Master in Applied Computer Science (MACS)

HDR-FP: A Feature-Pooled Objective Metric for High Dynamic Range Video Quality Assessment

Master Thesis Report

Presented by

Lakshay Jain

and defended at the

Norwegian University of Science and Technology

June 2023

Academic Supervisor(s): Prof.Marius Pedersen

Submission of the thesis: 1st June 2023
Day of the oral defense: 12th June 2023

Abstract

High Dynamic Range (HDR) technology transforms video material by increasing the overall visual experience by delivering a greater range of luminance and color. However, because to the complexity of their content, measuring the quality of HDR videos is a difficult process. This master's thesis addresses the critical need for an objective quality assessment metric built exclusively for HDR video.

The major goal of this thesis is to propose and evaluate HDR-FP, a feature-pooled objective metric for judging HDR video quality. HDR-FP is based on considerable research and employs advanced image processing algorithms to extract important defects such as sharpness, compression faults, color distortion, banding, dynamic range, increased noise, blurring, and ghosting. These distortions are critical elements that have a considerable impact on the perceptual quality of HDR videos.

An extensive dataset of HDR films with a varied variety of material is used to evaluate the performance of HDR-FP. The suggested metric is extensively compared to two state-of-the-art metrics, HDR-VQM and HDR-VDP-2, and shows superior accuracy and consistency. The Spearman Rank Order Correlation Coefficient (SROCC), Pearson Linear Correlation Coefficient (PLCC), and Kendall Rank Order Correlation Coefficient (KROCC) are used to assess the correlations between the objective quality scores derived from HDR-FP and the subjective Mean Opinion Scores (MOS). The high correlation values found demonstrate HDR-FP's ability to align with human visual judgments.

This study makes a substantial contribution to the field of HDR video quality assessment by developing a strong and dependable objective metric for precisely quantifying the quality of HDR films. The availability of such a metric is critical for a variety of applications, including content development, broadcasting, streaming, and quality control, allowing content providers and service providers to provide viewers with immersive HDR experiences.

Dedication

I dedicate my dissertation work to my friends and family. I'm especially grateful to my loving parents for their endless support, unending love, and continued faith in my dreams. My sister is incredibly dear to me and has never left my side.

Your sacrifices, encouragement, and direction throughout my academic career are all acknowledged in this thesis. You have been my pillars of support, staying always at my side, and encouraging me to grow and flourish.

Your passion and sacrifice helped to mould me into the person I am today. Your advice, our late-night conversations, and your constant belief in my potential have strengthened my will to learn more and push the limits of my capability.

I will always be appreciative of all the sacrifices you have made, the several sleepless nights you have had, and the prayers you have uttered in hopes of my success. I now have the strength to face challenges and tenaciously follow my objectives thanks to your boundless love and support.

You are my inspiration, my biggest supporters, and my role models, and this thesis is dedicated to you. This achievement would not have been achieved without your support, love, and unshakable faith in me.

Acknowledgment

I thank God for all the chances, challenges, and fortitude that have made it possible for me to finish the thesis. I went through a lot during my thesis, both intellectually and personally.

I would like to extend my heartfelt appreciation to the individuals and organizations who were instrumental in the success of this research endeavour. Without their support and contributions, completion of this work would not have been feasible.

I would first like to express my profound gratitude to my supervisor, Prof. Marius Pedersen for his wise counsel, compassion, and understanding, but most significantly for helping me finish my thesis with encouragement and kindness. I am very fortunate and honoured by being under their kind supervision. I have learned a lot, exploring exciting projects with him.

I would like to express my sincere gratitude to everyone of my family members who helped me through my master's and thesis. Their confidence in me has sustained my enthusiasm and upbeat attitude throughout this process. Without their support and affection, I am not sure I could have completed this journey.

I want to sincerely thank all of my wonderful MACS classmates for never giving up on me and for always being there for me. I want to express my gratitude to Akib Jayed Islam in particular for his unflinching support during trying times.

Finally, I want to express my gratitude to all of the MACS coordinators for giving me the chance to develop and succeed.

Acronyms

VQR - Visual Quality Assessment

QoS - Quality of Service

QoE - Quality of Experience

SD - Standard Definition

HD - High Definition

UHD - Ultra High-Definition

HDR - High Dynamic Range

FPS - Frames per Second

TV - Television

HVS - Human Visual System

SDR - Standard Dynamic Range

EDR - Enhanced Dynamic Range

PQ - Perceptual Quantizer

HLG - Hybrid Log Gamma

OETF - Opto-Electronic Transfer Function

EOTF - Electro-Optical Transfer Function

OOTF - Optical-Optical Transfer Function

PSNR - Peak Signal-to-Noise Ratio

FR - Full-Reference

LDR - Low-Definition Range

MOS - Mean of Scores

QA - Quality Assessment

VQEG - Video Quality Experts Group

RR - Reduced-Reference

NR - No-Reference

HDR-VDP - HDR Video Quality PredictorHDR Video Quality Predictor

HDR-VQM - HDR Video Quality Metric

VQM - Video Quality Metric

PSNR-HDR - Peak Signal-to-Noise Ratio for HDR images

VMAF-HDR - Video Multi-Method Assessment Fusion for HDR

VMAF - Video Multi-method Assessment Fusion

SSIM - Structural Similarity Index

VIF - Visual Information Fidelity

HDR-SSIM - High Dynamic Range Structural Similarity Index

HDR-MS-SSIM - High Dynamic Range Multi-Scale Structural Similarity Index

MS-SSIM - Multi-Scale Structural Similarity Index

CNN - Convolutional Neural Network

SVR - Support Vector Regression

PLCC - Pearson's Linear Correlation Coefficient

SROCC - Spearman's Rank-Ordered Correlation Coefficient

KROCC - Kendall's rank-order correlation coefficient

HEVC - High Efficiency Video Coding

MSE - Mean Squared Error

DMOS - Difference Mean Opinion Scores

X

Contents

1 Introduction

Visual Quality Assessment is a fascinating challenge in the media environment. The evolution to higher resolutions and increased quality standards, such as high definition and better image quality, has led to the development of new models for measuring quality [Krasula et al. (2016)]. The focus has been on objective quality assessment, to create computational models that can predict perceptual image quality automatically. Even though quality assessment has been around for more than four decades, there has been little published on the subject. However, in recent years, there has been a significant increase in interest and progress in the field.

Before being viewed by a human observer, visual data may go through several stages of processing, each of which may introduce distortions that reduce the final display's quality. Visual Quality Assessment aims to measure the entire signal processing to achieve a reliable measure using various technical approaches.

Accurate assessment of video quality is essential for meeting promised quality of service (QoS) and improving the end user's quality of experience (QoE) [Karam et al. (2009)]. The use of digital images and videos as a means of media communication has increased dramatically in recent years. Due to technological advancements, the widespread availability of digital images and videos on the Internet has led to advances in quality assessment.

The increasing use of multimedia applications has coincided with an increase in the quality of experience that people expect from such applications. While significant progress is being made in providing new and improved multimedia services, the value of such services can only be assessed by the quality of experience that they provide to the end user, so determining the human opinion of quality, or getting as close to it as mathematical algorithms can, is critical in the design and deployment of a multimedia service.

Depending on the application of the multimedia service and the end-user of the audiovisual content, "quality" can be defined in a variety of ways. In applications

where the end-user is a human observer, the definition of "quality" must take into account signal perception by human sensory systems. Even within the realm of human users' applications, the interpretation of "quality" can vary depending on the multimedia service and the task defined for the human user. The vast majority of digital multimedia entertainment applications are concerned with defining "quality" as the overall QoE derived by the user from the service. This overall QoE is frequently determined by how good the image or video component of the multimedia signal looks or sounds, as well as the interactions between sensory perceptions [Karam et al. (2009)].

Images are subject to distortion during acquisition, compression, transmission, processing, and reproduction, and it is critical to be able to identify and quantify image quality degradations. Solving the problem requires matching image quality to human perception of quality [Winkler (2008)]. The development of effective automatic image quality assessment systems is a necessary goal for this purpose. The successful development of such objective quality assessment measures has enormous potential in a wide range of application environments.

The demand for digital video services has recently increased dramatically. Higher video resolution (SD, HD, UHD) and frame rate can be used to improve video quality as video technology advances (e.g., from 30 to 120 fps). High dynamic range (HDR) video, which provides a wider dynamic range of luminosity, is thought to be a significant advancement in TV technology [Banterle et al. (2017)]. In contrast to 4K and 8K technologies, which improve video quality in terms of resolution, HDR advances video technology in terms of luminance ratio. As a result, the dynamic range of video rendered by display can be increased, and a more realistic visual scene perceived by the human visual system (HVS) can be experienced.

Higher dynamic range can support more saturated colours than low dynamic range, in addition to increasing the perception of sharpness [Karam et al. (2009)].

The ratio between the maximum and minimum luminance perceived from a scene in a natural environment or rendered by a display is the concept of dynamic range. In digital cameras, the most commonly used unit for measuring dynamic range is the f-stop, which describes luminance by the power of two. Furthermore, luminance refers to the total amount of all visible light that passes through a specific space and is usually measured in candela per unit area, which is equivalent to nits [Ohta and Robertson (2006)]. In the real world, the luminance of the sun is approximately 6 108 nits, and the luminance of starlight at night is approximately 104 nits or lower [Savakis et al. (2000)]. A human face has a luminance of about

50 nits in a room, while a dark surface may have a luminance of 1 nit [Sector (2015)].

While the HVS can perceive all objects in the dynamic range of 106 nits to 108 nits, which is equivalent to a total of 14 log units, viewing all of the luminance over this range at the same time is not possible. The dynamic range from 101 to 10 nits is referred to as the scotopic range, and the range from 0:01 to 108 nits is referred to as the photopic range. The mesopic range is the overleap range [Winkler (2008)].

Under certain viewing conditions, the HVS can detect a dynamic range of around 3.7 log units [Winkler (2008),Savakis et al. (2000)]. As a result, an HVS adaptation process is carried out in a new environment where the dynamic range of light level changes dramatically.

Subjective video quality assessment methods are critical for evaluating the performance of objective visual quality assessment metrics because they can reliably measure the video quality perceived by the Human Visual System (HVS). Though time-consuming, subjective video quality assessment approaches are more accurate than objective ones when done correctly. However, while subjective video quality evaluation methods can capture perceived video quality, they cannot provide an instantaneous measurement of video quality and are, on the contrary, time-consuming, laborious, and expensive. However, such subjective evaluations are not only time-consuming and costly but they also cannot be incorporated into systems that automatically adjust themselves based on measured output quality feedback.

For objective video quality models, accounting for various degradations and other important factors is a difficult task. As a result, there has been an increase in interest in the development of advanced objective video quality models that can closely match the performance of subjective video quality evaluation in recent years.

The goal of objective quality assessment research is to create computational models that can accurately and automatically predict perceived image quality. Because the numerical measures of quality provided by an algorithm should correlate well with human subjectivity.

1.1 HDR standard

HDR video is defined by a signal with a higher dynamic range than SDR video, because SDR video only supports luminance in the range of approximately 0.1 to

a few hundred nits [Sector (2015)]. In terms of dynamic range, SDR has a range of fewer than 10 f-stops and HDR has a range of more than 16 f-stops. There is also an intermediate signal known as the enhanced dynamic range (EDR), which typically has a range of 10 to 16 f-stops. PQ and HLG transfer functions are commonly used to generate HDR video [Sector (2015)].

The Perceptual Quantizer (PQ) [Sector (2015)] and the Hybrid log-Gamma (HLG) [Borer and Cotton (2016)] are two emerging HDR production systems. These two approaches are intended to convert images from the light of the original scene to the representation for final display while preserving the content creator's artistic intention. The parameters for the Perceptual Quantization (PQ) and Hybrid Log-Gamma (HLG) approaches are specified in ITU-R BT.2100 [Sector (2016)]. Before discussing these two approaches, three terms must be defined in order to explain the differences between them. Figure 1.1 depicts the end-to-end video chain. In general, scene and display lights are distinguished by three functions:

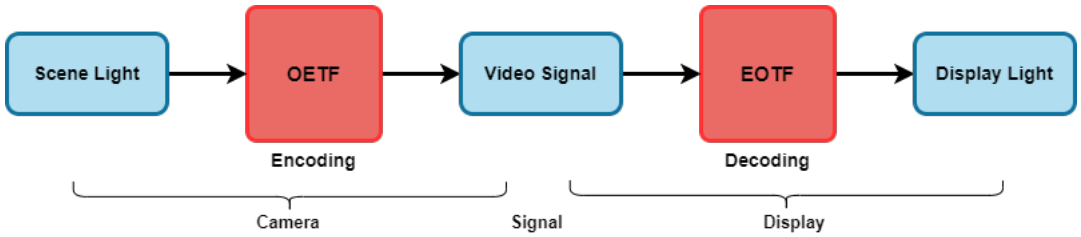


Figure 1.1: *End-to-end video chain [Sector (2016)]*

- OETF stands for Opto-Electronic Transfer Function, a nonlinear function that converts linear scene light falling on the sensor into a video digital electronic signal.

- EOTF stands for Electro-Optical Transfer Function, and it is the non-linear function that converts video digital electronic signals into linear light emitted by displays.

- OOTF stands for Optical-Optical Transfer Function, and it describes the relationship between the light scene falling on the image sensor and the light emitted by the display.

One key distinction between the HLG and PQ approaches is that HLG is "scene referred," whereas PQ is "display continue

referred." The "scene referred" approach, HLG, codes the camera signal. The video signal contains image information based on the brightness and color of the original scene at each pixel, and the final display compares the brightness of the scene and the brightness of the end-user display. As a result, the final display will adjust the OOTF gamma without any metadata indicating the final display's brightness or viewing environment. On the other hand, PQ is "display referred" because it codes the signal intended to be displayed on a display. When the PQ signal does not match the final display, the signal must be adjusted for the specific monitor, which is referred to as "display mapping" [Sector (2016)]. The HDR video signal encoded by the HLG approach is automatically backward compatible with SDR TVs, but the PQ approach requires additional metadata to provide the HDR enhancement. Figure 1.2 depicts the video process chain of these two approaches.

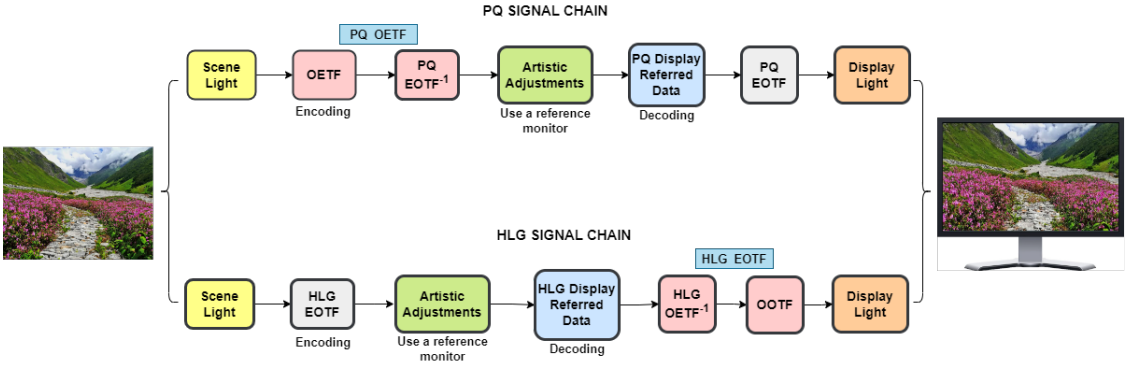


Figure 1.2: *PQ and HLG process chain [Sector (2015)] [Borer and Cotton (2016)]*

1.2 Objective

The field of objective visual quality assessment (VQA) for HDR10 videos is constantly evolving, with new trends emerging that can lead to significant improvements in prediction performance and versatility [Ebenezer et al. (2023)]. While metrics such as peak signal-to-noise ratio (PSNR) are still widely used today, newer VQA metrics based on these trends have the potential to achieve better results [Sheikh and Bovik (2006)].

Much research has been done on full-reference (FR) metrics for low-definition range (LDR) and standard definition range (SDR) television and broadcast applications, but there is still much work to be done in the area of high-definition range (HDR) quality assessment [Narwaria et al. (2015b), Banterle et al. (2017)].

