Quality evaluation of light fields

Ali Ak, Patrick Le Callet

CHAPTER

Quality Evaluation of Light Fields

1

Ali Ak[a] and Patrick Le Callet[a]

[a]*Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France*

ABSTRACT

The Light Field (LF) as an immersive imaging technology has received significant attention from both academia and industry in the last decade. LF enables a wide variety of applications in computer vision and computer graphics domain. Due to its novel representation it has unique challenges in the imaging pipeline. These challenges renders the quality assessment of the LF content a necessity. In this chapter, various aspects regarding the LF quality assessment is discussed including LF related distortion characteristics, subjective user studies and objective quality metrics. The chapter[1] provide an overview of the current state in LF quality assessment.

KEYWORDS

Light field, Subjective quality assessment, Objective quality metrics

1.1 Introduction

In the last century, a 7D plenoptic function was introduced to define the modern LF as introduced in Chapter 1. It is described as below:

$$L_{7d} = P(x, y, z, \theta, \phi, t, \lambda) \tag{1.1}$$

which represents the light ray from any given point (x, y, z) in 3D space, to any direction (θ, ϕ) in 3D space for any given time t and wavelength λ. Although the 7D plenoptic function has a comprehensive definition, it is not fully utilized in practical applications due to the high dimensionality of the data. A practically more desirable version, 4D plenoptic function, is introduced as a result. It represents each ray with 4 points defined on two parallel planes. The coordinates are denoted with (x, y) for the image plane and (a, b) for the camera plane.

LF can be represented in a variety of ways such as sub-aperture views, epipolar plane images (EPI), lenslet image, refocused image stack, pseudo-video sequence (with sub-aperture views or refocused image stack). For more details regarding LF representations can be found in Chapter 7. Figure 1.1 depicts the two most common

[1]This document is the authors copy of the Chapter 10 in the "Immersive Video Technologies" book. https://doi.org/10.1016/B978-0-32-391755-1.00016-X

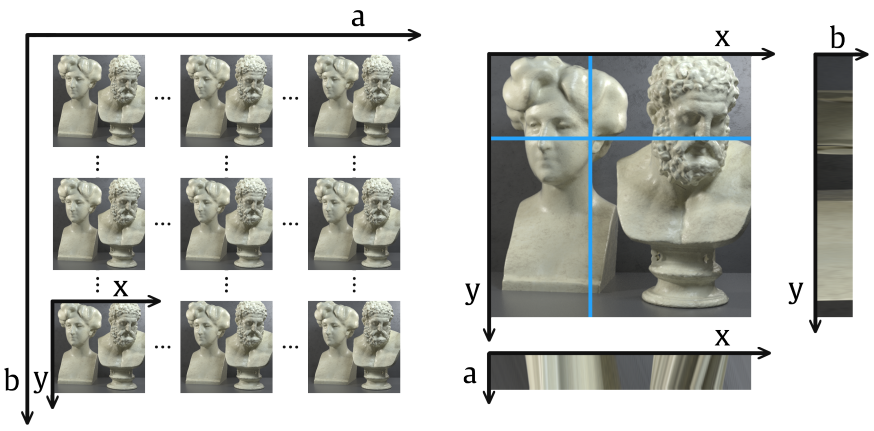

FIGURE 1.1

Sub-aperture views and EPI representations of a sample LF from the WIN5-LID dataset[1].

representations, sub-aperture views and EPIs. On the left, sub-aperture views are displayed on a grid representing the angular domain of the LF with (a, b) coordinates whereas (x, y) coordinates is used for the spatial dimension of the individual sub-aperture views. EPI slices are drawn on the spatial dimension of each sub-aperture view on horizontal or vertical axis of the angular dimension. In other words, by stacking horizontal or vertical sub-aperture views on top of each other, a rectangular slice can be used to generate a single EPI. By shifting the rectangular slice we can generate all the EPIs for the corresponding axis. For a LF with resolution $(a, b, x, y, 3)$ (where 3 represents the red, blue and green color channels), each sub-aperture view has the resolution $(x, y, 3)$, each vertical EPI has the resolution $(a, x, 3)$, and each vertical EPI has the resolution $(b, y, 3)$.

