

Article

Objective Quality Assessment Metrics for Light Field Image Based on Textural Features

Huy PhiCong ^{1,2,3,4,*} , Stuart Perry ² , Eva Cheng ² and Xiem HoangVan ¹

¹ Faculty of Electronics and Telecommunications, VNU-University of Engineering and Technology, Vietnam National University, Hanoi 10000, Vietnam; xiemhoang@vnu.edu.vn

² School of Electrical and Data Engineering, University of Technology Sydney, Sydney 2007, Australia; stuart.perry@uts.edu.au (S.P.); eva.cheng@uts.edu.au (E.C.)

³ Faculty of Multimedia, Posts and Telecommunications Institute of Technology, Hanoi 10000, Vietnam

⁴ UTS-VNU Joint Technology and Innovation Research Centre (JTIRC), VNU-University of Engineering and Technology, Hanoi 10000, Vietnam

* Correspondence: congphuy.phi@student.uts.edu.au

Abstract: Light Field (LF) imaging is a plenoptic data collection method enabling a wide variety of image post-processing such as 3D extraction, viewpoint change and digital refocusing. Moreover, LF provides the capability to capture rich information about a scene, e.g., texture, geometric information, etc. Therefore, a quality assessment model for LF images is needed and poses significant challenges. Many LF Image Quality Assessment (LF-IQA) metrics have been recently presented based on the unique characteristics of LF images. The state-of-the-art objective assessment metrics have taken into account the image content and human visual system such as SSIM and IW-SSIM. However, most of these metrics are designed for images and video with natural content. Additionally, other models based on the LF characteristics (e.g., depth information, angle information) trade high performance for high computational complexity, along with them possessing difficulties of implementation for LF applications due to the immense data requirements of LF images. Hence, this paper presents a novel content-adaptive LF-IQA metric to improve the conventional LF-IQA performance that is also low in computational complexity. The experimental results clearly show improved performance compared to conventional objective IQA metrics, and we also identify metrics that are well-suited for LF image assessment. In addition, we present a comprehensive content-based feature analysis to determine the most appropriate feature that influences human visual perception among the widely used conventional objective IQA metrics. Finally, a rich LF dataset is selected from the EPFL dataset, allowing for the study of light field quality by qualitative factors such as depth (wide and narrow), focus (background or foreground) and complexity (simple and complex).

Keywords: light field coding; light field imaging; image quality assessment



Citation: PhiCong, H.; Perry, S.; Cheng, E.; HoangVan, X. Objective Quality Assessment Metrics for Light Field Image Based on Textural Features. *Electronics* **2022**, *11*, 759. <https://doi.org/10.3390/electronics11050759>

Academic Editors:
Constantinos Papaodysseus,
Dimitris Arabadjis,
Michail Panagopoulos and
Yiannis Boutalis

Received: 19 January 2022

Accepted: 26 February 2022

Published: 1 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, Light Field (LF) imaging has been applied to many areas such as biomedicine, e.g., otoscope [1], microscopy [2], vision-based robot control [3] and velocimetry [4]. However, LF image data have a complex structure and high dimensionality that needs to be analyzed and explored, specifically in quality assessment and representation. LF images can be degraded due to many types of distortions at different LF image processing stages such as acquisition, pre-processing, reconstruction/compression (mid-processing) and rendering/display. Hence, an effective LF Image Quality Assessment (LF-IQA) model is much needed, and this model must represent the characteristics unique to LF, such as digital focusing.

A LF describes the set of light rays traveling in every angular direction at every point in 3D space [5] and thus includes a massive amount of information about each ray of light including location x, y, z , angle Θ, O , wavelength γ and the time t . LF image data

are of high dimensionality and in general can be described by the 7D plenoptic function, $LF(x, y, z, \Theta, O, \gamma, t)$ [5]. This high dimensionality poses many challenges in capture and processing. In practice, a set of constraints is introduced to the plenoptic function, reducing it to a 4D function.

$$P = L(u, v, s, t) \tag{1}$$

where the light intensity, P is a function of the sub-aperture images (u, v) and the position (s, t) within the sub-aperture image, respectively.

However, the visualization of the 4D LF is still difficult. Thus, the most popular parameterization of the spatial and angular dimensions of LF is called the two-plane LF model to help visualize the LF. Consider a two-plane LF model with (s, t) as the plane of a set of cameras with their focal plane in the (u, v) plane. Based on this definition, there are two different perspectives that can be used to understand this model. Firstly, the light rays are collected from each camera that passes to the (u, v) plane from a specific point on the (s, t) plane, and this is denoted as a Sub-Aperture Image (SAI), as shown in Figure 1a. Hence, the 4D LF can be visualized as a 2D array of images. Secondly, at a certain point on the (u, v) plane, all points on the (s, t) plane represent the range of angles of the light rays incident at that point on the (u, v) plane, so the different viewpoints on the (s, t) plane map to the same point on the (u, v) plane, as shown in Figure 1b. This is denoted as an LF sub-view. To summarize, this two-plane model uses the (s, t) plane to refer to angular dimensions, while the (u, v) plane refers to spatial dimensions [5].

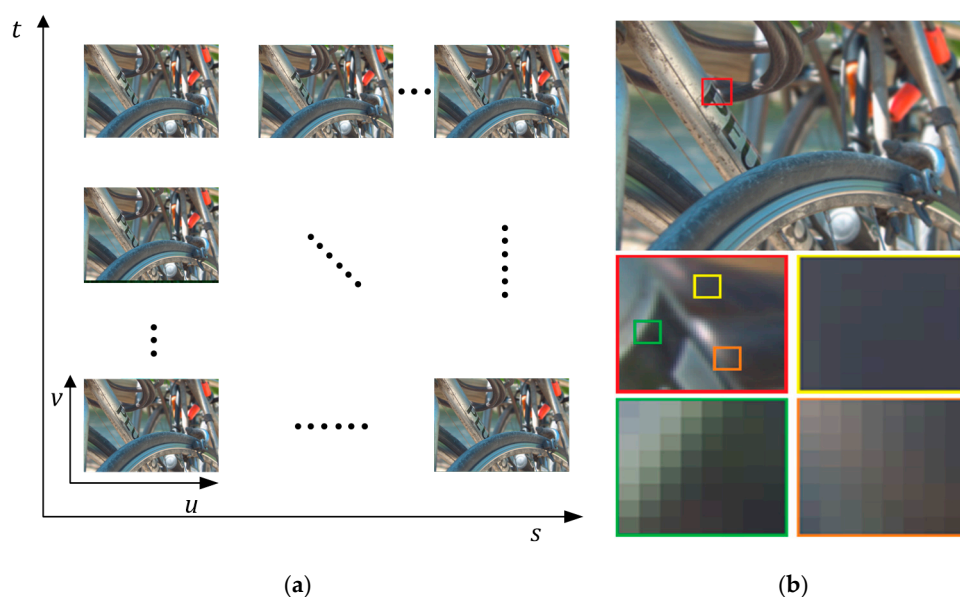


Figure 1. (a) Sub-aperture image (SAI) represented by the (s, t) and (u, v) planes; (b) LF sub-view represented by gathering samples with fixed (u, v) .

Regarding LF quality assessment, there are few works specifically concerning LF quality assessment. Most commonly, subjective and objective IQA metrics applied widely on conventional 2D image quality assessment are directly utilized for LF images. The benefits of this approach are fast prediction, high accuracy and correlation with human visual perception; however, limitations persist.

Subjective quality assessment has been considered the most reliable method of measuring the visual quality of multimedia content. In the literature, various works related to subjective quality assessment have been proposed as ways to explore the compressed and uncompressed LF content through subjective tests. For instance, the Double Stimulus Impairment Scale (DSIS) protocol and the Double Stimulus Continuous Quality Scale (DSCQS) [6,7] are two subjective testing methodologies that can be used to evaluate the

impact of encoding on LF content and the performance of different LF image compression methods, respectively. These subjective testing methods are used to examine all-in-focus views and refocused views to evaluate the visual quality of LF content. However, there remain some limitations, e.g., different impacts on the perceptual quality of blurry regions in refocused views and insufficient information to describe the overall quality of the LF image due to the need to examine many sub-aperture views. Thus, more interactive methods of evaluating LF images have been considered [8]. In a recent study [8], two methods for visualization, passive (using refocused pseudo video) and interactive (allowing the observer to select different focus points), were utilized, and the interactive method provided a better Quality of Experience (QoE) for observers. However, the limitation remains that although more interactive subjective quality methodologies provide a better QoE for observers, the greater interactivity decreases the control that the experimenter has over the number and order of viewpoints each observer visualizes. This result supports passive rendering techniques for the subjective quality assessment of LF images. It is also noted that there are various LF datasets that have been introduced for different purposes, e.g., Visual Quality Assessment Light Field Image Dataset (VALID) [9], SMART [10] and five windowed 5 degrees of freedom LFI Database (Win5-LID) [11]. However, the subjective scores are not fully aligned across the multiple datasets during the various subjective quality assessments. In this context, the subjective quality assessment with LF content thus needs further study and evaluation to identify and develop more suitable evaluation approaches.

In objective quality assessment, the aim is to devise and use a mathematical algorithm capable of predicting the quality of multimedia content with high correlation to the subjective quality judgements of the average human observer. Compared to the subjective methods, objective quality assessment provides a faster prediction of visual quality, and, notably, it can be efficiently deployed in quality driven optimization frameworks of LF processing chains. The accuracy of a subjective method is measured based on the human observers and influenced by many factors, such as the personal attitude of the observers, lighting condition, viewing distance and the configuration of the display device, while the accuracy of an objective method is measured based on its consistency, with subjective scores as the anchor. Thus, a suitable subjective method can support and strengthen the results of objective methods. Currently, the objective quality assessment for LF images is an active area of research [12–14]. There is a focus on the design of LF-IQA metrics that consider the LF characteristics (i.e., changing color information with view angle and depth information) through the analysis and extraction of LF features. However, these metrics cost additional computational complexity over the quality prediction performance. Thus, among the current well-established LF-IQA metrics, there are many widely used conventional metrics which provide low computational complexity and can be applied for LF images [7], e.g., the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) [15]. These two metrics are also used by the JPEG and MPEG groups under the JPEG Pleno [16] and MPEG-I [17] activities to evaluate the performance of LF image compression methods. However, there exists a need to evaluate the performance and complexity of the conventional existing metrics for LF quality assessment.