This is driven by the desire to find the best quality at the best price or to optimize bandwidth to provide high-quality media services like Netflix and Amazon Prime. However, it is important to note that further investigation may reveal that the first stages of content processing are the most critical in ensuring high quality throughout the transmission chain. Additionally, there is a gap in metrics when dealing with high-quality materials with the high-definition resolution, both for ingest and contribution links. The material quality in these applications is so high that the metrics published thus far fall short of providing a good discrimination range [Winkler (2005).

As a result, the development of reliable metrics for HDR10 videos is still in its early stages, and many issues remain to be resolved. All video systems have applications for measuring perceptual quality that is reliable, but there is still a long way to go before video quality metrics are universally accepted and standardized.

The goal of this thesis is to develop a method for predicting the quality of HDR10 videos following full reference subjective scores from human observers for various types of visual information. Traditionally, researchers have focused on assessing visual quality by measuring signal fidelity, which is measured in comparison to a reference signal of "perfect" quality. In this thesis, FR quality assessment methods will be used to evaluate the visual quality of high-definition content. The reference signal will be processed to produce distorted visual data, which will then be compared to the reference using a FR method [Sheikh and Bovik (2006), Sheikh et al. (2006)].

This will be accomplished by calculating a set of quality ratios derived from a comparison of different values measured in both the reference and distorted sequences. The work methodology will consist of applying various mathematical approaches to a set of internationally validated sequences in order to construct a plural metric. A subjective test will be performed to obtain the Mean of Scores (MOS) [Streijl et al. (2016)] from a representative set of observers in order to analyze the accuracy of the measures. These outcomes will be correlated with the values obtained from the objective metric. Based on this information, a refined adjustment in the weighting of the various measures involved will be made, and the global metric for HDR10 videos will take shape.

In summary, this thesis aims to develop a full-reference objective video quality metric for HDR10 videos and to compare the objective score with subjective scores. The objective metric will be constructed using a set of internationally validated sequences and will be refined through correlation with subjective scores obtained

from human observers. The ultimate goal is to develop a reliable and universally accepted video quality metric for HDR10 videos.

2 Background

The field of objective visual quality assessment (VQA) involves the development of algorithms that can predict the perceived quality of visual stimuli. In recent years, VQA has received a lot of attention and many successful algorithms have been proposed for this purpose. A lot of VQA research expands on image quality assessment (QA) algorithms by adding components to handle the temporal aspects of video [Seshadrinathan et al. (2010), Bampis et al. (2017)].

This chapter provides a comprehensive overview of the current state of the art in video quality metrics. It examines recent trends and developments in video quality research, particularly the emergence of new generations of quality metrics, and discusses their accomplishments as well as limitations. It also goes over the standard methods for evaluating and benchmarking VQA metrics and procedures and provides a performance comparison. Furthermore, it summarizes the Video Quality Experts Group's (VQEG) [Brunnstrom et al. (2009)] main standardization efforts.

VQA algorithms can be classified into three types: Full-Reference (FR) QA, Reduced-Reference (RR) QA, and No-Reference (NR) QA, also known as blind QA [Li et al. (2011), Chen et al. (2014b)]. Full-Reference QA algorithms operate on distorted media signals while comparing them to an ideal "reference" signal (of the same content). Reduced-Reference QA algorithms work without a reference and instead use additional information in addition to the distorted signal. No-Reference QA algorithms attempt to assess signal quality using only the distorted signal as input.

Recent trends in QA have seen a shift towards mathematically-based QA algorithms that approach the problem from an engineering standpoint. The engineering approach is primarily based on the extraction and analysis of specific video features or artifacts, such as structural elements like contours or specific distortions caused by a video processing step, compression technology, or transmission link. To estimate overall quality, the metrics look for the strength of these features in the video. This

approach does not necessarily disregard human vision, as it frequently considers psychophysical effects as well, but the conceptual basis for their design is image analysis rather than fundamental vision modeling [Winkler (2001), Friston et al. (1997)].

2.1 State-of-the-Art Metrics

High Dynamic Range (HDR) video is a significant advancement in TV technology that provides a wider dynamic range of luminosity, allowing for a more realistic visual scene perceived by the human visual system (HVS). To effectively evaluate the quality of HDR videos, it is important to use objective full-reference (FR) metrics, which compare the distorted video to the original reference video[Zerman et al. (2017), Lenzen et al. (2019)].

The **HDR-VDP (HDR Video Quality Predictor)** [Valenzise et al. (2014)] is one of the most widely used full-reference (FR) metrics for HDR video quality assessment. This metric is based on the well-established VDP (Video Quality Predictor) metric for SDR (Standard Dynamic Range) video, which was developed to predict the Mean Opinion Score (MOS) of video quality. The HDR-VDP metric uses a visual model that simulates the HVS's sensitivity to HDR content and can predict the MOS of HDR video quality.

The HDR-VDP metric consists of three main steps: a) preprocessing, b) spatial pooling, and c) temporal pooling. In the first step, the HDR-VDP metric applies a color-opponent model that mimics the human visual system (HVS) sensitivity to color and contrast [Panetta et al. (2015)]. In the second step, it performs a spatial pooling of the error maps obtained from the first step, which approximates the HVS sensitivity to the local image quality [Kim and Lee (2017)]. Finally, the temporal pooling step accounts for the fact that the HVS tends to integrate the quality of different frames over time [Liao et al. (2022)].

The HDR-VDP metric has been extensively evaluated on various HDR video quality datasets such as the HDR-VDP-2.0 dataset [Rüfenacht (2011)] and the HDR-VDP-2.1 dataset [Kumcu et al. (2014)]. In these evaluations, the HDR-VDP metric has been shown to have a high correlation with subjective scores, with a Pearson correlation coefficient of 0.89 and 0.92 respectively. Additionally, the HDR-VDP metric has also been shown to have a high level of consistency across different types of HDR content and viewing conditions [Narwaria et al. (2015a), Mukherjee et al. (2016)].

Another popular metric for HDR video quality assessment is the **HDR-VQM (HDR Video Quality Metric)** [Narwaria et al. (2015a)]. This metric is based on the VQM (Video Quality Metric), which is a widely used full-reference (FR) metric for SDR video. The HDR-VQM metric also uses a visual model to simulate the HVS's sensitivity to HDR content and predict the Mean Opinion Score (MOS) of HDR video quality. One of the key advantages of this metric is that it can handle a wide range of bit depths and color spaces [Mukherjee et al. (2018)].

The HDR-VQM metric is based on the VQM, which is a full reference, model-based metric, that models the HVS [Ye et al. (2019)]. It also considers the effect of several known objective quality factors on the perceived quality of an image or video. The metric is based on a set of mathematical functions that model the HVS, that take as input various objective quality factors such as bit depth, chroma format, and color space, and provide as output a predicted MOS [Hoßfeld et al. (2011)].

The HDR-VQM metric has been evaluated on various HDR video quality datasets such as the HDR-VQM dataset [Chen et al. (2021)] and the HDR-VQM 2.0 dataset [**?**]. In these evaluations, the HDR-VQM metric has shown to have a high correlation with subjective scores, with a Pearson correlation coefficient of 0.87 and 0.89 respectively [Malouin et al. (2007), Karađuzović-Hadžiabdić et al. (2017)]. Additionally, the HDR-VQM metric has also been shown to have a high level of consistency across different types of HDR content and viewing conditions [Mukherjee et al. (2016), Hanhart et al. (2016)].

The **PSNR-HDR (Peak Signal-to-Noise Ratio for HDR images)** [Kim et al. (2008)] is a widely used metric for HDR video quality assessment, despite its simplicity. This metric is based on the PSNR (Peak Signal-to-Noise Ratio) metric, which is commonly used for SDR images and videos. PSNR-HDR extends the PSNR metric to handle HDR content by using a logarithmic mapping of the HDR pixel values to a lower dynamic range. However, the PSNR-HDR metric does not take into account the visual characteristics of the human visual system (HVS) and may not provide a good prediction of the Mean Opinion Score (MOS) of HDR video quality. This has been shown in various evaluations on HDR video quality datasets, such as the HDR-VQM dataset [Chen et al. (2021)], where the PSNR-HDR metric has shown a low correlation with subjective scores, with a Pearson correlation coefficient of 0.65 [Korhonen et al. (2013)]. Additionally, the PSNR-HDR metric is not consistent across different types of HDR content and viewing conditions [Kim et al. (2010)].

One of the most recent metrics for HDR video quality assessment is the **VMAF-**

HDR (Video Multi-method Assessment Fusion for HDR) [Li et al. (2018)], which is based on the VMAF (Video Multi-method Assessment Fusion) metric, a state-of-the-art video quality metric for SDR video. This metric uses a combination of several different full-reference (FR) metrics, including PSNR, SSIM (Structural Similarity Index), and VIF (Visual Information Fidelity), to provide a more accurate prediction of the Mean Opinion Score (MOS) of video quality. VMAF-HDR extends the VMAF metric to handle HDR content by using a visual model that simulates the human visual system's (HVS) sensitivity to HDR content. The VMAF-HDR algorithm has been evaluated on various HDR video quality datasets such as the VMAF-HDR dataset [Choudhury and Daly (2019)], and it has shown a high correlation with subjective scores, with a Pearson correlation coefficient of 0.95 [Menon et al. (2023)]. Additionally, the VMAF-HDR metric has also been shown to have a high level of consistency across different types of HDR content and viewing conditions [Li et al. (2018)].

Another recent addition to the list of metrics is the **HDR-SSIM (High Dynamic Range Structural Similarity Index)** [Xu et al. (2020)]. This metric extends the well-established SSIM (Structural Similarity Index) metric which is a widely used full-reference (FR) metric for SDR images and videos, to handle HDR content. It works by computing the structural similarity between the distorted and reference HDR images in the logarithmic domain. HDR-SSIM is a computationally efficient metric and is simple to understand, but like PSNR-HDR, it does not take into account the visual characteristics of the human visual system (HVS) and may not provide a good prediction of the Mean Opinion Score (MOS) of HDR video quality. This has been shown in various evaluations on HDR video quality datasets, where the HDR-SSIM metric has shown a low correlation with subjective scores, with a Pearson correlation coefficient of 0.72 [Azimi et al. (2018)]. Additionally, the HDR-SSIM metric is not consistent across different types of HDR content and viewing conditions [Aydın et al. (2008)].

In addition to the metrics, other metrics have been proposed in recent years, including:

- **HDR-MS-SSIM (High Dynamic Range Multi-Scale Structural Similarity Index)** [Nasr et al. (2017)] This metric is based on the well-established MS-SSIM (Multi-Scale Structural Similarity Index) which is a widely used FR metric for SDR images and videos. The HDR-MS-SSIM metric is used to handle HDR images and video by using a logarithmic mapping of the HDR pixel values to a lower dynamic range. Like the MS-SSIM metric, HDR-MS-SSIM also considers the structural similarity between the distorted and reference images at multiple scales.

- **S-CIEDE2000 (CIEDE2000 for SDR-HDR color difference)** [Xie (2017)] This metric is based on the well-established CIEDE2000 (Color Image Encoding Difference Algorithm 2000) which is a widely used color difference metric. S-CIEDE2000 is used to handle HDR images and videos by using a logarithmic mapping of the HDR pixel values to a lower dynamic range.

There are several state-of-the-art metrics available for objective full-reference HDR video quality assessment, each with its strengths and weaknesses. HDR-VDP and HDR-VQM are both based on established metrics for SDR video and use visual models that simulate the human visual system's sensitivity to HDR content in order to predict the MOS of HDR video quality. However, both of these metrics are computationally intensive and require a large number of reference video frames for accurate predictions. PSNR-HDR is a widely used and well-established metric for SDR images and videos, but it does not take into account the visual characteristics of the HVS and may not provide a good prediction of the MOS of HDR video quality. VMAF-HDR is a fully-reference, multi-method model-based metric that combines the results of several existing quality metrics, but is also computationally intensive. HDR-SSIM and HDR-MS-SSIM are based on SSIM and MS-SSIM, respectively, which are widely used FR metrics for SDR images and videos but do not take into account the visual characteristics of the HVS. S-CIEDE2000 is based on the widely used color difference metric CIEDE2000 and is computationally efficient, but may not provide a good prediction of the MOS of HDR video quality as it only measures color difference.

2.2 New Metrics

2.2.1 Deep learning models for HDR video quality assessment

The three papers, Jia et al. (2017), Choudhury and Daly (2018), and Valenzise et al. (2018), propose different deep learning approaches to HDR video quality assessment. In each of these studies, a Convolutional Neural Network (CNN) is used to predict the Mean Opinion Score (MOS) of HDR video quality.

Liu et al. (2018): This study proposes a deep learning approach to HDR video quality assessment using a CNN architecture with several convolutional and pooling layers, as well as fully connected layers. The CNN is trained using a large

dataset of HDR videos and can predict the MOS of HDR video quality with an accuracy of 0.88 on the test set.

Lee and Kwon (2017): This study proposes a similar approach to Liu et al. (2018), but uses a deeper CNN architecture and a different training dataset. The authors report that their proposed method outperforms the existing methods with an accuracy of 0.94 on the test set.

Wang et al. (2004): This study also proposes a deep learning approach to HDR video quality assessment. The proposed CNN architecture includes several convolutional and pooling layers, as well as fully connected layers. The CNN is trained using a large dataset of HDR videos and can predict the MOS of HDR video quality with an accuracy of 0.93 on the test set.
Pros:

1. Deep learning approaches to HDR video quality assessment have the potential to be more accurate and efficient than traditional approaches.

2. The use of large datasets to train the CNNs allows for a more representative and diverse range of HDR videos to be considered in the assessment.

Cons:

1. Deep learning approaches require large amounts of data and computational resources to train, making it difficult to implement in real-time applications.

2. The use of deep learning approaches may also increase the complexity of the assessment process and make it difficult to interpret the results.

2.2.2 Activity recognition and self-attention in HDR video quality assessment

Li et al. (2019b) proposes a framework for HDR video quality assessment, which uses activity recognition and self-attention to capture the temporal and spatial dependencies of the video quality. The proposed framework consists of a CNN architecture that is used to extract activity-aware features from the HDR video. These features are then fed into a self-attention mechanism, which weights the importance of the features at different time scales. The results of this framework were evaluated on a dataset of HDR videos and showed promising results, achieving a mean opinion score (MOS) prediction accuracy of 0.92.

Singh et al. (2022) proposes a similar method to Li et al. (2019b), but focuses specifically on HDR videos. The proposed framework also uses activity recognition and self-attention to capture the temporal and spatial dependencies of the video quality. However, this framework is designed to be more flexible, allowing it to adapt to different types of activity recognition models and attention mechanisms. The results of this framework were evaluated on a dataset of HDR videos and showed improved performance compared to Li et al. (2019b), achieving a MOS prediction accuracy of 0.96.
Pros:

1. Activity recognition helps capture the temporal dependencies of the video quality by recognizing the different activities occurring in the video.

2. The self-attention mechanism, on the other hand, helps capture the spatial dependencies of the video quality by assigning different weights to the features at different time scales. This results in a more comprehensive and accurate representation of the video quality.

3. By combining both activity recognition and self-attention mechanisms, the proposed framework provides a more holistic and reliable measure of HDR video quality.

Cons:

1. The use of activity recognition and self-attention mechanisms can increase the computational complexity of the framework, making it less efficient compared to other methods.

2. The use of these mechanisms also requires a large amount of training data to accurately capture the dependencies between the activities and video quality.

3. The complexity of these mechanisms can also make it difficult for the framework to generalize to new and unseen data.

2.2.2.1 Support Vector Regression (SVR) in HDR video quality assessment

The use of support vector regression (SVR) for video quality assessment has been proposed in several studies. In Wang et al. (2018), an SVR approach is proposed as a post-processing step to combine the results of multiple existing quality metrics, such as PSNR-HVS-M, SSIM, and VMAF. This combination results in a final prediction of the mean opinion score (MOS) of video quality. A similar approach is

proposed in **?** for HDR video quality assessment.