1.2 Characteristics of LF related distortions

Understanding the characteristics of distortions is essential for both objective and subjective quality assessment. In this section, we investigate the impact of various steps in LF imaging pipeline and the impairments occurring at each step.

1.2.1 Acquisition related distortions

Various methods can be used to capture the LF content which are introduced in detail in Chapter 6. Depending on the use-case, certain capture method may be more

advantageous than alternatives. We can categorize them as multiplexed, sequential and multi-sensor acquisition methods [2]. Multiplexed acquisition relies on cameras known as plenoptic cameras that utilizes a micro-lens array placed between the image sensor and main lens. Sequential acquisition methods relies on a single image sensor placed often on a robotic arm or any other moving system to capture multiple views of the scene with a sequential manner. Time-sequential methodology can only be used to capture static scenes. And finally, multi-sensor acquisition utilizes multiple image sensors and captures the scene from multiple views at the same time. Therefore, it is suitable for LF videos and moving scenes.

1.2.1.1 Multiplexed acquisition

Spatially multiplexed cameras are often referred as plenoptic or light-field cameras. Currently, Raytrix [3] is the most commonly known example of such devices. Although plenoptic cameras provides a compact device to capture a light field, they have to have a constant trade-off between the spatial and angular resolution. In other words, increasing the spatial resolution of individual micro-lenses provides a high spatial resolution/low angular resolution LF. Conversely, increasing the number of micro-lenses decreases the spatial resolution and increases the angular resolution of captured LF. Due to the spatial/angular resolution trade-off additional processing steps (e.g. spatial/angular super resolution) might be required that affect the quality of the acquired LF.

LF acquired with plenoptic cameras often suffer from lack of information on the corner sub-aperture views due to micro lens array arrangements. These distortions are visible similar to a vignetting. To overcome such distortions, corner views are often omitted from the acquired LF since filling this information is not straightforward.

When it comes to LF video acquisition, current plenoptic camera technology suffers from the lower frame-rates (3-4 fps). Although plenoptic cameras can shoot with higher fps than sequential acquisition methods, it has a lower fps than multi-sensor acquisition methods. To overcome this limitation, temporal super resolution methods can be used. These methods often introduce specific distortions on the temporal dimension of the processed video.

1.2.1.2 Time-sequential acquisition

Time-sequential acquisition utilizes a predefined trajectory containing sub-aperture views uniformly distributed on a planar or spherical grid. A robotic arm is often utilized to move the image sensor on this predefined trajectory in order to capture the LF. The advantage of this method mainly lies within the high spatial angular resolution and the flexibility of the baseline distance. Main disadvantage is the inability to capture LF videos and dynamic LF scenes (i.e. still LF image with moving objects). In the presence of a moving object in the scene, LFs captured with time-sequential method might contain ghosting artefacts due to location difference between consequent captures.

1.2.1.3 Multi-sensor acquisition

The multi-sensor capture requires a set of image sensors ordered on a planar or spherical grid. Ordered image sensors captures an image simultaneously to generate the LF. Each image sensor in this setup corresponds to a sub-aperture view. Main disadvantage of this method is the cost due to high number of image sensors required to capture. On the other hand, based on the number of image sensors, captured LFs do not suffer from low spatial and angular resolutions. Consequently, multi-sensor capture does not lead to any specific impairments. However, baseline of the captured LF is limited by the physical size of the image sensors.

1.2.2 Processing related distortions

Capturing a high spatio-angular resolution light field with high frame rate is limited by at least one of the dimensions in each capturing method. Research often relies on super resolution techniques on various dimensions to mitigate the resolution limitations. For more information, interested readers may refer to Chapter 8.

1.2.2.1 Spatial super resolution

Spatial resolution of a LF content have a similar impact on the QoE as the traditional 2D content. Although it is not a LF specific concern, the content acquired via plenoptic cameras has always a trade-off between the spatial and angular resolution. To this end, spatial super resolution is an alternative solution to cope with this trade-off. Typically, spatially super resolved images contain blur-like distortions and aliasing problems. Learning-based super resolution algorithms may introduce novel distortions depending on the approach. Moreover, super resolution methods developed for 2D images may introduce additional distortions due to inconsistencies on the angular domain and inconsistency on the temporal domain in cases of LF videos.