Although the above-mentioned image quality assessment metrics, SSIM and PSNR, have partly considered image/video content in their formulation, they were mainly designed for images with natural content captured by a generic camera. Therefore, the current state-of-the-art image quality assessment metrics may not be suitable for LF content, which meet neither criterion. To achieve high accuracy quality assessment through a metric specific to LF images, we propose in this paper an improved method for objective LF-IQA based on textural features. This objective LF-IQA is evaluated on a rich LF dataset selected from the EPFL dataset [18], which takes various information based on the characteristics of the LF images (i.e., depth, complexity and focus) for subjective quality assessment. The key contributions of this paper can be summarized as:

- *A novel content-adaptive LF quality assessment method: a wide range of conventional objective quality metrics are modelled and analyzed when augmented with different*

textural features of LF image content; in particular, Haralick texture features [19], the Discrete Cosine Transform (DCT) [20] based features and the Features-from-Accelerated-Segment-Test (FAST) method [21]. Additionally, extensive experiments are conducted on each objective quality metric to identify the most appropriate metric based on the LF image content.

- *A comprehensive feature analysis for LF images with various conditions:* for an in-depth analysis of textural features, various conditions of a rich LF dataset selected from the EPFL dataset [18] are explored and augmented by various information from the LF image such as depth (narrow and wide), complexity (simple and complex) and focus (background and foreground). Based on these conditions, the above-mentioned features are analyzed and discussed in detail.

The remainder of this paper is organized as follows. Section 2 briefly describes the related work on LF visualization techniques and objective quality metrics for LF images in general. Section 3 introduces the rich subjective quality dataset of this paper, while Section 4 proposes our improved objective LF quality assessment metrics. Section 5 presents the experimental results for discussion, and Section 6 summarizes the work and contributions of this paper and describes future work.

2. Related Work

In this section, related work is briefly reviewed and described, especially concerning the way LF visualization is used to generate LF datasets and work related to objective quality assessment of LF data.

2.1. LF Visualization Techniques

LF image display is a challenging task because dedicated LF displays are expensive and not yet widely available. The raw LF images cannot be viewed by a regular 2D or stereo 3D image display due to the complex high-dimensionality data of LF images. Thus, LF images generally need to be converted from a sub-views format to sub-aperture images and then subsampled into 2D images to present on a 2D display. In this context, many visualization techniques have been introduced to address this display issue, mainly including all-in-focus images, the display of particular Sub-Aperture Images (SAIs), the display of Refocused Images (RIs) at different focal points and Pseudo Video Sequences (PVS), where, generally, the observer is shown a sequence of SAIs [10]. A hybrid of the PVS and RI techniques are applied in this paper due to their common use in LF subject quality assessment.

The PVS of a 4D-LF image is generated and created from the SAIs, which is considered as a series of frames in the stimuli video. Each SAI in the LF is selected and displayed in a specific order and speed [22]. Thus, a PVS is similar to normal 2D video that uses motion perception to indicate the disparity information within the LF [23]. In contrast, the RI technique can be considered as similar to changing the different focal depths of an LF image. Re-focusing is achieved by the transformation and superposition of SAIs based on a specific slope related to the disparity of objects in the scene at the intended focal point and enables the control of focused and defocused regions of the LF image. Disparity refers to the distance between two corresponding points in the left and right images of a stereo pair. The effect of disparity can be observed as the degree of shift of the subject position as the viewpoint (or SAI in the case of LF images) changes. The effect is greatest for subjects closer to the eye or camera and is negligible for subjects at infinity. In an LF image, this manifests as an increment in pixels between the position of the same point in the scene as we move between adjacent SAIs. This pixel level increment is referred to as the slope. Thus, this means that, at the focal point of interest, the slope refers to the degree of shift needed between SAIs as they are averaged to produce the final re-focused image. This may be done interactively or presented as video sequence. However, the interactive viewing is more common for the RI technique.

In the literature, both PVS and RIs have been explored. For instance, the concept of six different PVS playback options is introduced in the work of Battisti [24] and illustrated in Figure 2. Each playback option affects the LF performance of subjective quality assessment and coding differently, because the angular and spatial information of LF images are fully represented through each SAI in the PVS. Meanwhile, Viola [8] identified that the interactive assessment based on the RI technique provides significant information for LF quality assessment, such as depth information, spatial structure information and semantic information. Therefore, Viola [8] introduces an interactive method based on the RI technique [25]. The interaction between the participants and LF content is enabled through a graphical user interface, where the participants are able to change the focal point of the LF. Figure 3 illustrates the change in different slopes of the LF image, generated using the MATLAB LF Toolbox ver. 04 [26].

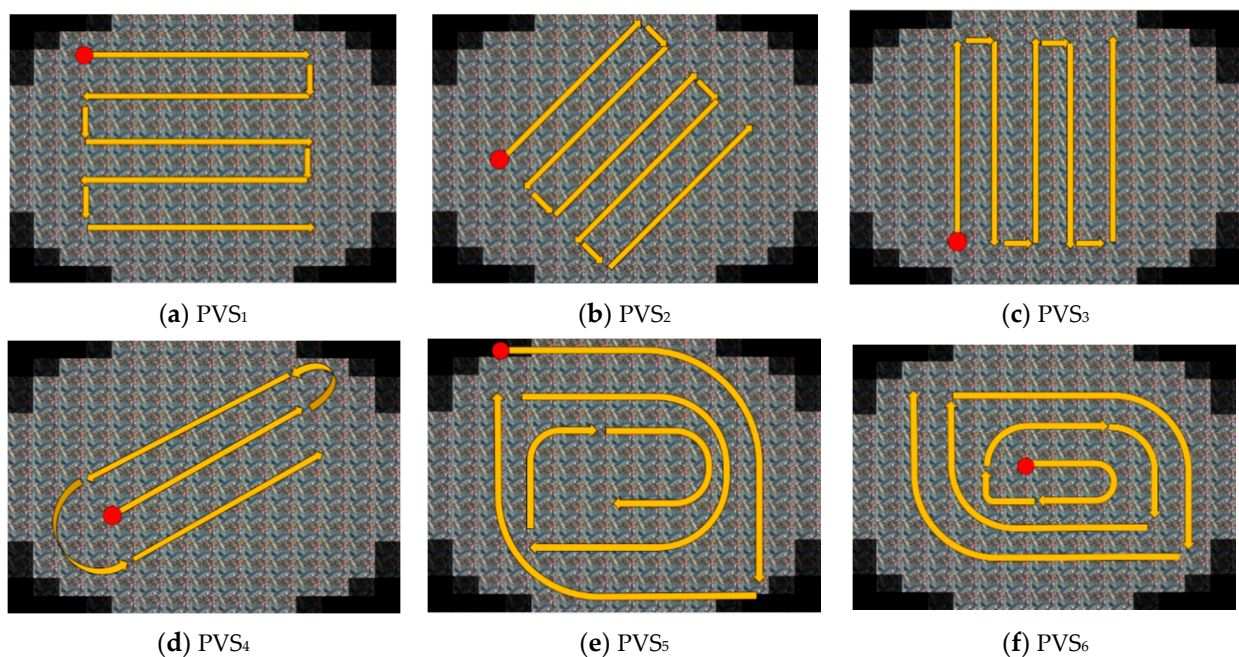


Figure 2. The concept of different PVS playback options.



Figure 3. Three RIs with different slopes: LF image—Wheat & Silos.

2.2. Objective Quality Metrics for LF Images

Objective quality metrics can provide faster prediction for visual quality; thus, there are multiple LF-IQA models introduced based on subjective quality assessment. Similarly to the 2D IQA, LF-IQA models are categorized into three types: Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR). Recently, the works of Min et al. and Tian et al. propose an FR metric to measure the quality degradation caused by LF operations [27,28]. These works consider the LF characteristics by exploiting the depth, inter-view and angular

information, e.g., symmetry feature extraction, depth feature extraction, global spatial quality, local spatial quality and angular quality estimation. Likewise, in Shan et al. [29], the author presents an NR metric that exploits 2D and 3D characteristics (i.e., the brightness, hue, saturation and texture features) of the LF with support vector regression and the depth information to obtain the final prediction score. In addition, an RR metric has been proposed by Viola [25], which takes the LF depth map as a reference. This metric measures the structural similarity of depth maps between the original and distorted LF to predict the quality of the distorted LF. Regarding the state-of-the-art LF quality assessment methods based on quality scores, there are also several recent studies [30–34] that present the overview of LF-IQA with NR or blind LF approaches based on LF characteristics (i.e., spatial-angular information) and also provide a flexible coding scheme based on block Krylov subspace approximation for LF displays.