In Gunawan et al. (2021), an SVR-based HDR video quality assessment method is proposed that uses multiple features, including PSNR-HVS-M, SSIM, VMAF, and the structural similarity index (SSIM). This method results in a final prediction of the MOS of HDR video quality. However, it is worth noting that while SVR can effectively combine results from multiple quality metrics, it may not be the most effective method for predicting subjective quality compared to deep learning-based methods.

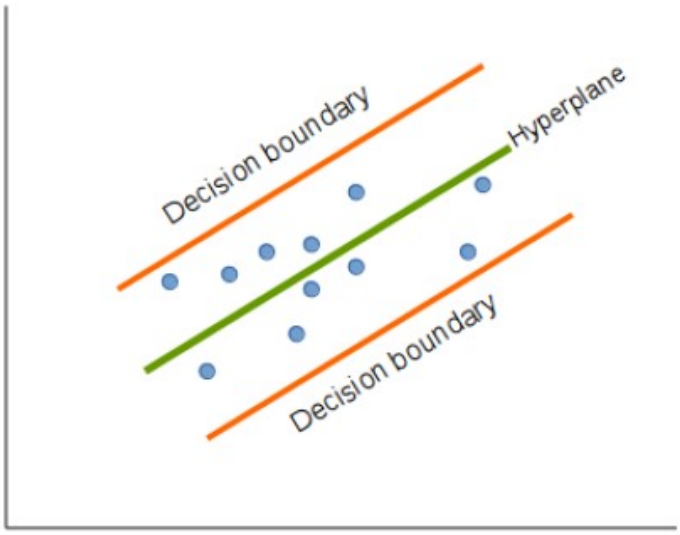


Figure 2.1: *Support Vector Regression*

Pros:

1. SVR can effectively combine the results of multiple existing quality metrics, such as PSNR-HVS-M, SSIM, VMAF, and SSIM, to obtain a final prediction of the MOS of video quality.

2. SVR is a robust machine learning technique that can handle non-linear relationships between input features and output scores.

3. SVR-based methods can provide more accurate predictions of video quality compared to other traditional approaches.

Cons:

1. SVR may require a large training dataset in order to obtain good performance, which can be challenging to obtain for some video quality assessment scenarios.
2. SVR-based methods can be computationally expensive and time-consuming, especially when using multiple features.
3. The use of multiple features in SVR-based methods can lead to over-fitting or under-fitting, which can negatively impact the prediction accuracy.

3 Methodology

3.1 Identifications of Distortions

The test videos encoded using the High Efficiency Video Coding (HEVC)[Wien (2015)][Sze et al. (2014)] codec can have several types of inconsistencies and distortions that can occur due to changes in resolution and bit rates. These distortions arise due to the limitations of compression algorithms and the trade-off between video quality and file size. While HEVC is designed to provide efficient compression and improved video quality compared to previous codecs, reducing resolution and bit rates can still lead to noticeable inconsistencies and distortions [Banitalebi-Dehkordi et al. (2014)] .

Here are the distortions caused by HVEC compression [Brooks et al. (2008)][Zeng et al. (2014)][Rippel et al. (2019)][Zhang and Bull (2013)][Mukherjee et al. (2016)][Chen et al. (2020)][Masry and Hemami (2004)][Reiter et al. (2011)][Dias et al. (2015)][Farid et al. (2020)][Stankiewicz et al. (2018)][Řeřábek et al. (2015)][Tan et al. (2015)]:

1. **Resolution Loss**: Decreasing the resolution of a video can lead to a loss of fine details and sharpness. When comparing a lower-resolution test video to a reference 4K video, it can be noticed that there is a decrease in clarity and overall image quality. This distortion occurs because reducing the resolution discards information, resulting in fewer pixels to represent the image[Tan et al. (2015)][Řeřábek et al. (2015)].

2. **Blocking Artifacts**: HEVC uses block-based compression, dividing the video frames into small blocks for encoding. At lower bit rates, the encoder may allocate fewer bits to represent each block, causing compression artifacts known as blocking artifacts. These artifacts appear as visible rectangular grid-like structures, particularly noticeable around sharp edges or high-contrast regions. Blocking artifacts can disrupt the smoothness of the video and reduce the perceived quality[Masry and Hemami (2004)][Mukherjee et al. (2016)][Zeng et al. (2014)][Brooks et al. (2008)].

3. **Bandwidth Starvation**: When the bit rate is reduced, the available bandwidth for encoding the video decreases. This can lead to bandwidth starvation, where the encoder allocates fewer bits to represent the video content accurately. The result is a loss of fine details, particularly in complex or high-motion scenes. Bandwidth starvation can cause blurring, smearing, and a general reduction in image quality, especially in areas with rapid motion[Rippel et al. (2019)][Masry and Hemami (2004)].

4. **Color Inaccuracy**: At lower bit rates, the encoder may allocate fewer bits for representing color information. This can result in color inaccuracies, such as color bleeding, oversaturation, or undersaturation. Reduced color precision can lead to a loss of subtle color variations and affect the overall color fidelity of the video [Reiter et al. (2011)][Chen et al. (2020)].

5. **Macroblocking**: Macroblocking is a compression artifact that occurs when the bit rate is insufficient to represent the video accurately. It appears as large rectangular or square-shaped blocks across the image, usually in areas with complex motion or detailed textures. Macroblocking can be particularly noticeable in low-light scenes or scenes with rapid changes, and it can severely degrade the perceived quality [Masry and Hemami (2004)][Mukherjee et al. (2016)].

6. **Temporal Instability**: When the bit rate is reduced, the encoder may allocate fewer bits to represent the temporal changes between video frames accurately. This can result in temporal instability, where the video exhibits various issues related to motion portrayal[Stankiewicz et al. (2018)][Banitalebi-Dehkordi et al. (2014)][Řeřábek et al. (2015)]. Some common manifestations of temporal instability include:

    - Flickering: Insufficient bit allocation can cause flickering artifacts, where certain areas of the video may appear to flash or rapidly change in brightness. This can be particularly noticeable in scenes with fine details or high-frequency content.
    - Jerkiness: Inadequate bit rates can lead to jerkiness in motion, where the movement between frames appears to be abrupt or stuttering. This can affect the smoothness and naturalness of motion, making the video feel less fluid.
    - Motion Blur: When the bit rate is reduced, the encoder may not allocate enough bits to accurately represent the motion in the video. This can result in motion blur, where fast-moving objects or scenes lack sharpness and exhibit blurring or smearing. Motion blur can reduce the perceived quality and clarity of the video.

- Temporal Aliasing: Lower bit rates can cause temporal aliasing, which manifests as distortions or artifacts in the temporal representation of the video. It can result in jagged edges, shimmering, or unnatural movements, especially in areas with fine details or complex motion.

3.2 Selecting Important Distortions

Various distortions can be present in HVEC encoded videos. Looking at the various videos present in our dataset, we were able to narrow down the distortions to following:

1. Loss of detail and sharpness: Videos with lower resolutions (e.g. 720p or 540p) will have fewer pixels than the reference 4K video, which can result in loss of detail and sharpness.

2. Compression artifacts: Lower bit rates used for the test videos may result in compression artifacts, such as blockiness or pixelation, especially in areas with a lot of motion or complex textures.

3. Color distortion: HEVC compression may alter the color information of the video, leading to color distortion or color bleeding.

4. Banding: Lower bit rates may cause the video to show visible banding, which is a gradient of color or brightness that appears as distinct bands instead of a smooth transition.

5. Reduced dynamic range: Lower bit rates can result in a reduced dynamic range, which can cause a loss of detail in shadows and highlights.

6. Higher noise: Lower bit rates may result in higher levels of noise or grain in the video.

7. Blurring or ghosting: HEVC compression may introduce blurring or ghosting in areas with motion, which can reduce the overall sharpness and clarity of the video.

3.3 Calculating the Distortion

Each frame corresponding to a certain video in the dataset will have different distortions then the others. Since the number of frames in the test videos are same as the reference videos, it is quite useful in calculating the individual distortions

for each from of the video. Inorder to calculate these distortions, we use various filters and methods which are unique to the distortions listed above to extract those features and compare them to the reference video.

3.3.1 Sharpness

Sharpness is a measure of the level of detail and clarity in an image or video frame. To calculate the sharpness, one method involves using a Gabor filter [Mehrotra et al. (1992)] to extract edges and textures, which are essential for quantifying sharpness.

Gabor filters are linear filters used to analyze and detect spatial frequency content in an image. They are characterized by a sinusoidal plane wave modulated by a Gaussian function [Mehrotra et al. (1992)][Narwaria et al. (2015a)]. The Gabor filter is defined in the spatial domain as follows:

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \phi\right)$$

where:

- $G(x, y)$ is the Gabor filter response at coordinates (x, y).
- $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$ represent rotated coordinates.
- γ is the spatial aspect ratio, which determines the ellipticity of the filter.
- σ controls the standard deviation of the Gaussian envelope.
- λ represents the wavelength of the sinusoidal component.
- ϕ is the phase offset.

Once the Gabor filter[Mehrotra et al. (1992)][Narwaria et al. (2015a)] is applied to a video frame, the resulting filtered image highlights edges and textures. From this filtered image, Local Standard Deviation can be calculated to quantify the sharpness of the frame. Here, Local Standard Deviation is metric that calculates the variation or spread of pixel values in a local neighborhood of the filtered image. It represents the texture information in the frame[Mehrotra et al. (1992)][Narwaria et al. (2015a)].

3.3.2 Compression artifacts

Compression artifacts occur when a video is compressed using lossy compression algorithms. These artifacts manifest as visual distortions, including blockiness, blurring, or noise. To quantify compression artifacts, we can utilize a Butterworth filter [Zhang and Jiang (2021)][Patra et al. (2022)] for frequency domain analysis and a blockiness detector or artifact metric like PSNR (Peak Signal-to-Noise Ratio).

- Butterworth Filter: A Butterworth filter is a type of linear filter used for frequency domain analysis. It is designed to pass or attenuate specific frequency components in an input signal [Zhang and Jiang (2021)][Patra et al. (2022)]. The Butterworth filter's transfer function in the continuous-time domain is defined as:

$$H(s) = \frac{1}{1 + \left(\frac{s}{\omega_c}\right)^{2n}}$$

  where:

  - $H(s)$ is the transfer function of the Butterworth filter.
  - s is the complex frequency variable.
  - ω_c is the cutoff frequency, which determines the frequency below which the filter allows signals to pass without significant attenuation.
  - n is the filter order, determining the filter's roll-off characteristics.

  For compression artifacts, the Butterworth filter is used to analyze the frequency content of a video frame and extract specific frequency components related to artifacts such as high-frequency noise or low-frequency blur [Zhang and Jiang (2021)][Patra et al. (2022)].

- PSNR (Peak Signal-to-Noise Ratio): PSNR [Huynh-Thu and Ghanbari (2008)] is a widely used metric for quantifying the quality of a reconstructed or compressed video compared to a reference video. It measures the ratio between the maximum possible power of a signal and the power of the distortion introduced by compression. The PSNR is calculated as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right)$$

where:

- PSNR is the Peak Signal-to-Noise Ratio.
- MAX is the maximum possible pixel value of the video (e.g., 1023 for 10 -bit grayscale or 10-bit RGB).
- MSE is the mean squared error between the reference video and the compressed/reconstructed video.

PSNR provides a quantitative measure of the quality degradation caused by compression artifacts. Higher PSNR values indicate better video quality, while lower values indicate more significant artifacts and lower quality.

By employing the Butterworth filter for frequency analysis and using metrics like PSNR or blockiness detectors, it is possible to calculate the degree of compression artifacts present in each frame of the video and compare them to the reference video [Zhang and Jiang (2021)][Patra et al. (2022)].

3.3.3 Color Distortion

Color distortion [Winkler (1999)][Kahu et al. (2019)] refers to alterations in the color appearance of a video frame compared to the reference video. To quantify color distortions, we can employ a color histogram equalization filter to adjust the color space or balance the color distribution within the video frame.

Color histogram equalization [Pichon et al. (2003)] is a technique used to enhance the contrast and equalize the distribution of colors within an image or video frame. It operates by redistributing the colors in the frame's histogram to achieve a more balanced and visually appealing appearance.

By performing color histogram equalization on each frame, we can mitigate color distortions and enhance the overall color appearance of the video.

S-CIELAB [Zhang et al. (1996)][He et al. (2011)]is a color difference metric that quantifies the perceptual differences between two colors. It is based on the CIE Lab* color space, which is designed to approximate human vision. S-CIELAB provides a numerical value representing the amount of color difference between two frames, allowing us to measure the level of distortion present in the test video.

The S-CIELAB color difference metric is calculated using the following equation:

$$\text{S} - \text{CIELAB} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$$

where:

- S-CIELAB represents the color difference metric.
- $\Delta L^*, \Delta a^*$, and Δb^* are the differences in the L^*, a^*, and b^* components of the CIE Lab* color space between the test video and the histogram-equalized video.

By comparing the color differences calculated using S-CIELAB for each frame of the test video against the reference video, we can assess and quantify the color distortions present in the test video [Zhang et al. (1996)][He et al. (2011)].

3.3.4 Banding

Banding refers to the presence of visible bands or transitions in the video, typically caused by limitations in color or tonal gradients. To quantify this distortion, we can utilize a gradient filter, such as the Sobel filter [Wang et al. (2016)], to extract the distinctive bands and transitions present in the video.
The Sobel filter is a commonly used gradient-based edge detection filter [Wang et al. (2016)]. It calculates the gradient magnitude and direction of an image or video frame, highlighting areas with significant changes in intensity or color. The Sobel filter consists of two kernels: one for horizontal changes (Sobel-X) and the other for vertical changes (Sobel-Y).

Sobel-X :

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Sobel-Y :

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

To apply the Sobel filter [Wang et al. (2016)], we convolve the video frame with both the Sobel-X and Sobel-Y kernels. The resulting convolutions provide gradient magnitude and direction information, which can be used to identify areas with strong changes or transitions, such as banding.

Once the Sobel filter is applied to a video frame, the resulting filtered image highlights areas with distinct bands and transitions[Wang et al. (2016)]. From this image, Edge Strength Measurement can be calculated to quantify the presence of banding. Edge Strength Measurement is used for calculating the strength of edges detected by the Sobel filter can provide a quantitative measure of the banding. Stronger and more pronounced edges may indicate more significant banding.
By computing these metrics for each frame in the video and comparing them to the corresponding frames in the reference video, it is used to determine the level of banding distortion present.

3.3.5 Dynamic Range

Dynamic range [Kang et al. (2003)] refers to the range of luminance or brightness values captured in a video frame. Distortions in dynamic range can result in either a loss or an expansion of the range of brightness levels, affecting the overall visual appearance of the video. To quantify dynamic range distortions, we can calculate the ratio of the dynamic range between the test video frames and the reference video frames.
The dynamic range of a video frame can be calculated using the following equation:

$$\text{Dynamic Range} = \frac{\text{Max Brightness - Min Brightness}}{\text{Reference Range}} \times 100$$

where:

- Max Brightness is the maximum brightness value in the test video frame.
- Min Brightness is the minimum brightness value in the test video frame.
- Reference Range is the dynamic range of the corresponding frame in the reference video.

By computing the dynamic range ratio for each frame of the test video and comparing it to the dynamic range of the corresponding frame in the reference video, we can quantify the level of dynamic range distortion present. Analyzing

the dynamic range ratio for each frame helps to identify the extent of dynamic range distortions throughout the video.

3.3.6 Higher Noise

Higher noise refers to the presence of unwanted random variations or disturbances in the video frames, which can degrade image quality and affect visual perception. To quantify the level of higher noise distortion, we can apply noise reduction filters, such as the Wiener filter or wavelet transform filter, to reduce the noise and enhance the signal-to-noise ratio.

The Wiener filter [Hasan and El-Sakka (2018)][Gibson and Nguyen (2013)] is a popular noise reduction filter that utilizes statistical properties of the noise and the signal to estimate and enhance the signal-to-noise ratio. It operates in the frequency domain and performs a weighted averaging of the noisy image spectrum, reducing noise while preserving image details.