1.2.2.2 Angular super resolution

Angular super resolution is commonly called view synthesis in the literature. Similar to spatial super resolution, angular super resolution also helps to cope with the spatial angular resolution trade-off. Moreover, it provides a solution to increase the density of the sub-aperture views of the sparsely captured LF content. View synthesis algorithms shall be categorized into two based on the presence of the depth map as an input. A recent overview of the view synthesis literature can be found in [2]. Quality of the depth map used for view synthesis may affect the intensity and type of distortions over the synthesized views.

Various factors may affect the performance of view synthesis algorithms. Consequently, the visual quality of the synthesized views may vary. One important factor is the occlusions in the captured scene. Occluded areas in the image needs to be filled with information that is unseen previously. Therefore the problem itself (sometimes referred as inpainting) is ill-posed and one of the common inverse imaging problems. Additionally, separating the occluded pixels from the occluding pixels can be an-

other challenge based on the input views. These challenges often leads to structural distortions around the contour of the occluding objects in the scene.

High baseline distances (i.e., the distance between the neighboring sub-aperture views of a LF) further increases the difficulty of the view synthesis task. By increasing the baseline distance, we also increase the disparity between sub-aperture views resulting in distortions with higher intensity and consequently a synthesized view with lower visual quality [4].

Lambertian reflectance model [5] assumes that the light reflects from a surface based on the surface normal and light directions, and the amount of light reflected from the surface is equally bright in every angle. However, complex real world scenes often do not have this property which is violated by specular, translucent and certain reflective surfaces. Pixels that belong to non-lambertian surfaces cannot be represented with a single depth value. Consequently, view synthesis algorithms encounter yet another challenge on non-lambertian surfaces. Inconsistencies along the angular dimension of non-lambertian surfaces can be observed as a result.

1.2.2.3 Temporal super resolution

Temporal super resolution is only a concern for LF video content. Due to slower adoption of LF videos in the literature, the current state of the research regarding to temporal super resolution is somehow limited. However, the characteristics of temporal super resolution distortions are generally common with its 2D counterpart.

Despite similar characteristics, there are two additional complexity for LF video temporal super resolution. First of all, consistency among the 4D LF need to be ensured as opposed to a 2D image. Furthermore, due to high dimensionality of the LF content, current plenoptic cameras can only shoot up to a limited frame rate. As an example, Lytro Illum 2.0 [6] camera shoots only 3 fps. As an alternative solution, Ting et al. [7] proposed a hybrid system containing a DSLR camera and a plenoptic camera to capture high frame rate LF videos. In cases where such hybrid systems are not applicable, direct use of existing LF cameras can be supported with temporal super resolution algorithms to capture LF videos with an acceptable frame rate.

1.2.2.4 Depth estimation

Depth estimation related distortions are not only visible on acquired depth maps but also on the LF that is processed with distorted depth maps. Therefore, the factors that influence the LF quality due to depth map estimation are harder to isolate and discuss in the imaging pipeline. Depending on the use-case, an accurate depth map estimation may become crucial.

Often, depth estimation is done only for the center sub-aperture view [2]. In scenarios where the depth maps for multiple views is estimated, the consistency between the individual depth maps may affect the overall quality. Optimization and evaluation of the acquired depth maps also may influence the quality assessment. Relying on mean squared error as measurement may lower the pixel based accuracy and promote blurry depth maps. Impact of these errors are also enhanced for LF

content with lower baseline. Finally, occluded regions and non-lambertian surfaces pose a challenge on the depth estimation algorithms as discussed earlier. An in-depth overview of the literature for depth estimation of the dense LF content can be found in [8].

1.2.3 Compression related distortions

Compression related distortions are one of the most common distortion types that are represented in LF image quality datasets [9].The type of compression scheme dictates the characteristics of the distortions whereas the compression rate often controls the intensity of distortions. A detailed overview of the LF compression is given in Chapter 7.