In evaluating LF-IQA metrics, both the computational complexity of the model and performance are considered. With these two criteria, traditional objective IQAs can be considered, such as PSNR, SSIM [15], Feature Similarity Indexing Method (FSIM) [35], Multi Scale-Structural Similarity Index Method (MS-SSIM) [36], Information Content Weighted-PSNR (IW-PSNR) and Information Content Weighted-SSIM (IW-SSIM) [37]. Notably, recent research on the performance of the objective metrics for LF [38] provided an extensive analysis of the widely used objective IQA metrics. It was shown that the Most Apparent Distortion (MAD) [39], IW-SSIM [37] and the Gradient Magnitude Similarity Deviation (GMSD) [40] have the highest performance among the examined metrics for LF quality assessment, while some very popular IQA metrics (PSNR and SSIM) only rank 13th and 14th, respectively. Additionally, this comprehensive evaluation also concluded that NR metrics are not suitable for LF images. Motivated by this work, we explore how content-related features of the LF image can affect the performance of objective IQA metrics. Regarding the computational complexity, Tian et al. [14] have shown interesting results in that GMSD and PNSR have very low computational complexity compared to the other LF-IQA metrics. Based on this initial work, more evaluation and analysis of the features of LF images is needed to design an LF-IQA metric based on the conventional objective IQA metrics while balancing low complexity and accurate performance.

3. LF Dataset for Subjective Testing

3.1. LF Dataset Creation

In this study, various LF images are collected to generate the LF dataset for flexible subjective quality assessment. In our approach, a passive method that merges the traditional SAI-based PVS and interactive RI methods is used. This is based on a PVS technique, wherein 101 SAIs of the LF image are presented using PVS6 order (i.e., spiral scanning order, as shown in Figure 2) at a rate of 30 frames per second (fps), different focal points and depths of field.

A fully interactive method can provide a better QoE by giving the control of focus points to the observers and allowing interaction between observers and images. However, the limitation of the interactive method is the lack of control over the number of viewpoints that each observer visualizes. In our dataset, the user cannot control the focus points, but we provide all-in focus and multiple focus points (i.e., from background to foreground), different depth conditions (i.e., narrow and wide) and different complexity of subjects within each LF image to collect the most subjective judgements, along with control of the number of viewpoints presented to observers.

Our LF dataset is selected from the EPFL source dataset [18], with different Depth Of Field (DOF) types and Fields Of View (FOV) rendered for display to the observers. All of the LF images were taken using a Lytro Illum B01 (10-bit) camera and then extracted with the Matlab LF toolbox v0.4 [26] to obtain SAIs with 625×434 resolution. Table 1 presents an overview of this data and the justification for selection, with thumbnail visualization in Figure 4. These data are divided into two portions based on the complexity of the LF-FOV, i.e., the complex layer FOV and simple layer FOV (where the complex and simple

dichotomy refers to the number of scene objects that can be focused on and determined easily; more than two scene objects is considered as a complex layer). Each LF-FOV type contains 2 types of LF-DOF: narrow DOF and wide DOF (narrow and wide DOF refer to the depth ranges in focus in the scene, respectively). The configuration of LF slopes for changing focus is varied in the range from -0.6 (i.e., foreground focus) to 1.6 (i.e., background focus). To change the focus of the LF image, we changed the slope parameter of the LF image.

Table 1. Overview of LF data definition.

LF Slope Type	Description	Number of Samples
CWF	Complex layer with Wide DOF and focus on Foreground	10
CWB	Complex layer with Wide DOF and focus on Background	10
CNF	Complex layer with Narrow DOF and focus on Foreground	10
CNB	Complex layer with Narrow DOF and focus on Background	10
SWF	Simple layer with Wide DOF and focus on Foreground	8
SWB	Simple layer with Wide DOF and focus on Background	8
SNF	Simple layer with Narrow DOF and focus on Foreground	10
SNB	Simple layer with Narrow DOF and focus on Background	10

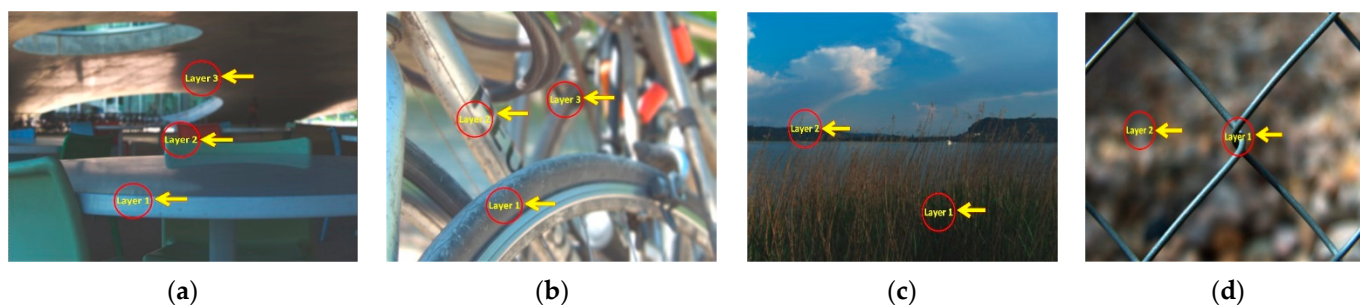


Figure 4. Thumbnails of the LF dataset with different conditions: (a) Complex layer with wide DOF, (b) Complex layer with narrow DOF, (c) Simple layer with wide DOF and (d) Simple layer with narrow DOF.

3.2. Subjective Evaluation Methodology

In our evaluation methodology, firstly, all the different slope types of the LF images were generated by the Matlab LF toolbox v0.4 [26]; then, the spiral scanning method was applied to create the video sequences with 30 fps and 12 s in total duration. The video sequences were encoded by the Versatile Video Coding codec (VVC) [41] with 3 Quantization Parameters-QPs (i.e., 22, 40, 50), using the simultaneous Double Stimulus Impairment Scale (DSIS) test method for visual quality comparison. This test method includes a hidden reference as a sanity check and uses a 5-level Likert rating scale: 1 for “very annoying”, 2 for “annoying”, 3 for “slightly annoying”, 4 for “perceptible, but not annoying” and 5 for “imperceptible”. The environment of test conditions was set up to follow ITU-R Recommendation BT.500-13 [6]. The viewing requirements used a Dell U2419H monitor with a 24-inch, full high definition (FHD) 1920×1080 resolution and a viewing distance of 1.2 m (15 cm). Regarding the video player, a customized version of the MPV player [42] was used for the passive subjective evaluation.

All of the videos were run through a randomized playback and scoring script developed within MATLAB version 2019b. At first, the observers were asked to complete a training session in order to familiarize them with the artifacts under assessment and the evaluation process. After finishing the playback, observers provided their rating scores. Both the original and processed LF (video) sequences were simultaneously presented to the observer. The visual quality of the processed video sequences was scored with respect to the original video sequences, which were clearly identified as such. A total of 236 scores were thus obtained per evaluation session: each subject assessed 38 test models degraded

by compression at 3 quality levels, plus hidden references. Outlier detection based on the ITU-R Recommendation BT.500-13 [6] was applied to the quality scores prior to analysis; however, no outliers were found. All scores are included in the analysis presented below, with 95% Confidence Intervals (CIs) on the Mean Opinion Scores (MOS).

3.3. LF Dataset Results and Analysis

Our subjective experiments were conducted at the VNU University of Engineering and Technology (VNU-UET) campus in Hanoi, Vietnam. There were 20 participants in total joining the test, 8 females and 12 males between the ages of 19 and 25 with normal or corrected to normal vision.

According to the LF dataset definition in Table 1, the dataset can be categorized according to three scene rendering conditions: complexity (simple and complex), depth (narrow and wide) and focus (background and foreground).

In our first test, we evaluated whether both the LF content and rendering affected the subjective judgements in the experiments. In other words, we tested whether the mean values of the MOS scores were different for the various conditions (i.e., CWF, CWB, SNF, SNB, etc., as per Table 1) and varying LF contents.

At this stage of the analysis, we were interested in the hypothesis that the rendering condition would affect the observer judgements and not interested in creating a predictive model of image quality. As such, the bitrates of LF data derived from the VVC codec were used as a proxy for quantifying the quality degradation caused by differing compression levels. In other words, at this stage of the investigation, we examined the dependence of content and rendering conditions on the MOS scores. We chose to model this dependence as a factor that would globally increase or decrease MOS scores dependent only on the content and rendering condition common to both the reference and reconstructed LFs, not dependent on the perceived quality difference between the reference and reconstructed LFs. Bitrate used in this manner is not a suitable predictor of LF quality, as it is dependent on the specific compression method used. However, the goal at this stage of the analysis is to examine the nature of the content dependence, not to create a predictive model. In Section 4, a predictive model is developed. The quality difference between the reference and reconstructed LFs was modeled as a linear function of bitrate. In Figure 5, the line of best fit for the MOS versus bitrate graph between different conditions is presented.

As observed in Figure 5, the slopes and intercepts of the lines of best fit to the MOS versus bitrate data are similar across different conditions, which implies that the conditions of the LF image (CNB, CNF and so forth) have little effect on the overall MOS score. In addition, Figure 5 implies that the rendering condition does not have a strong effect on the slope of the MOS versus bitrate relationship.

The Analysis of Variance (ANOVA) [43] results are shown in Table 2. In the ANOVA analysis, the degree of freedom (Df), the sum of squares (Sum Sq), the mean of squares (Mean Sq), the F-test (F-value) and the probability p -value (Pr) are reported. Importantly, we examine the p -value of the F statistic, where the null hypothesis is rejected if the p -value is lower than the threshold of 0.05.

Table 2. ANOVA analysis for the MOS scores with various conditions.