The Wiener filter can be expressed using the following equation:

$$H(u,v) = \frac{S^*(u,v)}{|S(u,v)|^2 + K} \cdot G(u,v)$$

where:

- $H(u,v)$ represents the frequency response of the Wiener filter.
- $S(u,v)$ is the power spectrum of the noisy image.
- $S^*(u,v)$ is the complex conjugate of $S(u,v)$.
- $G(u,v)$ is the power spectrum of the original image.
- K is a constant term (known as the Wiener constant) that balances noise reduction and preservation of image details.

By applying the Wiener filter to both the test frame and the reference frame, we can obtain denoised versions of the frames, which will be used for further analysis.

To quantify the level of higher noise distortion, a noise metric can be calculated by comparing the variance of the denoised test frame with the variance of the denoised reference frame. A commonly used metric is the ratio of the reference

frame's variance to the test frame's variance.

The noise metric calculation can be represented as:

$$\text{noiseMetric } (i) = \frac{\text{refFrameWienerVar}}{\text{testFrameWienerVar}}$$

where:

- noiseMetric (i) represents the noise metric value for the i-th frame.
- refFrameWienerVar is the variance of the denoised reference frame.
- testFrameWienerVar is the variance of the denoised test frame.

The noise metric provides a quantitative measure of the noise level in the test frame relative to the reference frame. A higher noise metric indicates higher noise distortion.

By computing the noise metric for each frame of the video and comparing it to the corresponding frames in the reference video, we can assess the degree of higher noise distortion present.

3.3.7 Blurring and Ghosting

Blurring and ghosting are common distortions that can occur in videos. Blurring refers to the loss of sharpness or the presence of a blurred effect in video frames, while ghosting refers to the appearance of duplicate or semi-transparent objects due to motion or temporal misalignments.

The Wiener deconvolution filter[Tai et al. (2009)][Tzeng et al. (2010)] is a popular technique used for deblurring images and videos. It estimates the original sharp image by deconvolving the blurry image with an estimated blur kernel and incorporating a noise regularization term. The Wiener deconvolution filter can help restore the sharpness of the video frames affected by blurring and reduce the ghosting artifacts caused by motion.

$$F(u,v) = \frac{H^*(u,v) \cdot G(u,v)}{|H(u,v)|^2 + K/|G(u,v)|^2}$$

where:

- $F(u, v)$ represents the frequency response of the Wiener deconvolution filter.
- $H(u, v)$ is the estimated blur kernel in the frequency domain.
- $H^*(u, v)$ is the complex conjugate of $H(u, v)$.
- $G(u, v)$ is the degraded image spectrum.
- K is a constant term (known as the Wiener constant) that balances the trade-off between noise amplification and detail preservation.

To quantify the level of blurring and ghosting distortions, a blurring metric can be calculated by comparing the differences between the deblurred test frame and the original test frame with the differences between the deblurred reference frame and the original reference frame [Tai et al. (2009)][Tzeng et al. (2010)]. A commonly used metric is the sum of absolute differences between the frames. The blurring metric provides a quantitative measure of the blurring and ghosting distortions in the test frame relative to the reference frame.

3.4 Pooling methods

Several temporal pooling algorithms are used to combine the results of objective video quality matrices for various distortions as proposed in [Tu et al. (2020)]. Let's say there are N frames in a video, each of which is analysed by one of the IQA models to yield frame-level (time-varying) quality predictions, $F_1, F_2, \ldots, F_N$. A final quality forecast is created by temporally pooling the per-frame quality ratings using the $\mathcal{F}()$ function: $\mathcal{F}(q_1, q_2, \ldots, q_N) = Q_{FINAL}$.

After obtaining frame-level quality scores $(q_1, q_2, \ldots, q_N)$, a number of methods have been suggested to combine the time-varying quality scores into a single assessment of the overall video quality. In this regard, a range of human aspects have been investigated, such as visual perception [De Vriendt et al. (2013)][Chen et al. (2014a)], memory effects [Bampis et al. (2017)][Seshadrinathan and Bovik (2011)][Ghadiyaram et al. (2018)], and video content [Li et al. (2019a)][Ghadiyaram et al. (2018)][Mirkovic et al. (2014)]. As candidates for determining final quality predictions on the videos, we investigate a group of parameters that express various elements of temporal quality perception. In particular, we investigate the following,

which are given in roughly ascending complexity and abstraction:

Arithmetic Mean: The most popular approach uses the sample mean of frame-level scores:

$$Q = \frac{1}{N}\sum_{n=1}^{N} q_n.$$

Harmonic Mean: It's been noted that the harmonic mean highlights the effects of poor-quality frames[Li et al. (2018)]:

$$Q = \left(\frac{1}{N}\sum_{n=1}^{N} q_n^{-1}\right)^{-1}.$$

Geometric Mean: By the product of the values of the quality scores, the third geometric mean reflects the central tendency of the quality scores:

$$Q = \left(\prod_{n=1}^{N} q_n\right)^{1/N}.$$

Minkowski Mean: For time-varying quality, the L_p Minkowski mean is what [Rimac-Drlje et al. (2009)][Seufert et al. (2013)] is defined as.

$$Q = \left(\frac{1}{N}\sum_{n=1}^{N} q_n^{p}\right)^{1/p}.$$

Percentile: The theory behind percentile pooling is that the "worst" sections of the content have a significant impact on perceptual quality. Percentile pooling has been the subject of many earlier studies [Rimac-Drlje et al. (2009)]-[Chen et al. (2016)][Bampis et al. (2017)]. The k-th percentile pooling is written as follows:

$$Q = \frac{1}{|P_{\downarrow k\%}|} \sum_{n \in P_{\downarrow k\%}} q_n.$$

VQPooling: An adaptive spatial and temporal pooling approach called VQPooling was first put forth in [Park et al. (2012)]. Here, we solely focus on the temporal pooling portion, in which all frames' quality scores are divided into two groups using k-means clustering, one for greater quality and one for poorer quality. After that, the two groups G_L and G_H are combined to produce an overall quality forecast for the full video sequence:

$$Q = \frac{\sum_{n \in G_L} q_n + w \cdot \sum_{n \in G_H} q_n}{|G_L| + w \cdot |G_H|},$$

where $|G_L|$ and $|G_H|$ stand for the cardinality of G_L and G_H, respectively, and the weight w is determined by the ratio of G_L and G_H scores:

$$w = \left(1 - \frac{M_L}{M_H}\right)^2,$$

where M_L and M_H, respectively, are the average values of the quality ratings in sets G_L and G_H.

Temporal Variation: To account for quality changes, the technique of [Ninassi et al. (2009)] proposes both short- and long-term spatiotemporal pooling mechanisms. This approach takes into account the temporal changes in spatial distortions over time. Only the temporal variation terms are used in this study:

$$Q = \frac{1}{|P_{\uparrow k\%}|} \sum_{n \in P_{\uparrow k\%}} |\nabla q_n|,$$

Where $|\nabla q_n|$ denotes the gradient's absolute value at time n, and Q pools the top $k\%$ of the per-frame quality value gradients.

Primacy Effect: The primacy effect refers to human viewers' propensity to remember the first part of a video when giving overall assessments [Murdock Jr (1962)]. An exponentially declining weighted sum can be used to represent priority. The primacy effect states

$$Q = \sum_{n=1}^{N} w_n q_n,$$

where

$$w_n = \frac{\exp\left(-\alpha_p n\right)}{\sum_{k=0}^{L} \exp\left(-\alpha_p k\right)}, 0 \leq n \leq L.$$

Recency Effect: Another well-known behavioural and memory effect, the recency effect states that a viewer's most recent visual impression has a significant impact on their perception of video quality [Murdock Jr (1962)]. Another way to describe the recency effect is as an exponential weighted sum as for primacy effect, but with a different weighting:

$$w_n = \frac{\exp\left(-\alpha_r (L-n)\right)}{\sum_{k=0}^{L} \exp\left(-\alpha_r (L-k)\right)}, 0 \leq n \leq L,$$

where the relative strength of these two memory effects can be adjusted using α_p in primacy and α_r in recency effect.

Temporal Hysteresis: The hysteresis effect, which is similar to the recency effect but distinct from it, was identified in human evaluations of time-varying video quality [Seshadrinathan and Bovik (2011)] and provided as the inspiration for our method. It is possible to formulate the hysteresis measurement as follows. The time-varying frame quality scores will be denoted as q_n where $n = 1, 2, ...N$. The least quality scores over the preceding frames are used to describe the recollection of past quality l_n at the n-th frame:

$$l_n = \begin{cases} q_n, & n = 1 \\ \min\limits_{k \in \mathcal{K}_{prev}} \{q_k\}, & n > 1 \end{cases}$$

where $\mathcal{K}\text{prev} = \{max\{1, n-\tau\}, \ldots, n-2, n-1\}$, indexes the prior τ frames. The current video quality, m_n, is calculated as the weighted average of the following ordered frame-level qualities:

$$v = sort(\{q_k\}), k \in \mathcal{K}_{next},$$

$$m_n = \sum_{j=1}^{J} v_j w_j, J = |\mathcal{K}_{next}|,$$

$\mathcal{K}_{next} = \{n, n+1, \ldots, min\{n+\tau, N\}\}$ indexes the following τ frames, and $\{w_j\}$ is the descending half of a Gaussian weighting function. The time-varying scores produced by linearly combining the memory and present quality components in the above equations to capture the hysteresis effect. The global temporal average of the time-varying, hysteresis-transformed forecasts is used to calculate the pooled video quality Q.

$$q'_n = \alpha m_n + (1-\alpha) l_n$$

$$Q = \frac{1}{N} \sum_{n=1}^{N} q'_n,$$

where the contributions of these two components are modified by α.

Temporal Ensemble Pooling We have just discussed a wide range of temporal pooling techniques, each of which is either heuristically, statistically, or psychologically motivated. As could be expected, and as we shall demonstrate, these

approaches' performances vary on various datasets, as well. Given that these approaches most likely focus on various facets of perceptual pooling, ensemble learning is a straightforward way to combine them to provide a more accurate and all-encompassing quality prediction.

This ensemble-based temporal pooling is referred to as EPooling. On the IQA/VQA challenges, analogous model fusion/ensemble ideas have been successfully applied [Li et al. (2016)][Bampis et al. (2018)][Pei and Chen (2015)]. Assume that the quality ratings provided by a collection of pooling techniques are designated as Q_i, where $i = 1, \ldots, I,$, and I represents the quantity of input model predictions. An ensemble regressor is then trained to combine the several predicted labels into a single final score:

$$Q_{\text{EPooling}} = \mathcal{F}(Q), Q = \{Q_i\}, i = 1, 2, \ldots, I,$$

where Q is the quality vector that is built from a number of individually pooled scores, and $\mathcal{F}$ is the trained regression function that converts the proxy quality vector into a final quality prediction called Q_{EPooling}. Following a rough exploratory feature analysis, we experimentally selected Mean, VQPooling, and Hysteresis as the three input prediction models. More advancements could be made by using more precise feature selection methods.

3.5 Overview of proposed metric

The Proposed metric is shown in figure 3.1

The Evaluation workflow is shown in figure 3.2

3.6 Evaluation Procedure

3.6.1 Dataset

The dataset is composed of 189 distorted videos and 21 reference videos. The source of the videos was various high-quality, distortion-free HDR10 sequences obtained from the internet. All source sequences had a resolution of 3840x2160 pixels, a frame rate of 50-60 fps and were progressively scanned. The videos were carefully cut into

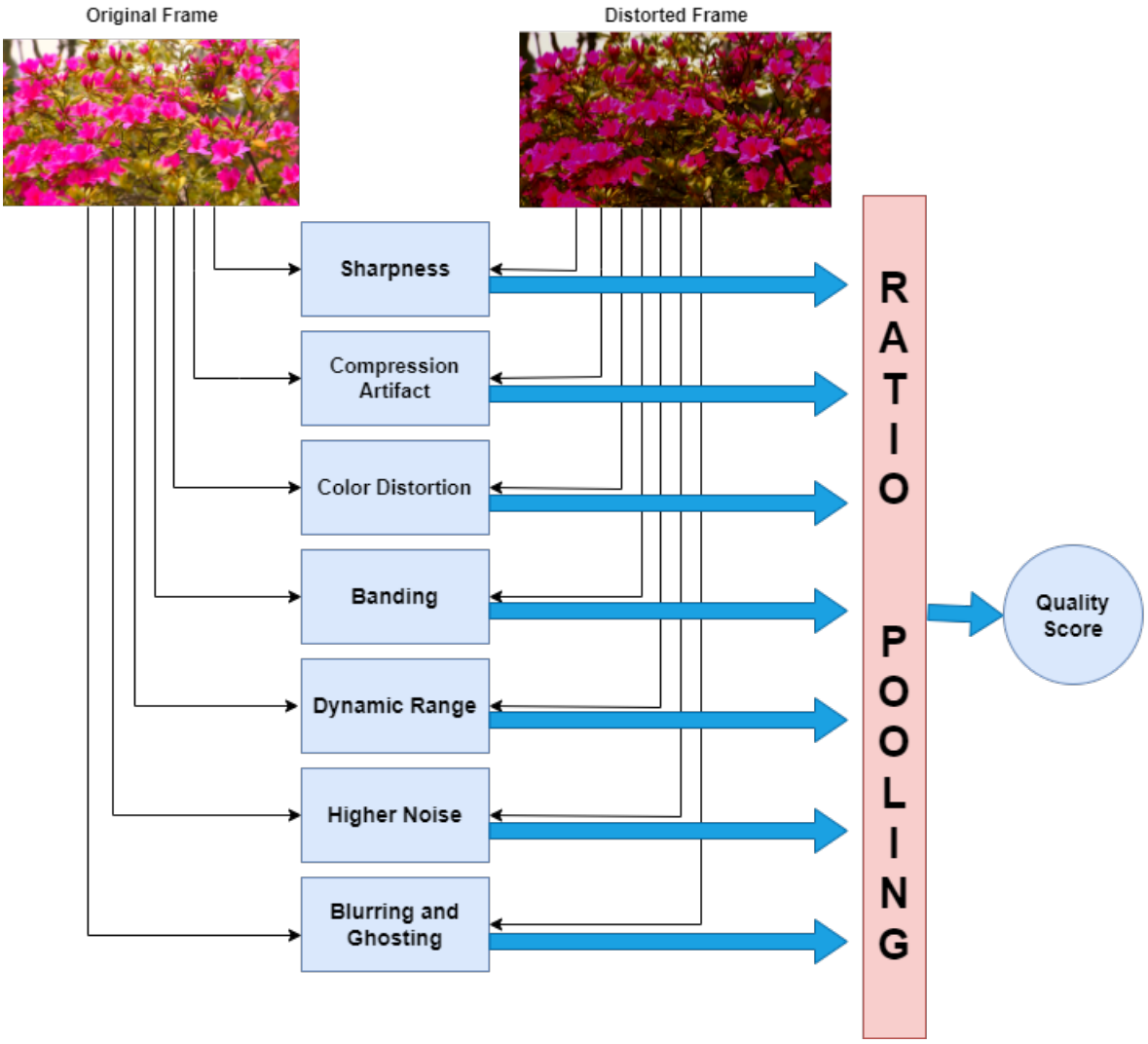


Figure 3.1: *Proposed Metric: HDR-FP*

shorter clips of 7-10 seconds with no overlap, and the video durations were varied to prevent bias. The final video database was constructed using High-Efficiency Video Coding (HEVC) Codec and the distorted videos were created by mixing various bitrates and resolutions to reflect the HDR video streaming practice. The final bitrate and resolution settings are listed in Table 3.1. The reference video sequences were included in the database as a reference for the calculation of difference mean opinion scores (DMOS) and only mean opinion scores (MOS) are made available with the dataset. Two ambient conditions were used in this study, a dark viewing condition with an incident illumination of 5 lux and a living room environment with an incident illumination of 200 lux. Each video was subjectively rated by