One of the common approaches for LF compression is treating the 4D LF as a pseudo video sequence (see Section 1.3.3.2 for more details regarding the pseudo video sequences) and relying on existing hybrid video coding approaches such as HEVC [10] and VVC [11]. Due to its simplicity and high efficiency, this method is widely adopted in the literature [12]. Distortions that occurs with this type of approaches share the same characteristics of 2D video compression distortions. They may occur on the spatial dimension with high compression rates. Distortions that occur as irregularities in the temporal domain instead occur in the angular domain for LFs.

Another common approach utilizes the view synthesis algorithms to efficiently compress LF content. A subset of the sub-aperture views (with additional geometry information) is used to reconstruct the full LF. Often, resulting artefacts depends on the utilized view synthesis methodology. Furthermore, occluded regions and non-lambertian surfaces may alter the geometry information, which may result in structural artefacts on affected regions.

Finally, we have seen an increasing amount of approach utilizing learning-based models in various part of the coding chain such as view synthesis, learning-based prediction and sparse prediction [12]. Their efficiency and the characteristics of the resulting distortions vary greatly from one to another.

1.2.4 Use-case specific influencing factors for LF distortions

As it is suggested in [13], based on the state of adoption several use-cases emerges in the LF domain such as industrial, medical, commercial, educational, cultural and communicational. A detailed analysis of the various key performance indicators [14] and their specific relations to each use case is given in [13]. Since each use-case may have different requirements to provide a greater user experience, understanding use-case specific parameters for LF quality help us to allocate resources into more beneficial parts of the imaging pipeline.

For example, in communication scenarios where the telepresence through mobile devices is concerned the angular quality and angular resolution of the LF visualization gain importance over other parameters. Medical use-cases on the other hand

Table 1.1 Publicly available LF IQA datasets.

	Method	Distortions	Display	# SRC	# Stim	# Obs	Raw Scores
MPI-LFA [15]	JOD	Compression Super Resolution	2D	14	350	40	✗
VALID [16]	DSIS	Compression Refocusing	2D	5	-	-	✗
SMART [17]	PC/JND	Compression	2D	16	4352 (pairs)	19	✗
WIN5-LID [1]	DSCQS	Compression Super Resolution Refocusing	3D	10	220	23	✗
LFDD [18]	DSIS	Compression Processing Contrast Enhancement	2D	10	-	16	-
Turntable [19]	SSCQE	Compression Additive Noise Gaussian Blur	3D LF	7	168	20	-
FVV [20]	ACR	Compression	2D	6	265	23	✓
VSENSE [21]	Eye Tracking	-	2D	20	-	21	-

prioritize the accuracy and precision of the displayed content over visual aesthetics. Therefore, device characteristics such as spatial resolution, brightness, contrast are more important for such scenarios.

1.3 Subjective quality assessment

Similar to the other multimedia types, subjective quality assessment is the golden standard for estimating visual quality of LF content. Subjective experiments to collect human preferences on the LF quality is conducted for this purpose. These experiments are often time-consuming and costly. However, they are essential to understand image quality and develop objective quality metrics that correlates with the human opinion. To this end, researchers collect subjective preferences on the quality of various LF content and made these datasets publicly available to promote research on quality evaluation of LFs. A non-exhaustive list of publicly available datasets is given in Table 1.1.

Unlike other multimedia types, standards and recommendations are not well established for subjective LF quality assessment as of today. Despite the lack of standards, individual efforts are shining light on some of the important research questions. In their study, Darukumalli et al. [22] conducted a series of experiments with different experiment methodologies. The experiments measures the impact of level of zoom on the visual comfort and overall subjective preference. Conducted experiments provides insight regarding the impact of subjective experiment design on QoE experiments conducted on projection based displays, the effect of level of zoom on the

subjective preferences, the impact of the presence (or the lack of) background on the perceived quality and the visual discomfort. Two experiment were conducted to measure the acceptance of zoom level. One experiment utilized the absolute category rating (ACR) method whereas second experiment is conducted with pair comparison (PC) method. The third experiment measures the visual comfort with a single stimulus (SS) method with varying level of zoom. Results indicate that the pair comparison method reveals the user opinions better than the other two methods. The visual comfort results acquired with the ACR and SS experiments highly correlate, indicating that the design choice between the two has minimal effect on the outcome of the experiment.