Factor	Df	Sum Sq	Mean Sq	F Value	Pr (>F)
Complexity	1	1.0	0.9654	0.666	0.415
Focus	1	0.3	0.3095	0.214	0.644
Depth	1	0.3	0.3293	0.227	0.634
Residuals	224	324.5	1.4488		

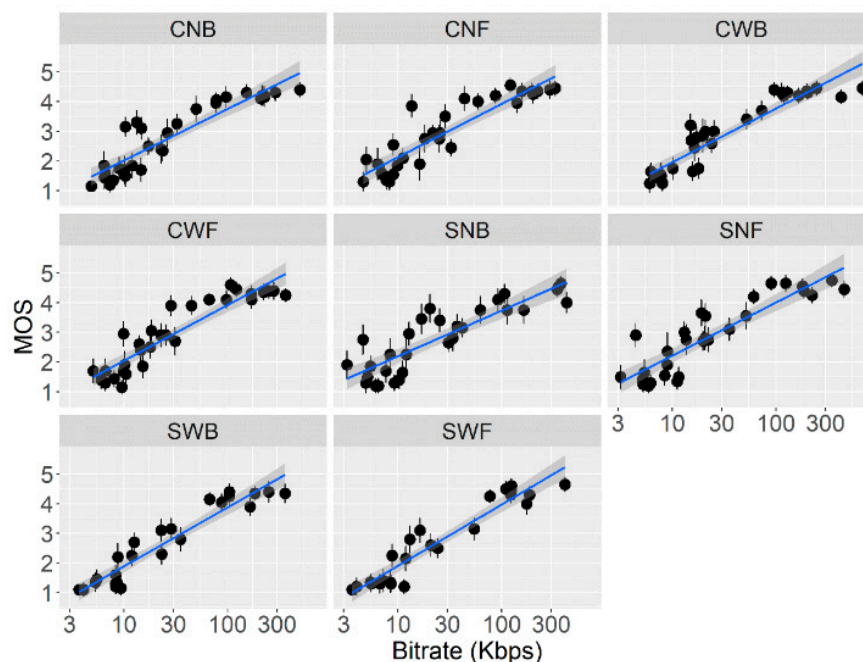


Figure 5. MOS score versus Bitrate for various conditions. The black dots in the scatter plots represent the data collected for each of the LF rendering conditions as defined in Table 1 such that each black dot presents a paired measurement of two variables (i.e., Bitrate and MOS), while the blue lines are the lines of best fit for each set of LF rendering conditions.

In Table 2, no rendering condition factor had a statistically significant effect on the MOS level, and, therefore, we can conclude that there is no effect on MOS value average due to the different conditions. In other words, the experimental conditions of complexity, depth and focus do not have an effect to shift MOS values up or down, so the LF rendering conditions do not need to be explicitly accounted for in the proposed model discussed further in the next section.

Next, we sought to examine the effect of LF content on the MOS scores. We started by considering the rendering of the LF content based on complexity (i.e., complex content and simple content). Although the analysis above showed no significant difference in mean MOS score due to differences in this condition, we sought to examine whether the spread of MOS scores was affected by the complexity of the LF content. Figure 6 shows the distribution of MOS scores for complex and simple content at different coding distortion levels, QP50 and QP22. Moreover, the well-known objective quality metric SSIM, which takes into account some image features (i.e., luminance, contrast and similarity), is also used in measuring the variation of SSIM with the LF contents (Figure 7).

As Figure 5 indicates, the rendering condition does not have a strong effect on the slope of the MOS versus bitrate relationship, but it shows variation in the quality degradation as a result of increasing or decreasing bitrate. In the case of the lower bitrate (or the poorer quality case), it is noticed that the MOS scores of simple contents (i.e., SWB and SWF) are slightly lower compared to the complex contents (i.e., CWB and CWF). Thus, to explore in detail this scenario, Figure 6 presents an alternative rendering of the distribution of MOS scores based on the different LF contents grouped by rendering conditions. LF images with simple and complex contents perform similarly in the case of good image quality, but this is not the case for poorer image quality. Notably, in the case of good quality, the complex contents provide similar MOS scores with slightly lower MOS scores compared to the simple contents. This effect may be explained by the fact that the observers more easily recognize the difference in the simple content to give a judgement compared to the complex content. However, this result changes in the case of poor quality LFs because the observers have difficulty recognizing the difference between the reference and distorted

4D-LF pseudo-sequences in the complex content images compared to the simple content images. In other words, the observers tended to give lower scores to the simple content in the case of poor quality because of the ease in recognizing the difference between the reference and distorted examples. Thus, it can be said that the image content has some effect on the observer scores; however, this is just a qualitative observation. To further explore these qualitative findings, the objective quality score SSIM is evaluated in Figure 7; interestingly, it illustrates an opposite variation of LF contents at both of the distortion levels compared to subjective judgement. This result can be explained by the SSIM metric just comparing mathematical similarity between the reference and distorted images. Hence, the simple content provides higher scores than the complex content in the good quality case due to less information being in the simple image content, while both image contents exhibit a similar score in poor quality compression because the content looks the same in this case. This means that the LF content affects the performance of the SSIM objective quality metric in a manner that is not consistent with subjective quality judgements by human subjects.

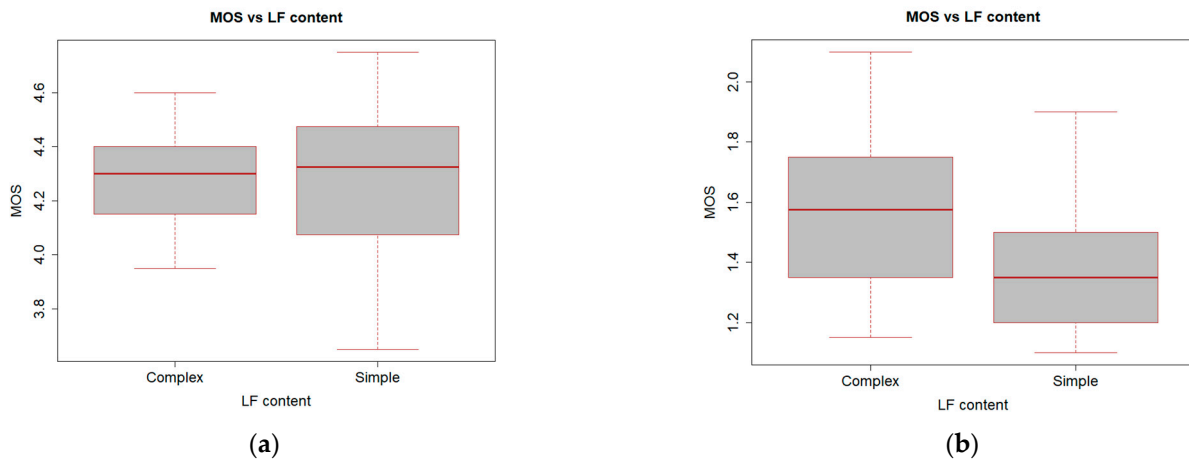


Figure 6. The variation of MOS scores and LF content with different distortion levels: (a) good quality case (QP22), (b) poor quality case (QP50).

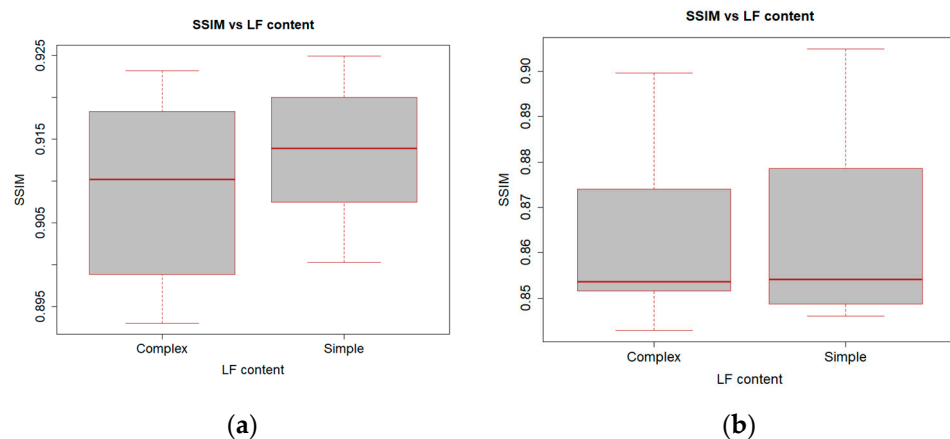


Figure 7. The variation of SSIM and LF contents with different distortion levels: (a) good quality case (QP22), (b) poor quality case (QP50).

To further analyze the effect of LF content on the MOS scores, a model fitting analysis approach was taken. Two models were compared. The first model is a Generalized Least

Squares (GLS) [44] technique that assumes that the MOS scores can be modelled as a simple non-linear function of bitrate:

$$Y_{BasicModel} = \alpha Bitrate + \rho \ln Bitrate + \beta + \epsilon \quad (2)$$

where a logarithmic function of bitrate is introduced to account for non-linearity. The variables α , ρ and β represent the coefficients of the fitted function, and ϵ is a residual error term. This model assumes no content dependence between individual LFs on the MOS scores.

The second model is a mixed effect model [45] that introduces a random intercept effect to the model:

$$Y_{MixedEffectModel} = \alpha Bitrate + \rho \ln Bitrate + \gamma Z + \beta + \epsilon \quad (3)$$

where Z and γ are the model matrices and coefficient array for the random effects component of the observations, respectively. For this model, the random effect is the content of the original LF. This produces a model wherein the intercept of the model is allowed to vary independently according to each individual LF, but the functional form of the model is assumed to be common to all LFs. In this way, we can separate and independently study the effect of LF content on the MOS scores, distinct from the effect of compression levels. Note that the above model yields a single value of α and β but one value of γ for each reference light field in the data set, indicating whether that light field elevated or lowered overall MOS scores. It is also of note that, although Equation (3) is not a predictive model, it serves to validate or exclude a specific form of the LF quality model hypothesis—that is, the hypothesis that the intercept of the quality model is dependent on specific LF content and not the degradation applied to the content, while the functional form of the MOS versus bitrate relationship is dependent only on the degradation applied to the content and not the content itself. Table 3 shows an ANOVA analysis of the two models (i.e., BasicModel and MixedEffectModel). The component parameters from this table, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), are two established mathematical methods for evaluating and scoring the fitting of data models. The lower the values of AIC and BIC are, the better the model fit. The term logLik stands for the Log-Likelihood densities of the entire dataset under each model with higher values of logLik, indicating the better a model fits a dataset. Finally, the likelihood ratio test (L. Ratio) is the assessment of the goodness of fit of two compared statistical models as determined by the ANOVA analysis.

