEG Joint Video Exploration Team*, 116, 2016.

[60] V. Zakharchenko, K. P. Choi, and J. H. Park. Quality metric for spherical panoramic video. In *Optics and Photonics for Information Processing X*, volume 9970, page 99700C. International Society for Optics and Photonics, 2016.

[61] Lin Zhang, Ying Shen, and Hongyu Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014. doi:10.1109/TIP.2014.2346028.

[62] Lin Zhang, Lei Zhang, and Alan C. Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015. doi:10.1109/TIP.2015.2426416.

[63] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011. doi:10.1109/TIP.2011.2109730.

[64] Yi Zhang, Lu Zhang, Wassim Hamidouche, and Olivier Déforges. A fixation-based 360° benchmark dataset for salient object detection. *2020 IEEE International Conference on Image Processing (ICIP)*, pages 3458–3462, 2020. URL: https://api.semanticscholar.org/CorpusID:210859107.

[65] Z. Zhang, Y. Xu, J. Yu, and S. Gao. Saliency detection in 360 videos. In *European Conference on Computer Vision*, pages 488–503. IEEE, 2018.

[66] Zicheng Zhang, Wei Sun, Xiongkuo Min, Tao Wang, Wei Lu, and Guangtao Zhai. A full-reference quality assessment metric for fine-grained compressed images. In *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–4, 2021. doi:10.1109/VCIP53242.2021.9675389.

[67] Jingbo Zhao, Robert S. Allison, Margarita Vinnikov, and Sion Jennings. Estimating the motion-to-photon latency in head mounted displays. In *2017 IEEE Virtual Reality (VR)*, pages 313–314, 2017. doi:10.1109/VR.2017.7892302.

[68] Wei Zhou, Jiahua Xu, Qiuping Jiang, and Zhibo 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*, 32(4):1778–1791, 2022. doi:10.1109/TCSVT.2021.3081182.