A recent work from Kara et al. [23] presents an overview of the viewing conditions of the LF video for subjective experiments. In their work, authors investigates the viewing conditions for static, video and interactive content separately. The impact of the viewing position of the observers, spatial and angular resolution of the display/content are discussed. The result indicate a higher tolerance of low angular resolution for static LF content compared to LF videos.

1.3.1 Characterization of LF content for subjective assessment

Image quality datasets are expected to be representative in terms of source content (SRC) and the type of distortions. The type of distortions that can occur in LF content were discussed in the previous section and this section will introduce a set of features that can be used to identify the pristine LF content. A detailed overview of the LF features can be found in [24].

Spatial perceptual information (SI) is a measurement of spatial information in a scene. It is one of the common features that is used for traditional 2D images [25]. Higher SI values indicates a higher spatial complexity for a scene. In SMART LF IQA dataset [17] it is used as the standard deviation over the Sobel [26] filtered luminance image.

Colorfulness (CF) is another feature that is also used for 2D images which impacts the overall aesthetic of a given content. Many descriptions exist for image colorfulness and a detailed overview supported by subjective experiments can be found in [27]. For LF content it is described with the mean and standard deviation of the red, green and blue pixel values [24].

Similar to SI and CF, contrast is also another feature that is commonly used for 2D and LF content. For natural images an extensive analysis of existing measures and the definition of root mean squared contrast can be found in [28]. For LF images, in [24], Gray Level Co-occurence Matrix (GLCM) is utilized as a contrast descriptor.

Disparity range can be used as a feature to incorporate the 3D information regarding the scene. In [24], the algorithm proposed in [29] is used to estimate the pixel disparity and the [minimum disparity, maximum disparity] range is used as a feature. Alternatively, 95% range can be used to increase the robustness towards the errors in disparity estimation.

Refocusing range is another feature described in [24] specificly for LF content.

Based on the refocusing implementation publicly available in Matlab Light Field Toolbox [30] a refocusing range is calculated.

The ratio of the occluded regions in the image can be used as another feature related to the 3D structure of the LF scene. In [24] number of occluded pixels in the LF is calculated with the Matlab Light Field Toolbox [30] and used as a feature. In order to prevent bias towards the spatial resolution of the LF image, number of occluded pixels can be normalized with the spatial resolution of a sub-aperture view.

1.3.2 Quality assessment on LF displays

LF displays are still not widely adopted in the commercial market and definitely not available at the consumer level. Consequently, there is a lack of publicly available LF quality datasets that are collected on LF displays.

LF displays can be categorized as back and front projection. Observers and the light source are located on the same side of the screen for front projection displays while the opposite holds for back projection displays. In comparison to back projection displays, front projection displays requires additional attention to experimental conditions in terms of positioning of the observers.

There are various parameters that affects the quality of a LF content when visualized on a LF display. These parameters can be also called key performance indicators in certain literature[14]. Common with the traditional 2D displays, physical size and the spatial resolution, brightness and color space of the LF displays greatly affect the QoE.

On another front, angular resolution is an influencing factor specific to LF displays that affects the visual quality. Angular resolution governs the smoothness of the parallax effect. Currently, commercially available LF displays are only takes horizontal parallax into consideration and HoliVizio C80 [31] has the highest angular resolution in the market with 0.5 degrees.

Depth budget defines the perpendicular distance around the LF display where an object can appear [14]. It is linearly related to the angular resolution and the pixel size of the display. Importance of the depth budget is highly content dependent and having higher depth may improve the QoE based on the visualized scene.

1.3.2.1 IQA datasets with LF displays

Currently, there are not many publicly available dataset that assess LF quality with LF displays. To the best of our knowledge, Turntable dataset [19] by Tamboli et al. is the only publicly available LF IQA dataset that utilizes a LF display. The angular resolution of the LF display was one view per degree. Holovizio HV721RC display was used for the experiment. This display allow users to experience corresponding viewpoints based on their angle to the display. During the rating task, each observer rated the stimuli on the 5 positions along the viewing arc. Observers rated the stimuli in a quality range from 1 to 5 with Single Stimulus Continuous Quality Evaluation (SSCQE) method.

1.3.3 Quality assessment on other displays

Subjective studies can also be conducted on other display types such as 2D and 3D stereoscopic displays. Majority of the publicly available datasets for LF image quality are collected with 2D displays thanks to their availability.