Table 3. ANOVA analysis to compare the models.

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-Value
BasicModel	4	269.9303	283.6477	−130.96517			
MixedEffectModel	5	195.1769	212.3236	−92.58845	1 vs. 2	76.75344	<0.0001

Based on the ANOVA analysis in Table 3, the mixed effect model allowing the intercept value to be adjusted separately for each reference LF is a more accurate fit to the MOS data than the basic model. Although the mixed effect model is not a predictive model, it does reinforce the conclusion that content dependence in the results can be modelled by a model with a fixed functional form whose zero point is adjusted according to LF content. The statistical goodness of fit measure R-squared (R2) [46] is used to compare the correlation results with 93% and 97% for the basic model and mixed effect model, respectively. It was found that 50.4% of the variance of the mixed effected model was explained by the random effect (LF content), indicating that accounting for LF content dependence has the potential to greatly improve existing LF quality metrics.

Thus, it can be concluded that the LF content, as represented by the reference LF alone, has an effect on the MOS scores and can be modelled as a shift in the mean of a quality

model where the shape of the quality model is determined by the difference between the reference and degraded (or processed) LF. The exact manner in which to quantify the LF content and the coefficients of such a model remains to be determined and is the focus of the next section.

4. Proposed Objective LF Quality Metric

4.1. Adaptive Content Based LF-IQA Model

Our proposed model is based on the addition of features relating to image content to adjust the performance of the objective quality metrics and the subjective judgements. According to the analysis in the previous section, the proposed model is as follows:

$$Obj_{new} = \alpha Obj_{old} + \gamma f_{content} + \epsilon \quad (4)$$

where Obj_{old} represents a conventional objective metric such as PSNR, SSIM, IW-PSNR, IW-SSIM, IW-MSE, Visual Information Fidelity (VIF), etc., and $f_{content}$ is the function of features representing the LF content. The Obj_{old} component of the model represents the difference in quality between the reference and degraded model, and $f_{content}$ represents an offset in the output of the quality dependent only on the reference light field to account for content dependence in the model.

As a first step to develop the proposed model, the content feature $f_{content}$ needs to be defined, as discussed below.

4.2. Feature Analysis

We initially need to determine the suitable content features for LFs as potential candidates for the $f_{content}$ component of the model. Thus, the feature selection procedure proposed is provided below:

- Step 1: Select a set of potential content descriptors such as Haralick features, DCT features and corner features using the FAST method.
- 1. Step 2: Compare the performance of the LF content features in terms of their ability to model the random effect intercept values, γ , as determined by the mixed effect model in Equation (3).
- Step 3: Identify the appropriate feature and generate the feature function.

Step 1:

In step 1, we firstly select the LF content features to work with. In this paper, three sources of content descriptive features are compared with the random effect function intercept values determined by the mixed effect model in Equation (3). The first set of features we considered were Haralick textural features [19], e.g., energy, contrast, correlation, homogeneity, entropy, difference entropy, etc. The second set were derived from the Discrete Cosine Transform (DCT) [20] as DCT features, i.e., Mean of Alternating Current (AC) coefficients (MAC), Mean of Direct Current (DC) coefficients (MDC), Variance of DC coefficients (VDC) and Variance of AC coefficients (VAC). The final set were corner features arising from the FAST method [21]. All features are extracted from the original (reference) LF image contents. The measures were computed from the SAIs of the 4D-LF pseudo-sequence and then averaged.

Regarding feature computation, the Haralick features are computed by selecting the first direction offset in the Gray Level Co-occurrence Matrix (GLCM), i.e., offset = [0, D], where D represents the distance from the pixel of interest [19]. Meanwhile the DCT features extract texture information from each of the SAIs using the DC and AC elements of the DCT. The DC coefficients represent the average energy of the image, while the AC coefficients contain higher frequency information to capture variation in the image spatial structure. The equations are defined as follows:

$$VDC = \sigma_{DC}^2 \quad (5)$$

where σ_{DC}^2 is the variance of the DC coefficient value of the SAIs in a 4D-LF pseudo-sequence.

$$VAC = 1/N \sum_{i=1}^N \sigma_{AC}^2(i) \quad (6)$$

where $\sigma_{AC}^2(i)$ is the variance of the AC coefficient value of the SAIs in a 4D-LF pseudo-sequence.

$$MDC = 1/N \sum_{i=1}^N DC(i) \quad (7)$$

where $DC(i)$ is the DC coefficient value of the SAI i in a 4D-LF pseudo-sequence.

$$MAC = 1/N \sum_{i=1}^N \sum_{k=1}^{H.W-1} AC_k(i) \quad (8)$$

where $AC_k(i)$ is the k th value of the AC coefficient in the SAI i .

Lastly, the corner feature is detected following the algorithm using the detectFASTFeatures [21] function of MATLAB.

Step 2:

Based on the results from the feature computation, we perform a performance comparison of individual features as measured against the random effect intercept values of the mixed effect model in Equation (3). Table 4 presents the feature performance through various statistical methods: Kendall's Rank Correlation Coefficient (KRCC), Spearman Rank Correlation Coefficient (SRCC) [47], Pearson Linear Correlation Coefficient (PLCC) [48], Mean Absolute Error (MAE), Root Mean Square Error (RMSE) [49] and R-squared (R2) [46]. The better the correlation of the texture features with the random effect intercept values (γ), the greater the potential for these features to describe the content dependence of the light fields in the final model. The KRCC and SRCC evaluate and predict monotonicity, PLCC evaluates the prediction accuracy, MAE and RMSE evaluate the prediction consistency and R2 evaluates the quality of the fit to a model. Higher KRCC, SRCC, PLCC and R2 values and lower MAE and RMSE values indicate better correlation in the range, with values from 0 to 1. To present the performance of the statistical scores clearly, we used the min-max normalization technique [50] for each feature value and then averaged them in the normalization column in Table 4.

From Table 4, it can be seen that the Haralick features demonstrate outstanding performance, and this feature can be utilized for modelling content dependence in our final proposed model. The difference entropy feature presents a stable and good performance on all statistical measures compared to other potential features such as homogeneity, contrast, entropy and IMC_I. Thus, the difference entropy feature can be identified as an appropriate feature to use in our model to represent how content dependence affects MOS.

Step 3:

Based on the feature performance comparison in Step 3, the most appropriate feature to affect the MOS is the difference entropy feature. It is assigned as $f_{content}$ in the proposed model and computed by the equation below

$$f_{content} = - \sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\} \quad (9)$$

where N_g is the number of distinct gray levels in the quantized image and p_x is the i th entry in the marginal probability matrix obtained by summing the rows of $p(i, j)$, $\sum_{j=1}^{N_g} p(i, j)$.

Table 4. Performance of individual features versus mixed model random effort intercept values.

Features	Sources	KRCC	SRCC	PLCC	R2	MAE	RMSE	Normalization
Homogeneity	Haralick Features [18]	0.54	0.73	0.65	0.42	0.16	0.22	0.92
Correlation		0.21	0.29	0.17	0.03	0.20	0.29	0.26
Contrast		0.51	0.72	0.61	0.37	0.16	0.23	0.86
Energy		0.32	0.46	0.60	0.37	0.17	0.23	0.74
Variance		0.24	0.33	0.40	0.16	0.20	0.27	0.42
Entropy		0.40	0.56	0.64	0.41	0.17	0.22	0.82
Sum_Variance		0.20	0.29	0.34	0.12	0.20	0.27	0.37
Sum_Average		0.23	0.32	0.41	0.17	0.19	0.27	0.45
Sum_Entropy		0.36	0.50	0.60	0.36	0.17	0.23	0.76
Difference_Variance		0.51	0.71	0.59	0.35	0.17	0.24	0.80
Difference_Entropy		0.53	0.72	0.70	0.48	0.15	0.21	1.00
Information_Measure_of_Correlation_I (IMC_I)		0.50	0.70	0.60	0.36	0.17	0.23	0.82
Information_Measure_of_Correlation_II (IMC_II)		−0.08	−0.10	0.15	0.02	0.21	0.29	0.09
Maximal_Correlation_Coefficient		0.18	0.24	0.11	0.01	0.20	0.29	0.22
VAC	DCT features [19]	0.24	0.35	0.41	0.17	0.20	0.27	0.42
VDC		−0.20	−0.26	0.03	0.00	0.21	0.29	0.00
MAC		0.21	0.33	0.33	0.11	0.21	0.28	0.33
MDC		0.26	0.37	0.31	0.10	0.21	0.28	0.33
Corner feature	FAST algorithm [20]	0.50	0.68	0.51	0.26	0.18	0.25	0.69

5. Experimental Evaluation and Results

5.1. Testing Conditions

To evaluate our proposed quality metric, we use a combined passive and partially interactive method based on the DSIS testing method to obtain the subjective judgements. All the LF image data were selected from the EPFL dataset and distorted using the VVC codec. The resolution of the dataset is $625 \times 434 \times 15 \times 15$ but was converted to $624 \times 432 \times 101$ to visualize using the PVS method on 2D devices, removing the vignetting effect from LF generation. The total number of LFs evaluated is thus 236 distortion samples selected from 38 original LF samples.

Regarding experimental settings, linear and logistic functions are used to map the objective quality scores. Then, the consistency between the subjective scores and the mapped quality scores is measured by statistical methods, i.e., KRCC, SRCC, PLCC, MAE, RMSE and R2. In the next sub-sections, the proposed method is compared with the objective quality metrics. Due to the light-based sensitivity of the human visual system, all of the objective metrics are converted to YUV color space, and the luminance value Y is selected before comparison. Additionally, the metric scores are computed as averages across the SAIs of the LF image.