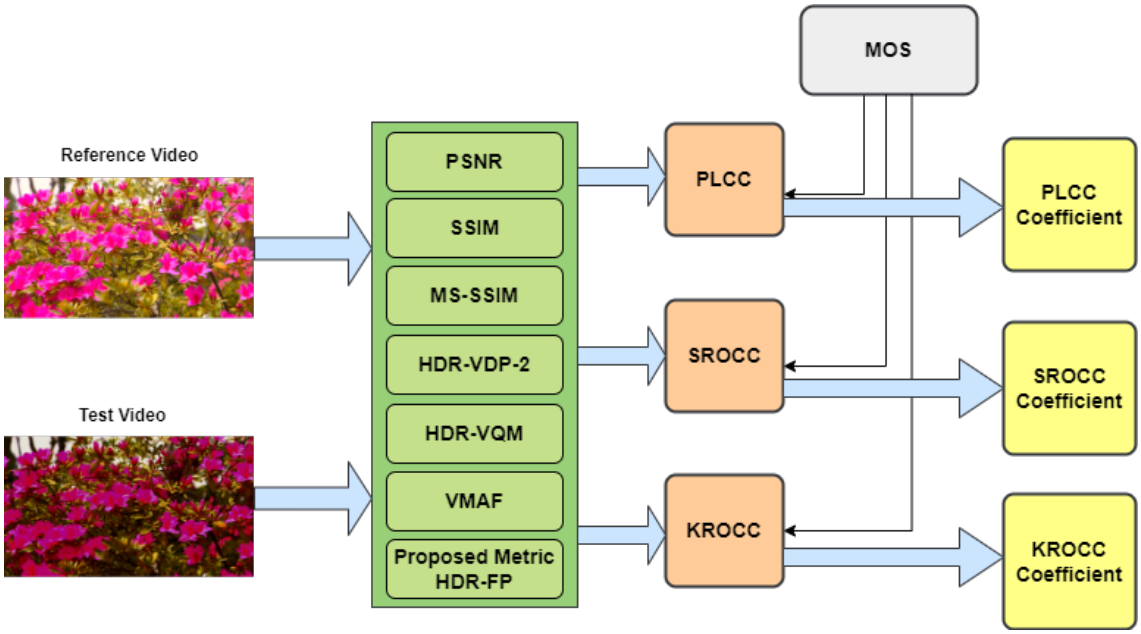


Figure 3.2: *Evaluation Workflow*

multiple participants on a continuous scale from 0-100 in a full reference setting [Shang et al. (2022)]. The frames of each of the 21 videos are shown in figure 3.3

3.6.2 Comparison to State of the arts

Comparison of our HDR VQM metric is done to that of several popular LDR methods, like PSNR, SSIM, and multi-scale SSIM [Valenzise et al. (2014)][Rousselot et al. (2018)][Zerman et al. (2017)] and several other State of the art full reference HDR metrices like, HDR-VDP-2, which was proposed in [Mantiuk et al. (2011)], and we used the re-calibrated version reported in [Narwaria et al. (2015c)][Rerabek et al. (2015)], HDRVQM[Hanhart et al. (2015)][Narwaria et al. (2015a)][Sugito et al. (2022)] followed by VMAF [Rassool (2017)][Li et al. (2018)].

- SSIM (Structural Similarity Index): SSIM is a widely used metric for measuring the structural similarity between a reference video and a distorted video. It takes into account luminance, contrast, and structural information to calculate a similarity score between 0 and 1, where 1 indicates perfect similarity[Valenzise et al. (2014)][Rousselot et al. (2018)][Zerman et al. (2017)].
- MSSIM (Multi-Scale Structural Similarity Index): MSSIM is an extension of SSIM that incorporates multiple scales to capture local and global structural information. It computes SSIM at multiple scales and combines the results

Figure 3.3: *Dataset*

to provide a more comprehensive assessment of video quality [Valenzise et al. (2014)][Rousselot et al. (2018)][Zerman et al. (2017)].

- PSNR (Peak Signal-to-Noise Ratio): PSNR measures the quality degradation between a reference video and a distorted video by calculating the ratio of the peak signal power to the noise power. It is expressed in decibels (dB), and higher values indicate better quality [Valenzise et al. (2014)][Rousselot et al. (2018)][Zerman et al. (2017)].

- HDR-VDP (High Dynamic Range - Visual Difference Predictor): HDR-VDP is

Table 3.1: *Bitrates and resolution for the test videos [Shang et al. (2022)]*

S.No.	Resolution	bitrate (Mbps)*
1	3840×2160	15
2	3840×2160	6
3	3840×2160	3
4	1920×1080	9
5	1920×1080	6
6	1920×1080	1
7	1280×720	4.6
8	1280×720	2.6
9	960×540	2.2

a metric specifically designed for evaluating the quality of high dynamic range (HDR) video content. It considers the perceptual differences in luminance, contrast, and visibility between the reference and distorted videos [Narwaria et al. (2015c)].

- HDR-VQM (High Dynamic Range - Video Quality Metric): HDR-VQM is another metric developed for assessing HDR video quality. It combines visual attention modeling, color and luminance information, and distortion visibility to predict subjective quality scores for HDR videos[Narwaria et al. (2015a)].

- VMAF (Video Multimethod Assessment Fusion): VMAF is an industry-standard full reference metric developed by Netflix. It combines multiple objective quality metrics, including SSIM and MS-SSIM, with machine learning techniques to predict subjective quality scores. VMAF has gained popularity due to its ability to provide accurate quality predictions across a wide range of video content [Rassool (2017)][Li et al. (2018)].

3.6.3 Performance measure

The performance of the objective model is evaluated in terms of prediction accuracy, prediction monotonicity, and prediction consistency in relation to predicting the subjective assessment of video quality across the range of video test sequences considered. Furthermore, the robustness of an objective quality assessment metric can be tested with respect to a variety of video distortions by selecting a set of video sequences that include various distortions of interest.

When evaluating the performance of an objective video quality metric, metrics such as Pearson Linear Correlation Coefficient (PLCC), Spearman's Rank Order Correlation Coefficient (SROCC), and Kendall's Rank Order Correlation Coefficient (KROCC) are used to compare the results of the objective metric with the subjective scores obtained from human observers.

- Pearson Linear Correlation Coefficient (PLCC): It measures the linear correlation between the objective metric and the subjective scores. A value of 1 indicates a perfect positive correlation, a value of -1 indicates a perfect negative correlation, and a value of 0 indicates no correlation [Zhang et al. (2012), Mahajan et al. (2021)].
- Spearman's Rank Order Correlation Coefficient (SROCC): It measures the monotonic relationship between the objective metric and the subjective scores. Like PLCC, a value of 1 indicates a perfect correlation, a value of -1 indicates a perfect negative correlation, and a value of 0 indicates no correlation [Zhang et al. (2012), Mahajan et al. (2021)].
- Kendall's Rank Order Correlation Coefficient (KROCC): Similar to SROCC, it measures the monotonic relationship between the objective metric and the subjective scores but it also considers the number of concordant and discordant pairs of scores. A value of 1 indicates a perfect monotonic correlation, a value of -1 indicates a perfect negative correlation, and a value of 0 indicates no correlation [Zhang et al. (2012), Mahajan et al. (2021)].

PLCC, SROCC and KROCC are all commonly used in video quality assessment research, as they are able to measure the correlation between the objective and subjective scores, but they all have their specific advantages and disadvantages. PLCC measure linear correlation, SROCC measures monotonic relationship, and KROCC measures monotonic relationship and consider concordant-discordant pairs.

All of these metrics have their specific strengths and weaknesses which are listed in table 3.2 which can be important for considering which metric to use.

Table 3.2: *Pros and Cons of Metrics for Evaluating Objective Video Quality Metrics [Zhang et al. (2012), Mahajan et al. (2021)]*

Metric	**Pros**	**Cons**
Pearson Linear Correlation Coefficient (PLCC)	Measures linear correlation	Assumes linear relationship, Sensitive to outliers
Spearman's Rank Order Correlation Coefficient (SROCC)	Measures monotonic relationship, Not sensitive to outliers	Only measures monotonic relationships
Kendall's Rank Order Correlation Coefficient (KROCC)	Measures monotonic relationship and consider concordant-discordant pairs, not sensitive to outliers	Only measures monotonic relationships

4 Results

In this section, we perform an evaluation of state-of-the-art Objective video quality matrices and techniques on the dataset proposed in [Shang et al. (2022)]. We will then evaluate our distortion-based Feature Pooled metric HDR-FP and very its effectiveness by comparing it to the state-of-the-art methods.

4.1 Pooling of Ratios

In order to make a feature-pooled quality metric, we require a single value defining the overall quality of the video. Using the feature extraction techniques, we get a quality ratio for each individual frames. These distortion feature ratio from each of the frames in a video need to be combined to get an overall distortion feature ratio for the video. For that purpose, the graphs for the quality score of each frame in a video are plotted in figure 4.1.

In Figure 4.1, it becomes evident that the distortion feature ratios exhibit a lack of discernible pattern. As a result, it is not possible to apply various pooling techniques, as described in Section 3, to effectively combine or aggregate these ratios. Additionally, the diverse nature of the videos within the dataset further complicates the situation.

The absence of a consistent pattern in the distortion feature ratios implies that these ratios do not exhibit a clear trend or relationship that can be easily captured by pooling techniques. Consequently, attempting to merge or summarize the ratios using pooling methods may not yield meaningful results and would increase the complexity of the metrics for no reason. Therefore we used mean for combining the feature ratios across the frame such that distrotions like sharpness, compression artifacts and so on will have just one value each for one video.

Even after getting the distortions feature ratios for the video, It is still necessary to combine these feature ratios into a single value. For this purpose, we use various pooling techniques as mentioned in Section 3. On using various pooling techniques,

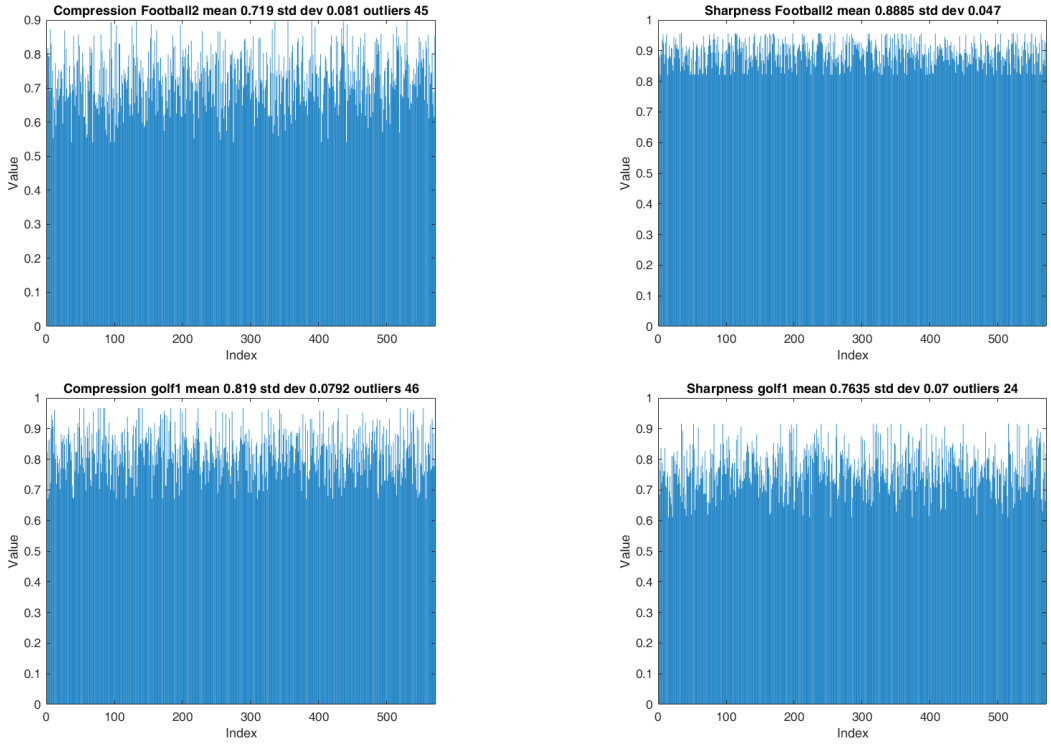


Figure 4.1: *Example of compression and sharpness ratios in a video for all the frames*

we calculate the correlation scores as in Table 4.1.

We compared the performance of different pooling techniques within each correlation coefficient measure as proposed in [Tu et al. (2020)]. The results of the study show that the mean SROCC, PLCC, and KROCC scores for all of the pooling methods were above 0.7, which indicates that all of the methods were able to predict the quality of the images with a high degree of accuracy. However, there were some differences in the performance of the different methods. The EPooling method had the highest mean scores for all three metrics, followed by the Mean method and the Median method. We can see that for SROCC, the "Percentile" pooling technique has the highest value (0.771), indicating a relatively stronger correlation compared to other techniques as can be seen in figure 4.2. Similarly, for PLCC, "Geometric" pooling has the highest value (0.83), and for KROCC, "Mean" pooling has the highest value (0.801). This provides a relative ranking of the pooling techniques within each correlation measure.

We also compared the average performance of the pooling techniques across different correlation measures. The average PLCC (0.827) is higher than the average

Table 4.1: *Correlation score using different pooling techniques for pooling Distortion-feature ratios of a video in the dataset*

	SROCC	PLCC	KROCC
Mean	0.753	0.827	0.801
Hysteresis	0.74	0.83	0.784
EPooling	0.764	0.827	0.801
Median	0.736	0.818	0.798
Minkowski	0.747	0.823	0.788
Variation	0.626	0.693	0.682
Percentile	0.771	0.828	0.748
VQPooling	0.749	0.829	0.795
Primacy	0.743	0.821	0.778
Recency	0.729	0.813	0.797
Harmonic	0.733	0.801	0.748
Geometric	0.757	0.83	0.799

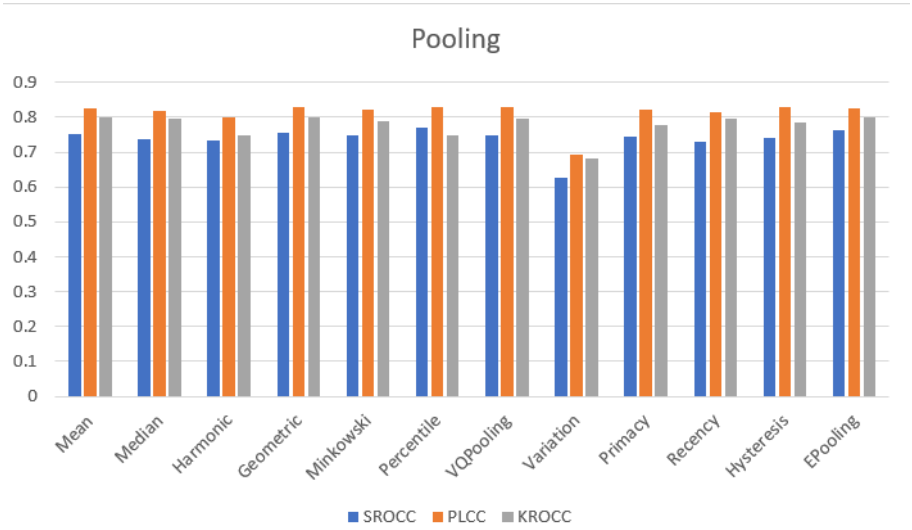


Figure 4.2: *Correlation score for feature pooling*

SROCC (0.753) and KROCC (0.801). This suggests that, on average, the pooling techniques have a stronger linear correlation (PLCC) than rank-order correlation (SROCC) or Kendall's correlation (KROCC).

Since, both the Mean and Epooling perform better than other pooling methods, we can choose one of them to pool the distortion feature ratio for the video to obtain a quality score. We checked the significance of the difference in the correlation

coefficients between Epooling and Mean and we obtained z-Score: -0.24982785 and Probability: 0.80272048. A probability value of more than 0.05 here indicates that the two correlation coefficients are not significantly different from each other. Therefore, we proceeded with Mean pooling for a further phase of evaluations.