LF content can be passively displayed on 2D displays as a pseudo video sequence. Frames of these video sequence are often individual sub-aperture views of the LF content on a pre-defined trajectory. In addition, frames of the pseudo video sequence can be refocused version of the LF scene. Passive methodology ensures that the same content is being delivered to all observers in the experiment. However, it fails to provide an interactive experience of the LF content.

Interactive experiments on 2D displays requires user input to determine the trajectory of the displayed sub-aperture views or refocused image on the focus stack. This allows observers to experience the LF content freely. Due to interactive viewing experience, displayed content varies from one observer to another resulting in higher variety in subjective preferences.

1.3.3.1 IQA datasets with other displays

Referring back to the Table 1.1, there are many LF IQA datasets in the literature that are collected on 2D displays. In this section, we will introduce these datasets and discuss some of the details. Interested readers are recommended to refer to the recent work from Ellahi et al. [9] for more details.

MPI-LFA dataset [15] consists of 5 real and 9 synthetic dense LF scenes. LF scenes in the datasets contains only horizontal parallax and has the resolution of $960 \times 720 \times 101$. HEVC compression algorithm and and various reconstruction related distortions were used to generate stimuli. In total, the dataset contains 350 LF in total with Just Objectionable Difference (JOD) scores. JOD scores defines the amount of objectionable difference between two stimuli. The experiment is conducted with Nvidia glasses to allow stereoscopic display. Acquired results suggests that optical flow based reconstruction outperforms both nearest neighborhood and linear interpolation methods. In addition, HEVC compression artefacts were easy to notice by participants in the majority of content.

VALID dataset [16] contains 5 LF scenes with 15×15 angular and 625×434 spatial resolution. Both passive and interactive methodology were utilized with DSIS methodology on a scale of 7 (-3 indicating the poorest quality, +3 indicating best quality). Compression related distortions and refocusing were evaluated in the dataset. Authors did not share any information regarding the participant demographics.

SMART dataset [17] contains 16 LF scenes that are selected based on content features such as spatial information, colorfulness, contrast, brightness, etc. LF scennes were captured with Lytro Illum plenoptic camera. 4 compression algorithms were used to generate the stimuli. Pairwise Comparison (PC) methodology were used to gather Just Noticeable Difference (JND) steps for each stimuli.

WIN5-LID dataset [1] contains 10 LF scenes (6 real, 4 synthetic scenes) and their quality scores collected with a 3D stereoscopic display. HEVC, JPEG2000

compression algorithms and linear, nearest neighborhood and two learning based angular super resolution algorithms were used to generate the distorted stimuli. In total, 220 stimuli were rated in the experiment. Picture quality and the overall quality of the LFs were collected with Double Stimulus Continuous Quality Scale (DSCQS) methodology.

LFDD dataset [18] consists of 8 synthetic LF scenes as the reference stimuli. JPEG, JPEG2000, VP9, AVC, HEVC are some of the compression algorithms that are used to generate the distorted stimuli. Additionally, simple noise and geometrical distortions were tested. Subjective experiment was conducted via crowdsourcing. DSIS was chosen as the subjective testing methodology. A predefined trajectory was used to generate the pseudo video sequences.

FVV dataset [20] contains 6 reference free viewpoint videos with 50 unique views presented as a video sequence from left to right to left in 100 frames. Compression and super-resolution artefacts are used to generate the distorted stimuli. Although it is not named as a LF dataset, the dataset shares the same characteristics with MPI-LFA dataset as it provides a high resolution horizontal parallax with LF related distortions as the stimuli.

VSENSE dataset [21] investigates the visual attention of refocused LF content. Focus stacks for refocusing were generated via the Fourier Disparity Level [32] method on 20 different LFs. Although the dataset does not provide any quality evaluation, we have mentioned the dataset due to close relation of visual attention to image quality. The experiment were conducted to investigate the effect of changes in focus on the visual attention of the observers. Visual attention was measured by collecting eye tracking data from the left eye of the observers during the experiment.