5.2. Performance Evaluation Using Objective Quality Metrics

Regarding quantitative comparison, the performance comparison results of the original objective quality metrics on the LF dataset are summarized in Table 5. The results present the correlation between the MOS score and the original objective quality metrics and are computed by the basic model described in Equation (2), with bitrate substituted with the original objective quality metrics. In Table 5, the metrics are reported individually on the whole dataset, and the key observations are as follows:

- The GMSD metric provides the best performance compared to the other objective metrics on our dataset because it combines the pixel-wise Gradient Magnitude Similarity (GMS) and the standard deviation of the GMS map to predict image quality more accurately, while the IW-MSE, IW-PSNR and PSNR rank 2nd, 3rd and 4th and measure the difference of images based on a Sum of Absolute Differences (SAD) approach. Notably, the PSNR metric shows a better performance compared to SSIM, which is more correlated with the quality perception of the human visual system than the PSNR metric. In other words, it can be said that the PSNR is a more suitable metric for prediction than SSIM in the case of subjective judgement in YUV color space.
- The JPEG Quality Score (JQS) presents the lowest performance since it is more suitable for still images than the PVS. The VIF and Information Fidelity Criterion (IFC) performs similarly to the metrics at the 3rd and 4th rank (i.e., IW-PSNR and PSNR). This may be due to the characteristics of these metrics, as VIF and IFC are based on the natural scene statistics, which are well represented in our dataset.

Table 5. Performance comparison of objective quality metrics on the LF dataset.

Distortion Type	Criteria	A	B	C	D	E	F	G	H	I	J	K	L	M
		PSNR	SSIM	MS-SSIM	FSIM	IW-SSIM	IW-PSNR	IW-MSE	JQS	UQI	IFC	VIF	VIFP	GMSD
VVC codec (QP22, QP40, QP50)	KRCC	0.6328	0.5583	0.5628	0.5935	0.5859	0.6351	0.6357	-0.2183	0.5080	0.6286	0.6217	0.5996	0.7081
	SRCC	0.8509	0.7659	0.7737	0.7985	0.7944	0.8569	0.8556	-0.2727	0.7068	0.8333	0.8269	0.8056	0.8952
	PLCC	0.8661	0.7761	0.7780	0.8210	0.8072	0.8687	0.8737	0.0664	0.7068	0.8184	0.8362	0.8246	0.8884
	R2	0.7502	0.6023	0.6053	0.6741	0.6515	0.7546	0.7633	0.0044	0.4996	0.6698	0.6993	0.6799	0.7893
	RMSE	0.5977	0.7541	0.7513	0.6827	0.7059	0.5923	0.5818	1.1933	0.8459	0.6871	0.6558	0.6766	0.5489
	MAE	0.4797	0.5690	0.5750	0.5222	0.5447	0.4728	0.4499	1.0569	0.6646	0.5782	0.5318	0.5340	0.4387

For qualitative comparison, Figure 8 shows the scatter plots of all compared metrics to visually illustrate the prediction ability of the objective quality metrics evaluated. The plots show similar results, as described in Table 5, but it can be said that all objective quality metrics are individually not good predictors for the LF dataset, as indicated by the wide spread of metric values about the fitted lines. To elaborate, the black dots in the scatter plots represent a paired measurement of two variables (i.e., Objective quality metric and MOS), while the blue lines are the line of best fit of each objective quality metric. If the distance between the black dots and the blue lines are far away or non-linear, it means that the variables are weakly correlated or not correlated. It also follows that if the variables are strongly correlated, the black dots should cluster close to the blue line. In other words, the predicted performance of a model with a tight scatter of data about the line of best fit will be superior to a model with a wider scatter.

5.3. Performance Evaluation Using the Proposed Model

In this section, the performance of the proposed model is described based on the same statistical techniques on all objective metrics as in the previous section. The proposed model is the objective measure with the addition of $f_{content}$. Figure 9 illustrates that the performance of our proposed model measured by PLCC and R2 consistently improved by about 0.05% and 0.06%, respectively. The exception is the significant improvement of JQS (0.4% improvement), which performs well only for still images, i.e., JPEG images. Therefore, to have a fair comparison, JQS is removed from Figure 9 so as to not affect the presentation of the other results. For brevity, the top four of the proposed objective metrics are presented in Figure 10. As explained in the previous sub-section, the tight scatter of the metrics about the line of best fit shown in Figure 10 indicates better performance compared to the wider scatter of the data in Figure 8. So, the results in Figure 10 show tighter scatter plots compared to the four best original metrics, i.e., PSNR, IW-PSNR, IW_MSE and GMSD. Once again, the proposed model applied to GMSD remains the most improved by 92% correlation on the PLCC method. It can be said that the textural features play an important role in improving the predicted performance of the proposed model with the LF image.

The information about the spatial arrangement of color or the intensities or selected region of the LF image are exploited to provide accurate results.

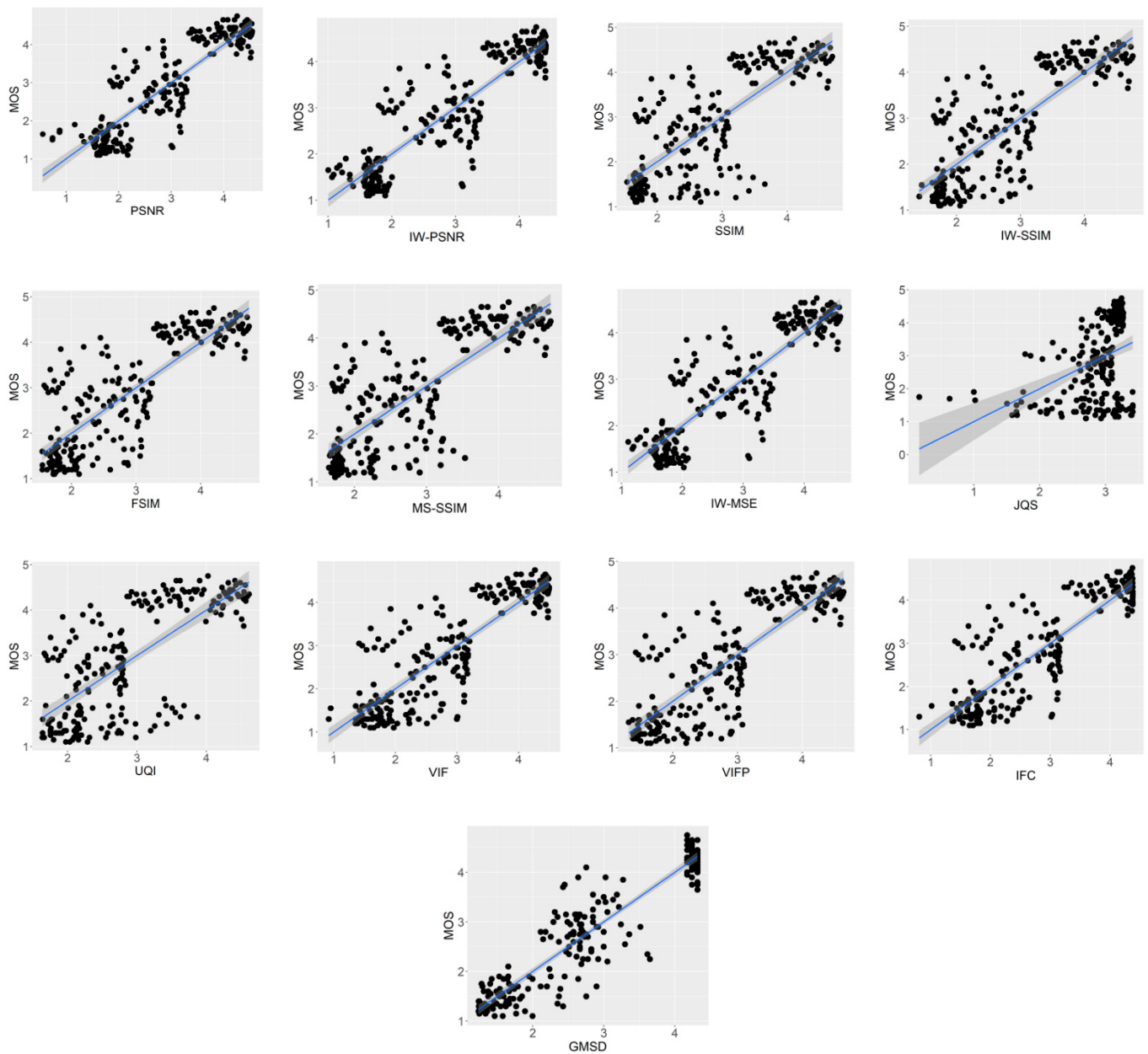


Figure 8. Scatter plots of all compared objective metrics on the LF dataset. The black dots in the scatter plots represent a paired measurement of two variables (i.e., Objective quality metric and MOS), while the blue lines are the fitted linear function on the LF dataset of each objective quality metric.

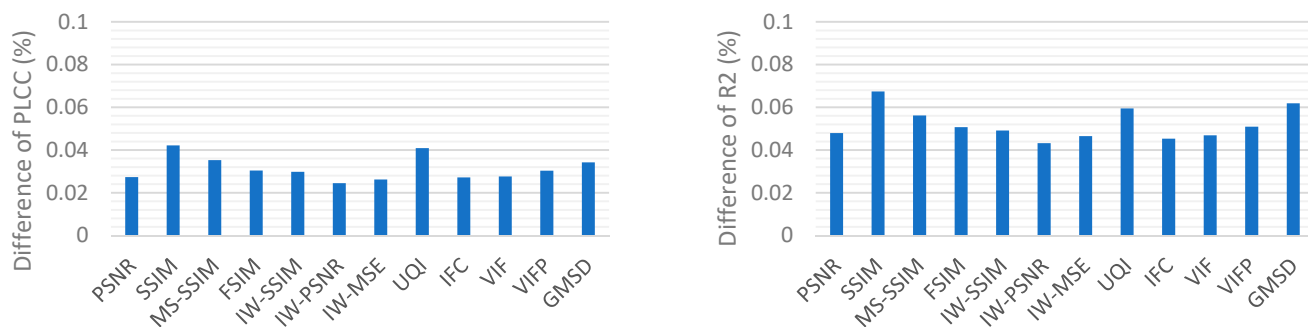


Figure 9. The difference between the PLCC and R2 performance of compared quality metrics with proposed metrics.