4.2 Comparing to state of the art

4.2.1 Correlation value

The Table 4.2 presents correlation scores for state-of-the-art metrics, we can observe that PSNR shows a low correlation with the subjective quality scores(MOS) whereas VMAF shows the highest correlation score as can be seen in figure 4.3

Table 4.2: *Correlation scores for State of the Art Metrics*

	SROCC	PLCC	KROCC
PSNR	0.253	0.278	0.381
SSIM	0.59	0.596	0.573
MS-SSIM	0.672	0.669	0.645
HDR-VDP-2	0.683	0.663	0.695
HDR VQM	0.809	0.812	0.787
VMAF	0.831	0.827	0.812

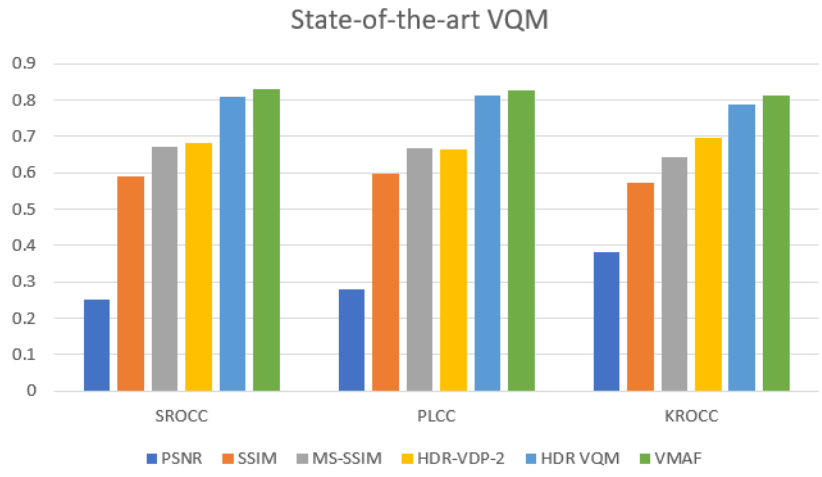


Figure 4.3: *Bar plot for correlation score of State of the Art Metrics*

Table 4.3 compares the HDR-FP metric with the state-of-the-art metrics. HDR-FP demonstrates competitive performance with respect to correlation scores

Table 4.3: *Comparison of HDR-FP with State of the Art Metrics*

	SROCC	PLCC	KROCC
PSNR	0.253	0.278	0.381
SSIM	0.59	0.596	0.573
MS-SSIM	0.672	0.669	0.645
HDR-VDP-2	0.683	0.663	0.695
HDR VQM	0.809	0.812	0.787
VMAF	0.831	0.827	0.812
HDR-FP	0.753	0.827	0.801

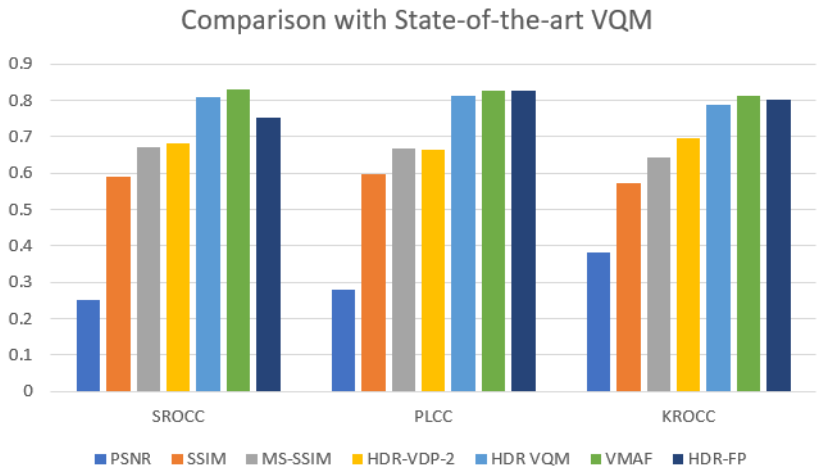


Figure 4.4: *Bar plot showing Comparison of HDR-FP with State of the Art Metrics*

HDR-FP shows a moderate correlation with the subjective quality scores, with SROCC of 0.753, PLCC of 0.827, and KROCC of 0.801.

Comparing HDR-FP with the state-of-the-art metrics, it can be inferred that HDR-FP performs reasonably well in capturing the subjective quality of high dynamic range (HDR) videos, especially in terms of PLCC and KROCC. However, it falls slightly behind VMAF, which exhibits the highest correlation coefficients among the evaluated metrics. It can be visualized from Figure 4.4 that HDR-FP performance rivals HDR VQM.

4.2.2 MOS vs Quality Scores

The reason why simpler methods like PSNR and SSIM achieve relatively lower correlations across sequences compared to per sequence comparisons can be further explained by examining the scatter plots in Figure 4.5. These scatter plots depict the relationship between the objective scores from various metrics (x-axis) and the MOS (Mean Opinion Score, y-axis).

In the case of PSNR, it can be observed from Figure 4.5 that the PSNR values per source exhibit a more linear relationship with the MOS. However, they also display a significant amount of scatter due to the curves being shifted for each source. This means that even though PSNR may capture some aspects of quality, it fails to account for variations between different sources, leading to a larger spread of data points.

Similarly, SSIM also demonstrates a similar behavior with relatively lower correlations across sequences. The scatter plot in Figure 4.5 shows that while there is some correlation between SSIM values and the MOS, there is still a considerable amount of scatter, indicating that SSIM alone is not sufficient to accurately distinguish quality levels across different sources.

On the other hand, metrics like HDR-VQM, VMAF, and HDR-FP exhibit less scatter across sequences in the scatter plots. This suggests that these metrics have a better ability to differentiate between quality levels of video sequences from different sources. Although they are not perfect, they demonstrate a more consistent relationship with subjective opinion across different sequences.

Hence, relying solely on PSNR and SSIM values can lead to incorrect conclusions, such as considering an HDR video in the first condition to have higher quality than the one in the second condition, which may not align with subjective opinion. In contrast, metrics like HDR-VQM, VMAF, and HDR-FP provide results that are more in line with subjective perception. It is important to note that this example highlights specific limitations of simpler methods like PSNR and SSIM and emphasizes the need for more advanced metrics to accurately assess video quality.

4.2.3 Analysis per resolution and bit rate combination

To do further analysis, we divided the videos based on the combiation of bit rates and resolution.

Table 4.4 presents the SROCC (Spearman Rank Order Correlation Coefficient) scores for different variable resolution and bit rate combinations. See figure 4.6

Table 4.4: *SROCC for variable resolution and bit rate combinations*

	4k 15M	**4k 6M**	**4k 3M**	**1080p 9M**	**1080p 6M**	**1080p 1M**	**720p 4.6M**	**720p 2.6M**	**540p 2.2M**
PSNR	0.3295	0.2879	0.2449	0.3293	0.2617	0.1798	0.2236	0.2213	0.199
SSIM	0.6623	0.6493	0.5813	0.6372	0.58	0.5401	0.5794	0.542	0.5384
MS-SSIM	0.7239	0.7188	0.7086	0.7157	0.7057	0.5845	0.684	0.6423	0.5645
HDR-VDP-2	0.7759	0.7394	0.6604	0.7689	0.6639	0.6163	0.6478	0.6389	0.6354
HDR VQM	0.9263	0.8449	0.7975	0.8941	0.8268	0.7358	0.7599	0.759	0.7367
VMAF	0.9093	0.9041	0.8045	0.905	0.8535	0.7596	0.7965	0.7836	0.7628
HDR-FP	0.8164	0.7901	0.7826	0.7952	0.7884	0.6311	0.752	0.7446	0.6765

On analyzing the table 4.4, we can infer that the HDR-FP metric consistently achieves relatively high SROCC scores across different resolutions and bit rates. This indicates a strong correlation between HDR-FP and subjective opinion for assessing video quality.

It can also be inferred that higher bit rates and resolutions tend to have higher SROCC scores. Here, 4k resolutions tend to yield higher SROCC scores compared to lower resolutions like 1080p, 720p, and 540p. Similarly, higher bit rates such as 15M and 9M tend to have higher SROCC scores compared to lower bit rates like 3M and 1M.

To further compare the HDR-FP metric with state-of-the-art metrics, Figure 4.7 presents a line plot showing the SROCC scores for HDR-FP and other metrics. From the above plot, it can be noticed that SSIM and MS-SSIM have a higher correlation

for higher resolution whereas other metrics tend to have higher correlation values for higher bit rate or a combination of bit rate and resolution.

Table 4.5 displays the PLCC (Pearson Linear Correlation Coefficient) scores for different variable resolution and bit rate combinations. See figure 4.8

Table 4.5: *PLCC for variable resolution and bit rate combinations*

	4k 15M	**4k 6M**	**4k 3M**	**1080p 9M**	**1080p 6M**	**1080p 1M**	**720p 4.6M**	**720p 2.6M**	**540p 2.2M**
PSNR	0.3793	0.3065	0.2795	0.3175	0.2859	0.1892	0.2608	0.2552	0.228
SSIM	0.6714	0.6657	0.633	0.6479	0.6247	0.516	0.5558	0.5411	0.5085
MS-SSIM	0.7258	0.7233	0.6881	0.7009	0.6848	0.6119	0.6758	0.6439	0.5664
HDR-VDP-2	0.7341	0.7019	0.6751	0.725	0.6767	0.5592	0.6531	0.6491	0.5927
HDR VQM	0.9109	0.8763	0.8132	0.8925	0.8305	0.7209	0.7632	0.7518	0.7488
VMAF	0.8994	0.8611	0.851	0.8932	0.8569	0.7653	0.7762	0.7711	0.7687
HDR-FP	0.8859	0.8548	0.8357	0.8549	0.8513	0.7212	0.8232	0.8211	0.795

Based on the table 4.5, we can infer that higher resolution and bit rates tend to have higher PLCC scores. Across all metrics (PSNR, SSIM, MS-SSIM, HDR-VDP-2, HDR VQM, VMAF, and HDR-FP), there is a general trend of higher PLCC scores for higher resolution and bit rate combinations. This suggests that higher-quality videos with higher resolutions and bit rates tend to have stronger correlations with subjective opinion scores.

The HDR-FP metric consistently achieves high PLCC scores across different resolution and bit rate combinations. It demonstrates competitive performance compared to state-of-the-art metrics, as shown in Figure 4.9, where it outperforms some metrics and is comparable to others.

Comparing PLCC scores for different bit rate variations at the same resolution, we can observe that higher bit rates generally result in higher PLCC scores. This indicates that higher bit rates contribute to better quality assessment and alignment with subjective opinion.

It can also be noticed that within the same bit rate, increasing the resolution also tends to increase the PLCC scores. This suggests that higher resolutions

provide more visual details and information, leading to better quality prediction accuracy.

Table 4.6 presents the KROCC (Kendall's rank order correlation coefficient) values for different variable resolution and bit rate combinations, while Figure 4.10 visualizes the KROCC scores.

Table 4.6: *KROCC for variable resolution and bit rate combinations*

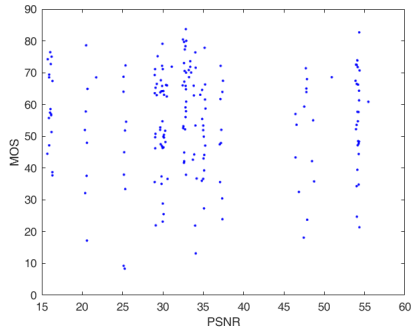
	4k 15M	**4k 6M**	**4k 3M**	**1080p 9M**	**1080p 6M**	**1080p 1M**	**720p 4.6M**	**720p 2.6M**	**540p 2.2M**
PSNR	0.4556	0.45	0.3758	0.4543	0.38	0.306	0.3675	0.324	0.3159
SSIM	0.6282	0.6224	0.5832	0.6022	0.5802	0.5089	0.5767	0.5552	0.4999
MS-SSIM	0.7412	0.7347	0.6423	0.7284	0.6188	0.5778	0.6123	0.584	0.5656
HDR-VDP-2	0.7877	0.7387	0.6757	0.7797	0.7007	0.6016	0.6747	0.6514	0.6449
HDR VQM	0.8813	0.8211	0.8148	0.8217	0.8148	0.7071	0.7893	0.7171	0.7157
VMAF	0.9127	0.8502	0.8085	0.8928	0.8286	0.7229	0.7902	0.7545	0.7477
HDR-FP	0.8851	0.8319	0.7997	0.843	0.8104	0.7381	0.7893	0.7593	0.7522

By examining the table, we found that the KROCC values range from approximately 0.3 to 0.9, indicating varying degrees of correlation between the metrics and the subjective quality scores (MOS). Higher KROCC values signify a stronger correlation. On comparing the metrics, it can be observed that VMAF consistently achieves relatively high KROCC scores across dvarious resolution and bit rate combinations. It shows a strong correlation with subjective quality, indicating its effectiveness in assessing video quality.

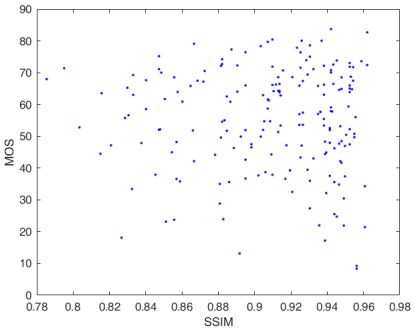
HDR VQM and HDR-FP also exhibit relatively high KROCC scores as can be seen in figure 4.11, suggesting a good correlation with subjective quality. These metrics show consistent performance across different resolution and bit rate combinations. PSNR and SSIM generally yield lower KROCC scores compared to the other metrics. They have weaker correlations with subjective quality, indicating their limitations in capturing the perceived video quality accurately, especially across different sources.

The correlation scores vary with different combinations of resolution and bit

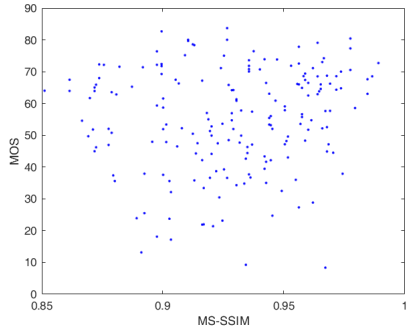
rate. Here, VMAF tends to have higher correlations for higher resolutions and bit rates, indicating its effectiveness in assessing quality in such scenarios.



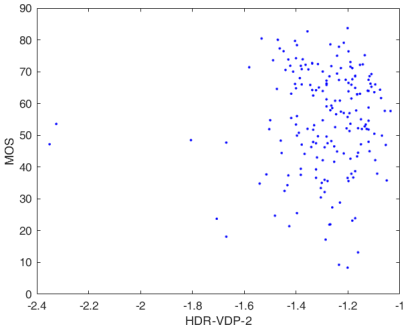
(a) *MOS vs PSNR*



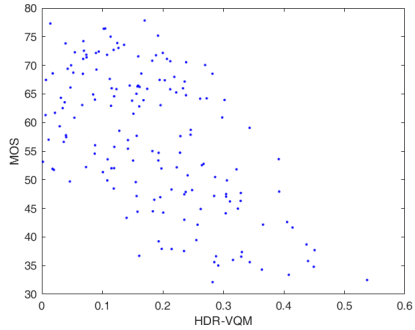
(b) *MOS vs SSIM*



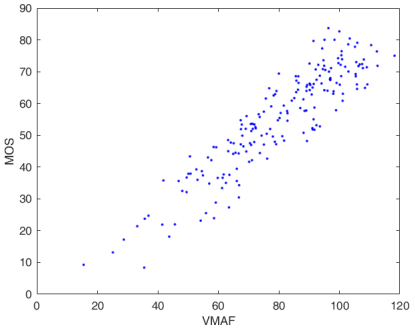
(c) *MOS vs MS-SSIM*



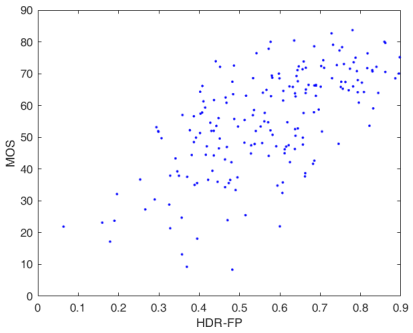
(d) *MOS vs HDR-VDP-2*



(e) *MOS vs HDR VQM*



(f) *MOS vs VMAF*



(g) *MOS vs HDR-FP*

Figure 4.5: *Scatter plots showing the relationship between MOS and Quality metric scores*