1.3.3.2 Impact of visualization trajectory

As stated previously, passive visualization of the LF content can be done by arranging sub-aperture views with a predefined trajectory as pseudo video sequence. In this case, the impact of trajectory can affect the visual quality of the LF. Figure 1.2 presents 6 different trajectories that are used in a comparative analysis in [33]. The trajectories are visualized with red arrows and explanations are quoted below:

- V1: spiral scan from the external to the internal views in clockwise direction
- V2: diagonal scan in a spiral fashion starting from the view on the left inferior corner
- V3: horizontal scan from left to right starting from the view on the left superior corner
- V4: spiral scan in counter-clockwise direction starting from the center view
- V5: vertical scan from bottom to top starting from the view on the left inferior corner
- V6: diagonal scan from left to right

12 LF content from the EPFL LF dataset [34] is used to generate the pseudo video sequences with the 6 trajectories. 28 participants were recruited to evaluate the quality of the generated video sequences with ACR methodology on a scale from 1 to 5 (1:

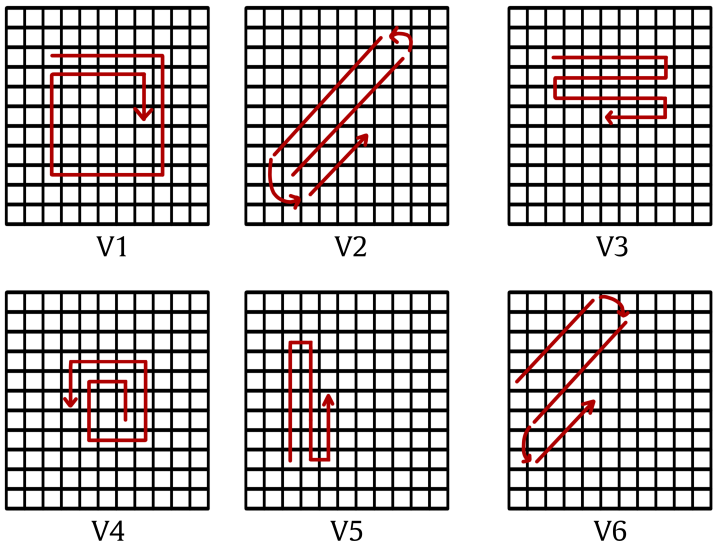

FIGURE 1.2

A common method to create pseudo video sequences for subjective experiments without user interaction. One of the predefined trajectories above can be used to generate a video sequence from the sub-aperture views.

bad, ..., 5: excellent). A training phase was utilized prior to the experiment in order to let the observer familiarize with the task. Mean Opinion Score (MOS) is used to indicate the quality of each video sequence along with the 95 percentile Confidence Interval (CI).

Figure 1.3 presents the result of the comparison between the 6 trajectories in terms of MOS and CI values. Overlapping CIs between the 5 trajectories (V1, V3, V4, V5, V6) except V2 indicates a non-significant difference. On the other hand, V2 has significantly lower MOS value than the other trajectories. Referring to Figure 1.2, lower MOS value can be explained, trajectory contains large disparity on both vertical and horizontal dimensions around the corner of the trajectories. This may disturb the observers and consequently lower the MOS value. A more detailed discussion and analysis of the results can be found in [33].

On another front, Figure 1.4 visualizes a sample rendering trajectory on the LF images with horizontal parallax only. The effect of rendering trajectory on the quality of experience is discussed in details for free viewpoint videos in [35]. Although they follow a different naming convention, free viewpoint videos are LF with horizontal parallax and the rendering trajectory of this LF is defined over temporal dimension. The results indicate that the observers shows significant preference towards certain rendering trajectories. Based on the regions of interest in the content, careful selection