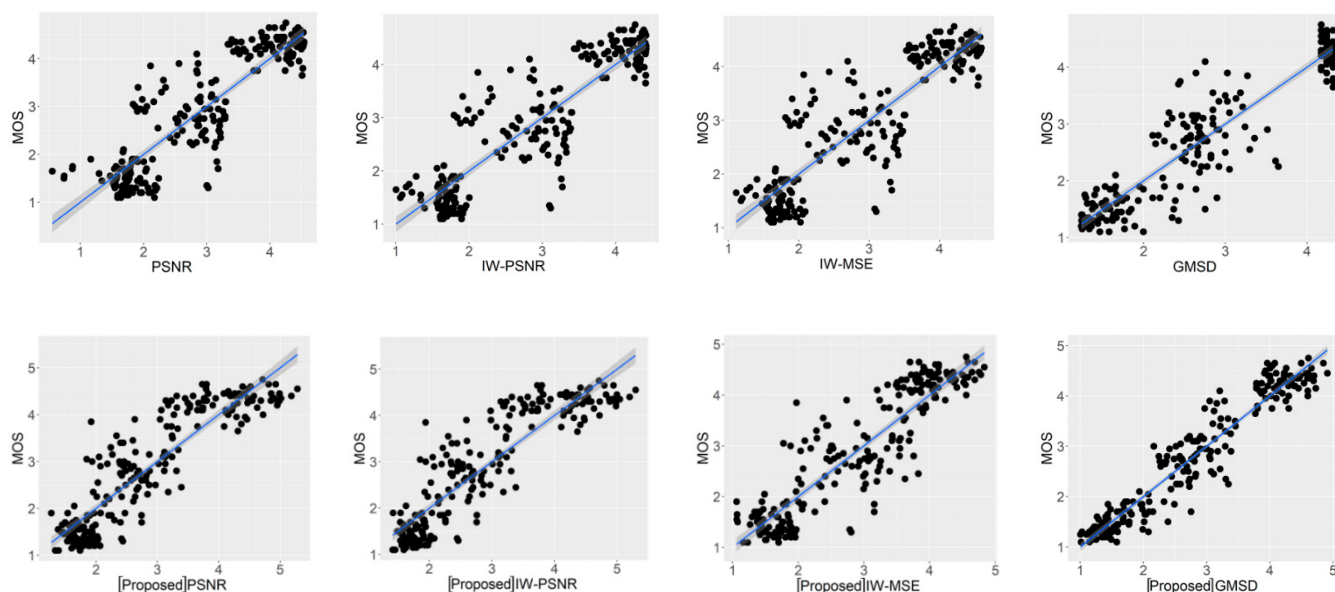


Figure 10. The comparison between the best four original fitted models and the proposed fitted models on the LF dataset. The black dots in the scatter plots represent a paired measurement of two variables (i.e., Objective quality metric and MOS), while the blue lines are the fitted linear function on the LF dataset of each objective quality metric. The top row presents the original metrics, and the bottom row presents the proposed model applying to the following objective metrics.

5.4. Further Evaluation of LF Content on the Proposed Method

Based on Table 4, we noticed that the Homogeneity feature performed similarly to Difference_Entropy. Therefore, it can be assumed that the combination of these two features may improve the performance of proposed model. To understand in more depth the performance of the proposed method, we evaluated an extended version of our proposed model:

- Opt. 1: Evaluating the proposed model only (Metric + Difference Entropy)
- Opt. 2: Evaluating the proposed model combined with the Homogeneity feature (Metric + Difference Entropy + Homogeneity)

To clarify, both extended version models are computed following Equation (4), but in the case of Opt. 1, the $f_{content}$ term includes only the Difference Entropy feature, while the $f_{content}$ includes the combination of Difference Entropy and Homogeneity features in Opt. 2.

Table 6 shows the performance comparison between the proposed model Opt. 1 and the proposed model Opt. 2 based on the top four of the proposed objective metrics, i.e., GMSD, PSNR, IW-PSNR and IW-MSE. According to Table 6, the proposed model Opt.

2 by the combination of features Homogeneity and Difference Entropy presents better performance in most criteria compared to the proposed model Opt. 1, i.e., GMSD, PSNR and IW-MSE. However, it is noticed that the metric IW-PSNR in the Opt.1 model gives different results, with 4 out of 6 criteria performing slightly better than the Opt.2 model. This difference in performance results from the different algorithms of the objective metrics. GMSD is based on the combination of the pixel-wise GMS and the standard deviation of the GMS map to predict the result, while PSNR, IW-PSNR and IW-MSE measure the difference of images based on a SAD approach. Therefore, Opt. 2 can be a potential model for some circumstances, depending on the objective metric used, e.g., GMSD, PSNR and IW-MSE. The use of Opt. 2 requires consideration of the complexity of the model due to the combination of content features. There is a tradeoff between the predicted performance and the computation complexity. Currently, the PSNR metric is considered to be the fastest computation metric due to its simple algorithm. Therefore, the complexity analysis of objective quality metrics may be the subject of prospective research in the future.

Table 6. Performance of the proposed model in different options.

Metrics	Criteria	Opt. 1	Opt. 2
GMSD	KRCC	0.7870	0.7849
	SRCC	0.9415	0.9411
	PLCC	0.9213	0.9226
	R2	0.8488	0.8512
	RMSE	0.4650	0.4613
	MAE	0.3711	0.3686
PSNR	KRCC	0.6880	0.6882
	SRCC	0.8879	0.8886
	PLCC	0.8934	0.8939
	R2	0.7981	0.7991
	RMSE	0.5374	0.5361
	MAE	0.4268	0.4292
IW-PSNR	KRCC	0.6994	0.6983
	SRCC	0.8943	0.8941
	PLCC	0.8932	0.8935
	R2	0.7978	0.7984
	RMSE	0.5378	0.5370
	MAE	0.4287	0.4311
IW-MSE	KRCC	0.6827	0.6840
	SRCC	0.8856	0.8868
	PLCC	0.8999	0.9016
	R2	0.8098	0.8129
	RMSE	0.5216	0.5173
	MAE	0.4109	0.4049

6. Conclusions

In this paper, we propose improved objective LF quality assessment approaches, which augment existing objective quality metrics to better account for content dependence within the light field content. In addition, a novel textural feature analysis that considers a variety of conditions for LF image content is provided. Taking into consideration the many aspects of LF subjective judgement, we select an efficient subjective LF dataset, which separates content according to the depth, focus and complexity of subjects within the LF image. The proposed solution shows an improvement in the performance of existing objective metrics and identifies the GMSD metric as the most suitable objective metric for LF images due to

the advantage of the combination between pixel-wise GMS and the standard deviation of the GMS map.

The proposed solution is one of few image quality assessment approaches specifically developed for LF data. There are many LF-IQA metrics explored in the literature that provide more accuracy and better performance than conventional 2D IQA metrics. However, the 2D IQA metrics such as SSIM or PSNR exhibit advantages such as very fast computation and easy implementation. Thus, the LF content-adaptive metric model provides an improved objective quality metric for future use in LF applications.

In our future work, the LF surface information is needed to study the effect of light reflection on the display quality of the LF image. Additionally, a wide range of LF images should be evaluated, including synthesized LF datasets and natural LF datasets.

Author Contributions: Conceptualization, H.P., S.P., E.C. and X.H.; methodology, H.P., S.P., E.C. and X.H.; software, H.P., S.P.; validation, H.P., S.P., E.C. and X.H.; formal analysis, H.P., S.P., E.C. and X.H.; investigation, H.P., S.P.; resources, H.P., S.P.; data curation, H.P., S.P.; writing—original draft preparation, H.P.; writing—review and editing, H.P., S.P., E.C. and X.H.; visualization, H.P., S.P.; supervision, H.P., S.P., E.C. and X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2020.15.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to reasons such as (i) the study does not relate to any medical aspects (e.g., drug-device, biologic, etc.) and human rights, (ii) the participants are students from VNU-UET, and they all agree with the study by signing in the list of the agreement form, (iii) the study is conducted at the VNU-UET lab with the approved environment.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: This work has been supported in part by the Joint Technology and Innovation Research Centre, a partnership between the University of Technology Sydney and Vietnam National University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bedard, N.; Shope, T.; Hoberman, A.; Haralam, M.A.; Shaikh, N.; Kovačević, J.; Balram, N.; Tošić, I. Light field otoscope design for 3D in vivo imaging of the middle ear. *Biomed. Opt. Express* **2017**, *8*, 260–272. [[CrossRef](#)] [[PubMed](#)]
2. Li, H.; Guo, C.; Jia, S. High-Resolution Light-Field Microscopy. *Front. Opt.* **2017**, *FW6D*, 3. [[CrossRef](#)]
3. Tsai, D.; Dansereau, D.G.; Peynot, T.; Corke, P. Image-Based Visual Servoing with Light Field Cameras. *IEEE Robot. Autom. Lett.* **2017**, *2*, 912–919. [[CrossRef](#)]
4. Lynch, K.; Fahringer, T.; Thurow, B. Three-Dimensional Particle Image Velocimetry Using a Plenoptic Camera. In Proceedings of the 50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Nashville, TN, USA, 9–12 January 2012.
5. Wu, G.; Masiá, B.; Jarabo, A.; Zhang, Y.; Wang, L.; Dai, Q.; Chai, T.; Liu, Y. Light Field Image Processing: An Overview. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 926–954. [[CrossRef](#)]
6. International Telecommunications Union. *Methodology for the Subjective Assessment of the Quality of Television Pictures*; ITU-R BT.500-13; International Telecommunications Union: Geneva, Switzerland, 2012.
7. Viola, I.; Rerabek, M.; Bruylants, T.; Schelkens, P.; Pereira, F.; Ebrahimi, T. Objective and Subjective Evaluation of Light Field Image Compression Algorithms. In Proceedings of the 2016 Picture Coding Symposium (PCS), Nuremberg, Germany, 4–7 December 2016; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2016; pp. 1–5.
8. Viola, I.; Řeřábek, M.; Ebrahimi, T. Impact of Interactivity on the Assessment of Quality of Experience for Light Field Content. In Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017.
9. Viola, I.; Ebrahimi, T. VALID: Visual quality Assessment for Light field Images Dataset. In Proceedings of the 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), Sardinia, Italy, 29 May–1 June 2018; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018; pp. 1–3.
10. Paudyal, P.; Battisti, F.; Sjöström, M.; Olsson, R.; Carli, M. Towards the Perceptual Quality Evaluation of Compressed Light Field Images. *IEEE Trans. Broadcast.* **2017**, *63*, 507–522. [[CrossRef](#)]