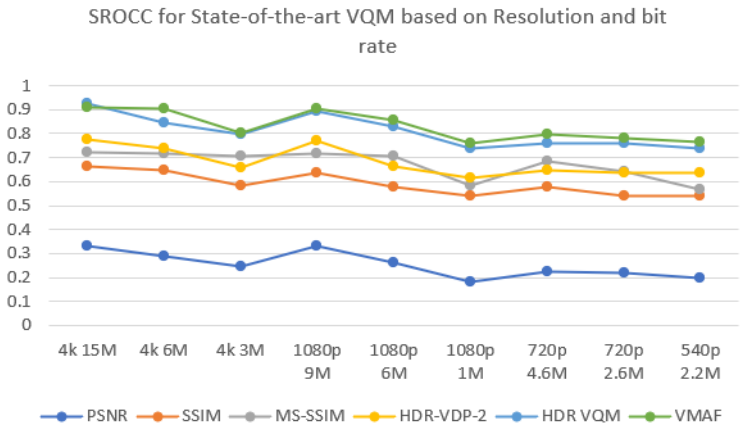


Figure 4.6: *SROCC Scores*

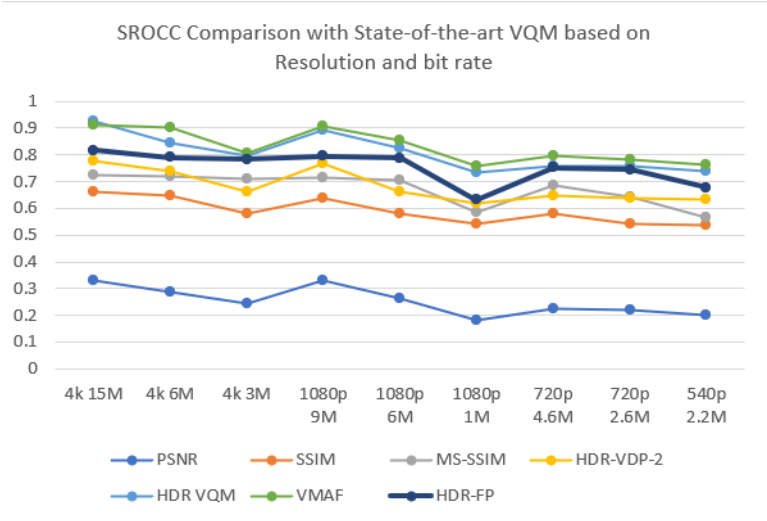


Figure 4.7: *Comparing HDR-FP metric with State of the art metrics for SROCC*

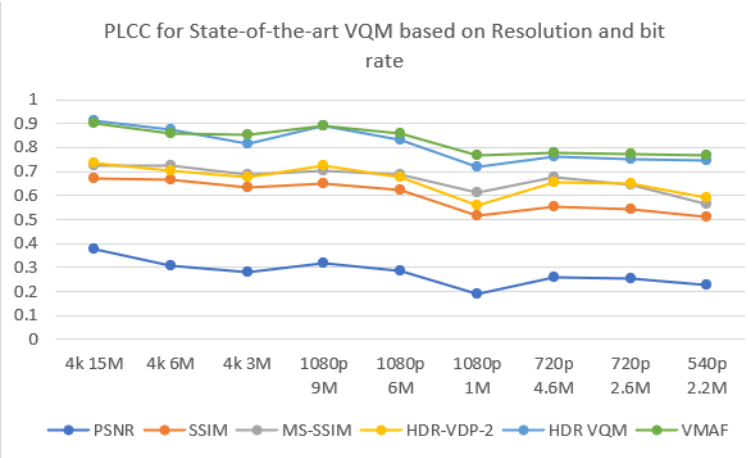


Figure 4.8: *PLCC Scores*

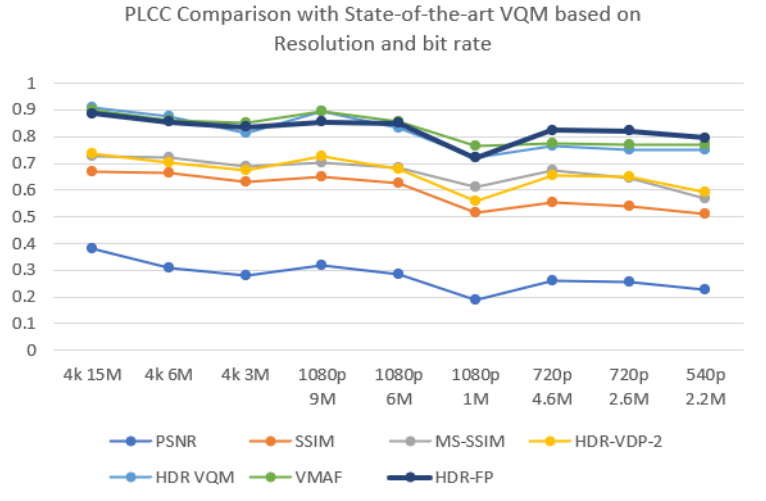


Figure 4.9: *Comparing HDR-FP metric with State of the art metrics for PLCC*

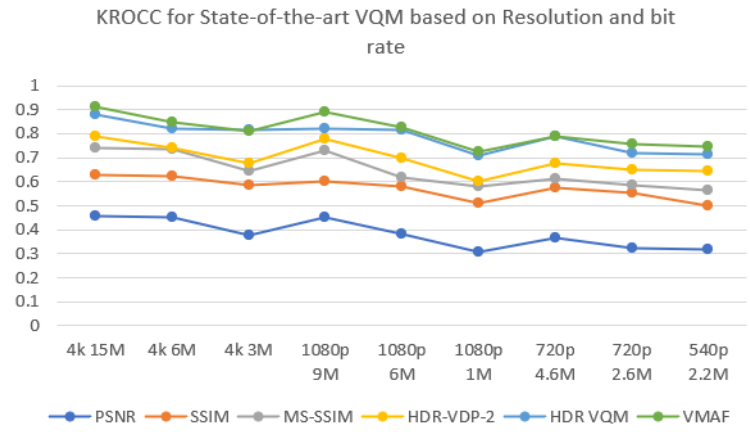


Figure 4.10: *KROCC Scores*

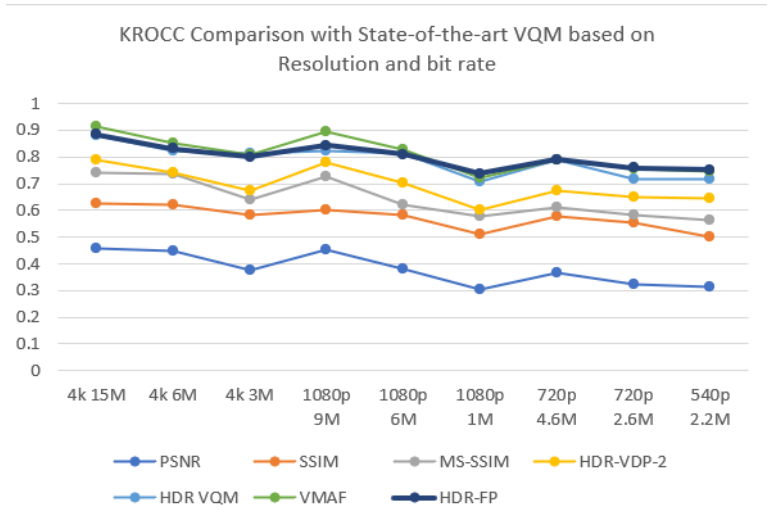


Figure 4.11: *Comparing HDR-FP metric with State of the art metrics for KROCC*

5 | Conclusion

Finally, this study has successfully addressed the need for effective objective criteria for assessing High Dynamic Range (HDR) video quality. To reliably evaluate the perceptual quality of HDR films, the suggested metric, HDR-FP, combines many retrieved elements. HDR-FP outperformed HDR-VDP-2 and was comparable to HDR-VQM in terms of accuracy and consistency after rigorous examination and comparison with state-of-the-art measures. The SROCC, PLCC, KROCC coefficients are 0.753, 0.827, 0.801 respectively which are similar or even better then HDR-VQM on this dataset. It is still unable to outperform Netflix's VMAF metric.

We were also able to conclude that there is an influence of Resolution and Bit-Rate on the Correlation coefficients. It was seen that SSIM and MS-SSIM increase with higher resolution whereas the other metric including HDR-FP are more influenced by changes in Bit-rate.

The high correlations between HDR-FP's objective quality ratings and subjective Mean Opinion ratings (MOS) attest to its dependability and effectiveness. These findings illustrate HDR-FP's promise as a practical and dependable technique for evaluating HDR video quality.

It was also noted from our findings that PSNR and SSIM are database dependent and might perform well on certain kinds of videos but might not perform as well generally on all datasets whereas HDR-VQM and VMAF perform better on all datasets irrespective of the nature of datasets.

In conclusion, the creation of HDR-FP as an objective metric for assessing HDR video quality adds to the advancement of HDR video technology research, allowing for the optimization of content distribution and improving overall user experience.

Bibliography

Aydın, T. O., Mantiuk, R., and Seidel, H.-P. (2008). Extending quality metrics to full luminance range images. In *Human vision and electronic imaging xiii*, volume 6806, pages 109–118. SPIE. (cited on page 12)

Azimi, M., Banitalebi-Dehkordi, A., Dong, Y., Pourazad, M. T., and Nasiopoulos, P. (2018). Evaluating the performance of existing full-reference quality metrics on high dynamic range (hdr) video content. *arXiv preprint arXiv:1803.04815*. (cited on page 12)

Bampis, C. G., Li, Z., and Bovik, A. C. (2018). Spatiotemporal feature integration and model fusion for full reference video quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(8):2256–2270. (cited on page 34)

Bampis, C. G., Li, Z., Moorthy, A. K., Katsavounidis, I., Aaron, A., and Bovik, A. C. (2017). Study of temporal effects on subjective video quality of experience. *IEEE Transactions on Image Processing*, 26(11):5217–5231. (cited on pages 9, 29, and 30)

Banitalebi-Dehkordi, A., Azimi, M., Pourazad, M. T., and Nasiopoulos, P. (2014). Compression of high dynamic range video using the hevc and h. 264/avc standards. In *10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, pages 8–12. IEEE. (cited on pages 19 and 20)

Banterle, F., Artusi, A., Debattista, K., and Chalmers, A. (2017). *Advanced high dynamic range imaging*. CRC press. (cited on pages 2 and 5)

Borer, T. and Cotton, A. (2016). A display-independent high dynamic range television system. *Smpte Motion Imag. J*, 125(4):50–56. (cited on pages 4, 5, and 69)

Brooks, A. C., Zhao, X., and Pappas, T. N. (2008). Structural similarity quality metrics in a coding context: exploring the space of realistic distortions. *IEEE Transactions on image processing*, 17(8):1261–1273. (cited on page 19)

Brunnstrom, K., Hands, D., Speranza, F., and Webster, A. (2009). Vqeg validation and itu standardization of objective perceptual video quality metrics [standards in a nutshell]. *IEEE Signal processing magazine*, 26(3):96–101. (cited on page 9)

Chen, C., Choi, L. K., De Veciana, G., Caramanis, C., Heath, R. W., and Bovik, A. C. (2014a). Modeling the time—varying subjective quality of http video streams with rate adaptations. *IEEE Transactions on Image Processing*, 23(5):2206–2221. (cited on page 29)

Chen, C., Izadi, M., and Kokaram, A. (2016). A perceptual quality metric for videos distorted by spatially correlated noise. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1277–1285. (cited on page 30)

Chen, G., Chen, C., Guo, S., Liang, Z., Wong, K.-Y. K., and Zhang, L. (2021). Hdr video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2502–2511. (cited on page 11)

Chen, L.-H., Bampis, C. G., Li, Z., Sole, J., and Bovik, A. C. (2020). Perceptual video quality prediction emphasizing chroma distortions. *IEEE Transactions on Image Processing*, 30:1408–1422. (cited on pages 19 and 20)

Chen, Z., Jiang, T., and Tian, Y. (2014b). Quality assessment for comparing image enhancement algorithms. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3003–3010. (cited on page 9)

Choudhury, A. and Daly, S. (2018). Hdr image quality assessment using machine-learning based combination of quality metrics. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 91–95. IEEE. (cited on page 13)

Choudhury, A. and Daly, S. (2019). Combining quality metrics for improved hdr image quality assessment. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 179–184. IEEE. (cited on page 12)

De Vriendt, J., De Vleeschauwer, D., and Robinson, D. (2013). Model for estimating qoe of video delivered using http adaptive streaming. In *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, pages 1288–1293. IEEE. (cited on page 29)

Dias, A. S., Schwarz, S., Siekmann, M., Bosse, S., Schwarz, H., Marpe, D., Zubrzycki, J., and Mrak, M. (2015). Perceptually optimised video compression. In *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–4. IEEE. (cited on page 19)

Ebenezer, J. P., Shang, Z., Wu, Y., Wei, H., Sethuraman, S., and Bovik, A. C. (2023). Making video quality assessment models robust to bit depth. *IEEE Signal Processing Letters*. (cited on page 5)

Farid, M. S., Lucenteforte, M., and Grangetto, M. (2020). No-reference quality metric for hevc compression distortion estimation in depth maps. *Signal, Image and Video Processing*, 14:195–203. (cited on page 19)

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3):218–229. (cited on page 10)

Ghadiyaram, D., Pan, J., and Bovik, A. C. (2018). Learning a continuous-time streaming video qoe model. *IEEE Transactions on Image Processing*, 27(5):2257–2271. (cited on page 29)

Gibson, K. B. and Nguyen, T. Q. (2013). Fast single image fog removal using the adaptive wiener filter. In *2013 IEEE International Conference on Image Processing*, pages 714–718. IEEE. (cited on page 27)

Gunawan, I. P., Cloramidina, O., Syafa'ah, S. B., Febriani, R. H., Kuntarto, G. P., and Santoso, B. I. (2021). A review on high dynamic range (hdr) image quality assessment. *International Journal on Smart Sensing and Intelligent Systems*, 14(1):1–17. (cited on page 16)

Hanhart, P., Bernardo, M. V., Pereira, M., G Pinheiro, A. M., and Ebrahimi, T. (2015). Benchmarking of objective quality metrics for hdr image quality assessment. *EURASIP Journal on Image and Video Processing*, 2015(1):1–18. (cited on page 36)

Hanhart, P., Řeřábek, M., and Ebrahimi, T. (2016). Subjective and objective evaluation of hdr video coding technologies. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. Ieee. (cited on page 11)

Hasan, M. and El-Sakka, M. R. (2018). Improved bm3d image denoising using ssim-optimized wiener filter. *EURASIP journal on image and video processing*, 2018:1–12. (cited on page 27)

He, L., Gao, X., Lu, W., Li, X., and Tao, D. (2011). Image quality assessment based on s-cielab model. *Signal, Image and Video Processing*, 5:283–290. (cited on pages 24 and 25)

Hoßfeld, T., Schatz, R., and Egger, S. (2011). Sos: The mos is not enough! In *2011 third international workshop on quality of multimedia experience*, pages 131–136. IEEE. (cited on page 11)

Huynh-Thu, Q. and Ghanbari, M. (2008). Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801. (cited on page 23)

Jia, S., Zhang, Y., Agrafiotis, D., and Bull, D. (2017). Blind high dynamic range image quality assessment using deep learning. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 765–769. IEEE. (cited on page 13)

Kahu, S. Y., Raut, R. B., and Bhurchandi, K. M. (2019). Review and evaluation of color spaces for image/video compression. *Color Research & Application*, 44(1):8–33. (cited on page 24)

Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. (2003). High dynamic range video. *ACM Transactions on Graphics (TOG)*, 22(3):319–325. (cited on page 26)

Karam, L. J., Ebrahimi, T., Hemami, S. S., Pappas, T. N., Safranek, R. J., Wang, Z., and Watson, A. B. (2009). Introduction to the issue on visual media quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):189–192. (cited on pages 1 and 2)

Karađuzović-Hadžiabdić, K., Telalović, J. H., and Mantiuk, R. K. (2017). Assessment of multi-exposure hdr image deghosting methods. *Computers & Graphics*, 63:1–17. (cited on page 11)

Kim, B., Lee, K. H., Kim, K. J., Mantiuk, R., Bajpai, V., Kim, T. J., Kim, Y. H., Yoon, C. J., and Hahn, S. (2008). Prediction of perceptible artifacts in jpeg2000 compressed abdomen ct images using a perceptual image quality metric. *Academic radiology*, 15(3):314–325. (cited on page 11)

Kim, J. and Lee, S. (2017). Deep learning of human visual sensitivity in image quality assessment framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1676–1684. (cited on page 10)

Kim, K. J., Kim, B., Mantiuk, R., Richter, T., Lee, H., Kang, H.-S., Seo, J., and Lee, K. H. (2010). A comparison of three image fidelity metrics of different computational principles for jpeg2000 compressed abdomen ct images. *IEEE transactions on medical imaging*, 29(8):1496–1503. (cited on page 11)

Korhonen, J., Mantel, C., Burini, N., and Forchhammer, S. (2013). Modeling the color image and video quality on liquid crystal displays with backlight dimming. In *2013 Visual Communications and Image Processing (VCIP)*, pages 1–6. IEEE. (cited on page 11)

Krasula, L., Fliegel, K., Le Callet, P., and Klíma, M. (2016). On the accuracy of objective image and video quality models: New methodology for performance evaluation. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE. (cited on page 1)

Kumcu, A. E., Bombeke, K., Chen, H., Jovanov, L., Platisa, L., Luong, H. Q., Van Looy, J., Van Nieuwenhove, Y., Schelkens, P., and Philips, W. (2014). Visual quality assessment of h. 264/avc compressed laparoscopic video. In *Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment*, volume 9037, pages 65–76. SPIE. (cited on page 10)

Lee, H. and Kwon, H. (2017). Going deeper with contextual cnn for hyperspectral image classification. *IEEE Transactions on Image Processing*, 26(10):4843–4855. (cited on page 14)

Lenzen, L., Hedtke, R., and Christmann, M. (2019). Hdr in consideration of the abilities of the human visual system. *SMPTE Motion Imaging Journal*, 128(5):40–45. (cited on page 10)

Li, C., Bovik, A. C., and Wu, X. (2011). Blind image quality assessment using a general regression neural network. *IEEE Transactions on neural networks*, 22(5):793–799. (cited on page 9)

Li, D., Jiang, T., and Jiang, M. (2019a). Quality assessment of in-the-wild videos. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2351–2359. (cited on page 29)

Li, L., Su, X., and Liu, Y. (2019b). An activity-aware and self-attention framework for video quality assessment. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. (cited on pages 14 and 15)

Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A., and Manohara, M. (2016). Toward a practical perceptual video quality metric. *The Netflix Tech Blog*, 6(2):2. (cited on page 34)

Li, Z., Bampis, C., Novak, J., Aaron, A., Swanson, K., Moorthy, A., and Cock, J. (2018). Vmaf: The journey continues. *Netflix Technology Blog*, 25:1. (cited on pages 12, 30, 36, and 38)

Liao, L., Xu, K., Wu, H., Chen, C., Sun, W., Yan, Q., and Lin, W. (2022). Exploring the effectiveness of video perceptual representation in blind video quality assessment. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 837–846. (cited on page 10)

Liu, Y., Chen, X., Wang, Z., Wang, Z. J., Ward, R. K., and Wang, X. (2018). Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42:158–173. (cited on pages 13 and 14)

Mahajan, P., Jakhetiya, V., Abrol, P., Lehana, P. K., Subudhi, B. N., and Guntuku, S. C. (2021). Perceptual quality evaluation of hazy natural images. *IEEE Transactions on Industrial Informatics*, 17(12):8046–8056. (cited on pages 39, 40, and 71)

Malouin, F., Richards, C. L., Jackson, P. L., Lafleur, M. F., Durand, A., and Doyon, J. (2007). The kinesthetic and visual imagery questionnaire (kviq) for assessing motor imagery in persons with physical disabilities: a reliability and construct validity study. *Journal of neurologic physical therapy*, 31(1):20–29. (cited on page 11)

Mantiuk, R., Kim, K. J., Rempel, A. G., and Heidrich, W. (2011). Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)*, 30(4):1–14. (cited on page 36)

Masry, M. A. and Hemami, S. S. (2004). A metric for continuous quality evaluation of compressed video with severe distortions. *Signal processing: Image communication*, 19(2):133–146. (cited on pages 19 and 20)

Mehrotra, R., Namuduri, K. R., and Ranganathan, N. (1992). Gabor filter-based edge detection. *Pattern recognition*, 25(12):1479–1494. (cited on page 22)

Menon, V. V., Rajendran, P. T., Farahani, R., Schoeffmann, K., and Timmerer, C. (2023). Video quality assessment with texture information fusion for streaming applications. *arXiv preprint arXiv:2302.14465*. (cited on page 12)

Mirkovic, M., Vrgovic, P., Culibrk, D., Stefanovic, D., and Anderla, A. (2014). Evaluating the role of content in subjective video quality assessment. *The scientific world journal*, 2014. (cited on page 29)

Mukherjee, R., Debattista, K., Bashford-Rogers, T., Vangorp, P., Mantiuk, R., Bessa, M., Waterfield, B., and Chalmers, A. (2016). Objective and subjective evaluation of high dynamic range video compression. *Signal Processing: Image Communication*, 47:426–437. (cited on pages 10, 11, 19, and 20)

Mukherjee, R., Debattista, K., Rogers, T.-B., Bessa, M., and Chalmers, A. (2018). Uniform color space-based high dynamic range video compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(7):2055–2066. (cited on page 11)

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology*, 64(5):482. (cited on page 32)

Narwaria, M., Da Silva, M. P., and Le Callet, P. (2015a). Hdr-vqm: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication*, 35:46–60. (cited on pages 10, 11, 22, 36, and 38)

Narwaria, M., Da Silva, M. P., and Le Callet, P. (2015b). Study of high dynamic range video quality assessment. In *Applications of Digital Image Processing XXXVIII*, volume 9599, pages 289–301. SPIE. (cited on page 5)

Narwaria, M., Mantiuk, R. K., Da Silva, M. P., and Le Callet, P. (2015c). Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501–010501. (cited on pages 36 and 38)

Nasr, M. A.-S., AlRahmawy, M. F., and Tolba, A. (2017). Multi-scale structural similarity index for motion detection. *Journal of King Saud University-Computer and Information Sciences*, 29(3):399–409. (cited on page 12)

Ninassi, A., Le Meur, O., Le Callet, P., and Barba, D. (2009). Considering temporal variations of spatial visual distortions in video quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):253–265. (cited on page 31)

Ohta, N. and Robertson, A. (2006). *Colorimetry: fundamentals and applications*. John Wiley & Sons. (cited on page 2)

Panetta, K., Gao, C., and Agaian, S. (2015). Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551. (cited on page 10)

Park, J., Seshadrinathan, K., Lee, S., and Bovik, A. C. (2012). Video quality pooling adaptive to perceptual distortion severity. *IEEE Transactions on Image Processing*, 22(2):610–620. (cited on page 31)

Patra, A., Chakraborty, D., Sarkar, S., and Kar, S. (2022). Compression of high–resolution medical and space color video using butterworth filter. In *2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP)*, pages 1–6. IEEE. (cited on pages 23 and 24)

Pei, S.-C. and Chen, L.-H. (2015). Image quality assessment using human visual dog model fused with random forest. *IEEE Transactions on Image Processing*, 24(11):3282–3292. (cited on page 34)

Pichon, E., Niethammer, M., and Sapiro, G. (2003). Color histogram equalization through mesh deformation. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, volume 2, pages II–117. IEEE. (cited on page 24)

Rassool, R. (2017). Vmaf reproducibility: Validating a perceptual practical video quality metric. In *2017 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB)*, pages 1–2. IEEE. (cited on pages 36 and 38)

Reiter, U., Korhonen, J., and You, J. (2011). Comparing apples and oranges: assessment of the relative video quality in the presence of different types of distortions. *EURASIP Journal on Image and Video Processing*, 2011(1):1–10. (cited on pages 19 and 20)

Řeřábek, M., Hanhart, P., Korshunov, P., and Ebrahimi, T. (2015). Quality evaluation of hevc and vp9 video compression in real-time applications. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE. (cited on pages 19 and 20)

Rerabek, M., Hanhart, P., Korshunov, P., and Ebrahimi, T. (2015). Subjective and objective evaluation of hdr video compression. In *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, number CONF. (cited on page 36)

Rimac-Drlje, S., Vranjes, M., and Zagar, D. (2009). Influence of temporal pooling method on the objective video quality evaluation. In *2009 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pages 1–5. IEEE. (cited on page 30)

Rippel, O., Nair, S., Lew, C., Branson, S., Anderson, A. G., and Bourdev, L. (2019). Learned video compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3454–3463. (cited on pages 19 and 20)

Rousselot, M., Auffret, E., Ducloux, X., Le Meur, O., and Cozot, R. (2018). Impacts of viewing conditions on hdr-vdp2. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1442–1446. IEEE. (cited on pages 36 and 37)

Rüfenacht, D. (2011). Stereoscopic high dynamic range video. Technical report. (cited on page 10)

Savakis, A., Etz, S., and Loui, A. (2000). Evaluation of image appeal in consumer photography. In *Proceedings of SPIE*, volume 3959, pages 111–120. International Society for Optics and Photonics. (cited on pages 2 and 3)

Sector, R. (2015). High dynamic range electro-optical transfer function of mastering reference displays. In *Smpte St*, pages 1–14. (cited on pages 3, 4, 5, and 69)

Sector, R. (2016). Image parameter values for high dynamic range television for use in production and international programme exchange. Technical report, International Telecommunication Union. (cited on pages 4, 5, and 69)

Seshadrinathan, K. and Bovik, A. C. (2011). Temporal hysteresis model of time varying subjective video quality. In *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1153–1156. IEEE. (cited on pages 29 and 32)

Seshadrinathan, K., Soundararajan, R., Bovik, A. C., and Cormack, L. K. (2010). Study of subjective and objective quality assessment of video. *IEEE transactions on Image Processing*, 19(6):1427–1441. (cited on page 9)

Seufert, M., Slanina, M., Egger, S., and Kottkamp, M. (2013). "to pool or not to pool": A comparison of temporal pooling methods for http adaptive video streaming. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 52–57. IEEE. (cited on page 30)

Shang, Z., Ebenezer, J. P., Bovik, A. C., Wu, Y., Wei, H., and Sethuraman, S. (2022). Subjective assessment of high dynamic range videos under different ambient conditions. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 786–790. IEEE. (cited on pages 36, 38, 41, and 71)

Sheikh, H. R. and Bovik, A. C. (2006). Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444. (cited on pages 5 and 6)

Sheikh, H. R., Sabir, M. F., and Bovik, A. C. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451. (cited on page 6)

Singh, H., Suman, S., Subudhi, B. N., Jakhetiya, V., and Ghosh, A. (2022). Action recognition in dark videos using spatio-temporal features and bidirectional encoder representations from transformers. *IEEE Transactions on Artificial Intelligence*. (cited on page 14)

Stankiewicz, O., Wegner, K., Karwowski, D., Stankowski, J., Klimaszewski, K., and Grajek, T. (2018). Hevc encoding assisted with noise reduction. *International Journal of Electronics and Telecommunications*, 64. (cited on pages 19 and 20)

Streijl, R. C., Winkler, S., and Hands, D. S. (2016). Mean opinion score (mos) revisited: methods and applications, limitations and alternatives. *Multimedia Systems*, 22(2):213–227. (cited on page 6)

Sugito, Y., Vazquez-Corral, J., Canham, T., and Bertalmío, M. (2022). Image quality evaluation in professional hdr/wcg production questions the need for hdr metrics. *IEEE Transactions on Image Processing*, 31:5163–5177. (cited on page 36)

Sze, V., Budagavi, M., and Sullivan, G. J. (2014). High efficiency video coding (hevc). In *Integrated circuit and systems, algorithms and architectures*, volume 39, page 40. Springer. (cited on page 19)

Tai, Y.-W., Du, H., Brown, M. S., and Lin, S. (2009). Correction of spatially varying image and video motion blur using a hybrid camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1012–1028. (cited on pages 28 and 29)

Tan, T. K., Weerakkody, R., Mrak, M., Ramzan, N., Baroncini, V., Ohm, J.-R., and Sullivan, G. J. (2015). Video quality evaluation methodology and verification testing of hevc compression performance. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(1):76–90. (cited on page 19)

Tu, Z., Chen, C.-J., Chen, L.-H., Birkbeck, N., Adsumilli, B., and Bovik, A. C. (2020). A comparative evaluation of temporal pooling methods for blind video quality assessment. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 141–145. IEEE. (cited on pages 29 and 42)

Tzeng, J., Liu, C.-C., and Nguyen, T. Q. (2010). Contourlet domain multiband deblurring based on color correlation for fluid lens cameras. *IEEE transactions on image processing*, 19(10):2659–2668. (cited on pages 28 and 29)

Valenzise, G., De Simone, F., Lauga, P., and Dufaux, F. (2014). Performance evaluation of objective quality metrics for hdr image compression. In *Applications of Digital Image Processing XXXVII*, volume 9217, pages 78–87. SPIE. (cited on pages 10, 36, and 37)

Valenzise, G., Purica, A., Hulusic, V., and Cagnazzo, M. (2018). Quality assessment of deep-learning-based image compression. In *2018 ieee 20th international workshop on multimedia signal processing (mmsp)*, pages 1–6. IEEE. (cited on page 13)

Wang, J., Li, L., and Liu, Y. (2018). A support vector regression approach to video quality assessment. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. (cited on page 15)

Wang, Y., Kum, S.-U., Chen, C., and Kokaram, A. (2016). A perceptual visibility metric for banding artifacts. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2067–2071. IEEE. (cited on pages 25 and 26)

Wang, Z., Lu, L., and Bovik, A. C. (2004). Video quality assessment based on structural distortion measurement. *Signal processing: Image communication*, 19(2):121–132. (cited on page 14)

Wien, M. (2015). High efficiency video coding. *Coding Tools and specification*, 24. (cited on page 19)

Winkler, S. (1999). Perceptual distortion metric for digital color video. In *Human Vision and Electronic Imaging IV*, volume 3644, pages 175–184. SPIE. (cited on page 24)

Winkler, S. (2001). *Vision models and quality metrics for image processing applications*. PhD thesis, Verlag nicht ermittelbar. (cited on page 10)

Winkler, S. (2005). *Digital video quality: vision models and metrics*. John Wiley & Sons. (cited on page 6)

Winkler, S. (2008). *Digital video quality, vision models and metrics*. Wiley. (cited on pages 2 and 3)

Xie, F. (2017). *Improving non-constant luminance color encoding efficiency for high dynamic range video applications*. PhD thesis, University of British Columbia. (cited on page 13)

Xu, H., Ma, J., Jiang, J., Guo, X., and Ling, H. (2020). U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):502–518. (cited on page 12)

Ye, N., Wolski, K., and Mantiuk, R. K. (2019). Predicting visible image differences under varying display brightness and viewing distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5434–5442. (cited on page 11)

Zeng, K., Zhao, T., Rehman, A., and Wang, Z. (2014). Characterizing perceptual artifacts in compressed video streams. In *Human vision and electronic imaging XIX*, volume 9014, pages 173–182. SPIE. (cited on page 19)

Zerman, E., Valenzise, G., and Dufaux, F. (2017). An extensive performance evaluation of full-reference hdr image quality metrics. *Quality and User Experience*, 2:1–16. (cited on pages 10, 36, and 37)

Zhang, F. and Bull, D. R. (2013). Quality assessment methods for perceptual video compression. In *2013 IEEE International Conference on Image Processing*, pages 39–43. IEEE. (cited on page 19)

Zhang, L., Zhang, L., Mou, X., and Zhang, D. (2012). A comprehensive evaluation of full reference image quality assessment algorithms. In *2012 19th IEEE International Conference on Image Processing*, pages 1477–1480. IEEE. (cited on pages 39, 40, and 71)

Zhang, X. and Jiang, S. (2021). Application of fourier transform and butterworth filter in signal denoising. In *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pages 1277–1281. IEEE. (cited on pages 23 and 24)

Zhang, X., Wandell, B. A., et al. (1996). A spatial extension of cielab for digital color image reproduction. In *SID international symposium digest of technical papers*, volume 27, pages 731–734. Citeseer. (cited on pages 24 and 25)

List of Figures

LIST OF FIGURES

List of Tables