11. Shi, L.; Zhao, S.; Zhou, W.; Chen, Z. Perceptual Evaluation of Light Field Image. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018; pp. 41–45.
12. Luo, Z.; Zhou, W.; Shi, L.; Chen, Z. No-Reference Light Field Image Quality Assessment Based on Micro-Lens Image. In Proceedings of the 2019 Picture Coding Symposium (PCS), Ningbo, China, 12–15 November 2019; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2019; pp. 1–5.
13. Zhou, W.; Shi, L.; Chen, Z.; Zhang, J. Tensor Oriented No-Reference Light Field Image Quality Assessment. *IEEE Trans. Image Process.* **2020**, *29*, 4070–4084. [[CrossRef](#)]
14. Tian, Y.; Zeng, H.; Hou, J.; Chen, J.; Ma, K.-K. Light Field Image Quality Assessment via the Light Field Coherence. *IEEE Trans. Image Process.* **2020**, *29*, 7945–7956. [[CrossRef](#)]
15. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
16. Pereira, F.; Pagliari, C.; Silva, E.D.; Tabus, I.; Amirpour, H.; Bernardo, M.; Pinheiro, A. JPEG Pleno Light Field Coding Common Test Conditions V3.3. In Proceedings of the JPEG Meeting, Brussels, Belgium, 2019; Document ISO/IEC JTC 1/SC29/WG1N84025. Available online: https://ds.jpeg.org/documents/jpegpleno/wg1n84049-CTQ-JPEG_Pleno_Light_Field_Common_Test_Conditions_v3_3.pdf (accessed on 26 February 2022).
17. Teratani, M.; Jin, X. How to Achieve Dense Light Field Video Compression? 2020. Available online: <https://mpeg.chiariglione.org/> (accessed on 5 December 2021).
18. Řeřábek, M.; Ebrahimi, T. New Light Field Image Dataset. In Proceedings of the 8th International Conference on Quality of Multimedia Experience, Lisbon, Portugal, 6 June 2016.
19. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [[CrossRef](#)]
20. Bae, H.-J.; Jung, S.-H. Image retrieval using texture based on DCT. In Proceedings of the ICICS, 1997 International Conference on Information, Communications and Signal Processing, Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat. No.97TH8237), Singapore, 12 September 1997.
21. Rosten, E.; Drummond, T. Fusing points and lines for high performance tracking. In Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 17–20 October 2005; pp. 1508–1515.
22. Paudyal, P.; Battisti, F.; Carli, M. Effect of visualization techniques on subjective quality of light field images. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2017; pp. 196–200.
23. Paudyal, P.; Battisti, F.; Carli, M. Reduced Reference Quality Assessment of Light Field Images. *IEEE Trans. Broadcast.* **2019**, *65*, 152–165. [[CrossRef](#)]
24. Battisti, F.; Carli, M.; Le Callet, P. A Study on the Impact of Visualization Techniques on Light Field Perception. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018; pp. 2155–2159.
25. Viola, I.; Řeřábek, M.; Ebrahimi, T. A new approach to subjectively assess quality of plenoptic content. In *Applications of Digital Image Processing XXXIX, Proceedings of the SPIE 9971, San Diego, CA, USA, 27 September 2016*; SPIE: Bellingham, WA, USA, 2016. Available online: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9971/1/A-new-approach-to-subjectively-assess-quality-of-plenoptic-content/10.1117/12.2240279.short?SSO=1> (accessed on 12 January 2021).
26. Dansereau, D. Light Field Toolbox for Matlab, Feb. 2015. Available online: <http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4> (accessed on 12 January 2021).
27. Min, X.; Zhou, J.; Zhai, G.; Le Callet, P.; Yang, X.; Guan, X. A Metric for Light Field Reconstruction, Compression, and Display Quality Evaluation. *IEEE Trans. Image Process.* **2020**, *29*, 3790–3804. [[CrossRef](#)]
28. Tian, Y.; Zeng, H.; Hou, J.; Chen, J.; Zhu, J.; Ma, K.-K. A Light Field Image Quality Assessment Model Based on Symmetry and Depth Features. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *31*, 2046–2050. [[CrossRef](#)]
29. Shan, L.; An, P.; Meng, C.; Huang, X.; Yang, C.; Shen, L. A No-Reference Image Quality Assessment Metric by Multiple Characteristics of Light Field Images. *IEEE Access* **2019**, *7*, 127217–127229. [[CrossRef](#)]
30. Huang, H.; Zeng, H.; Tian, Y.; Chen, J.; Zhu, J.; Ma, K.-K. Light Field Image Quality Assessment: An Overview. In Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Shenzhen, China, 6–8 August 2020; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2020; pp. 348–353.
31. Shi, L.; Zhou, W.; Chen, Z.; Zhang, J. No-Reference Light Field Image Quality Assessment Based on Spatial-Angular Measurement. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 4114–4128. [[CrossRef](#)]
32. Cui, Y.; Yu, M.; Jiang, Z.; Peng, Z.; Chen, F. Blind light field image quality assessment by analyzing angular-spatial characteristics. *Digit. Signal Process.* **2021**, *117*, 103138. [[CrossRef](#)]
33. Zou, Z.; Liu, C.; Zhang, L.; Qiu, J. Light Field Quality Assessment Based on Aggregation of Visual Features in Spatio-angular Domains. In *OSA Optical Sensors and Sensing Congress 2021 (AIS, FTS, HISE, SENSORS, ES)*; The Optical Society: Washington, DC, USA, 2021; p. JTU5A.11.
34. Ravishankar, J.; Sharma, M.; Gopalakrishnan, P. A Flexible Coding Scheme Based on Block Krylov Subspace Approximation for Light Field Displays with Stacked Multiplicative Layers. *Sensors* **2021**, *21*, 4574. [[CrossRef](#)] [[PubMed](#)]

35. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [[CrossRef](#)]
36. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402. [[CrossRef](#)]
37. Wang, Z.; Li, Q. Information Content Weighting for Perceptual Image Quality Assessment. *IEEE Trans. Image Process.* **2011**, *20*, 1185–1198. [[CrossRef](#)]
38. Mahmoudpour, S.; Schelkens, P. On the performance of objective quality metrics for light fields. *Signal Process. Image Commun.* **2021**, *93*, 116179. [[CrossRef](#)]
39. Larson, E.C.; Chandler, D.M. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* **2010**, *19*, 011006.
40. Xue, W.; Zhang, L.; Mou, X.; Bovik, A. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index. *IEEE Trans. Image Process.* **2014**, *23*, 684–695. [[CrossRef](#)] [[PubMed](#)]
41. Benjamin, B.; Jianle, C.; Shan, L.; Ye-Kui, W. Versatile Video Coding (Draft 10). In Proceedings of the 19th Meeting: By Teleconference, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Virtual, 22 June–1 July 2020.
42. MPV Video Player. Available online: <https://mpv.io> (accessed on 19 March 2021).
43. Lars, S.; Svante, W. Analysis of Variance (ANOVA). *Chemom. Intell. Lab. Syst.* **1989**, *6*, 259–272.
44. Aitken, A.C. IV. —On Least Squares and Linear Combination of Observations. *Proc. Proc. R. Soc. Edinb.* **1936**, *55*, 42–48. [[CrossRef](#)]
45. Bolker, B.M.; Brooks, M.E.; Clark, C.J.; Geange, S.W.; Poulsen, J.R.; Stevens, M.H.H.; White, J.-S.S. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* **2009**, *24*, 127–135. [[CrossRef](#)] [[PubMed](#)]
46. Figueiredo, D.; Silva, J.A.; Rocha, E. What is R2 all about? *Leviathan-Cad. Pesqui. Política* **2011**, *3*, 60–68.
47. Sheikh, H.R.; Sabir, M.F.; Bovik, A. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. [[CrossRef](#)]
48. Kirch, W. (Ed.) *Encyclopedia of Public Health, Pearson's Correlation Coefficient*; Springer: Dordrecht, The Netherlands, 2008.
49. Wang, W.; Lu, Y. Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. In *IOP Conference Series: Materials, Science and Engineering*; IOP Publishing: Bristol, UK, 2018; Volume 324. [[CrossRef](#)]
50. Han, J.; Kamber, M.; Pei, J. *Data Mining Concepts and Techniques*, 3rd ed.; Morgan Kaufmann, Elsevier Inc.: Wyman, MA, USA, 2012.