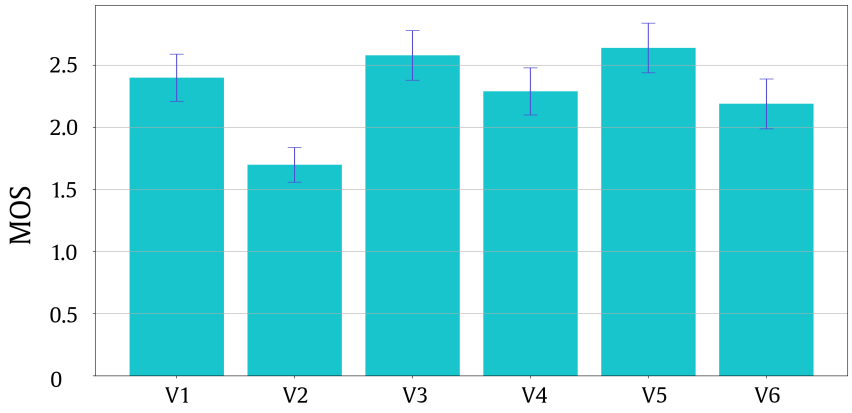

FIGURE 1.3

Comparison of the MOS scores and their 95% CIs for the 6 predefined trajectories

of rendering trajectory is required.

1.4 Objective quality assessment

Although subjective experiments are the golden standard for quality assessment automatically assessing the image quality is still necessary for many applications. It is also not practical to conduct a subjective experiment every time a quality evaluation is required. Additionally, determining the quality of LF content for real-time applications cannot be done via subjective experiments. To this end, objective quality metrics allow us to estimate the image quality in an automated fashion.

1.4.1 Visibility of LF related distortions on EPI

We have investigated the LF specific distortions and their sources in LF imaging pipeline in Section 1.2. Many of these distortions were considered when collecting LF image quality datasets in the literature. In this section, we will investigate these distortions on the selected datasets and their visibility on the EPI representations.

A visual inspection of the EPI patches and corresponding edge maps from 5 different LFs in MPI-LFA dataset [15](see Section 1.3.3.1) is given in Figure 1.5. Reference EPI patch (i.e., pristine, without distortion) is displayed at the top meanwhile the 4 EPI patches below contains distortions due to super-resolution algorithms indicated on the left. A brief examination of the EPI patches reveals the characteristics of each super-resolution method. Edge maps acquired with Canny edge detection algorithm [36] provides a binary map emphasising the visibility of this phenomenon.

Based on this observation, it is natural to investigate the benefit of EPI represen-

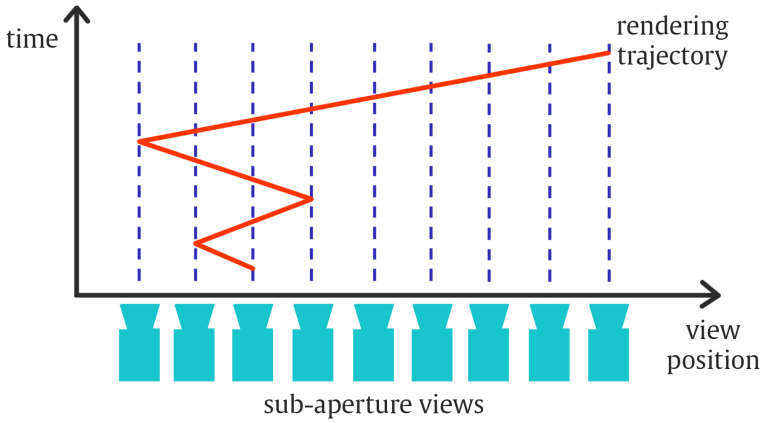

FIGURE 1.4

Rendering trajectory on LF with only horizontal parallax defined on the temporal dimension.

tations on the image quality metric performances. Therefore, next section introduces a set of image quality metrics that leverages structural information within an image to predict the overall quality. Afterwards, we compare the difference between metric performances on EPI and sub-aperture view representations.

1.4.1.1 Structural image quality metrics

Natural image counter evaluation (NICE) [37] is a full-reference structural image quality metric that relies on the edge maps of reference and distorted image. Extracted edge maps from the reference and distorted images are first dilated with a plus-sign kernel. Later, non-zero elements in the XOR maps between the dilated image pair is calculated to be pooled into a final quality score. Different edge maps might affect the performance of the metric.

Gradient magnitude similarity deviation (GMSD) [38] is another full-reference structural image quality metric that utilizes the directional image gradients. Both reference and distorted images are convoluted with Prewitt [39] filters (horizontal (h_x) and vertical(h_y)):

$$h_x = \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix}, h_y = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \end{bmatrix}$$

Later, gradient magnitudes at each pixel location for both images are calculated as follows:

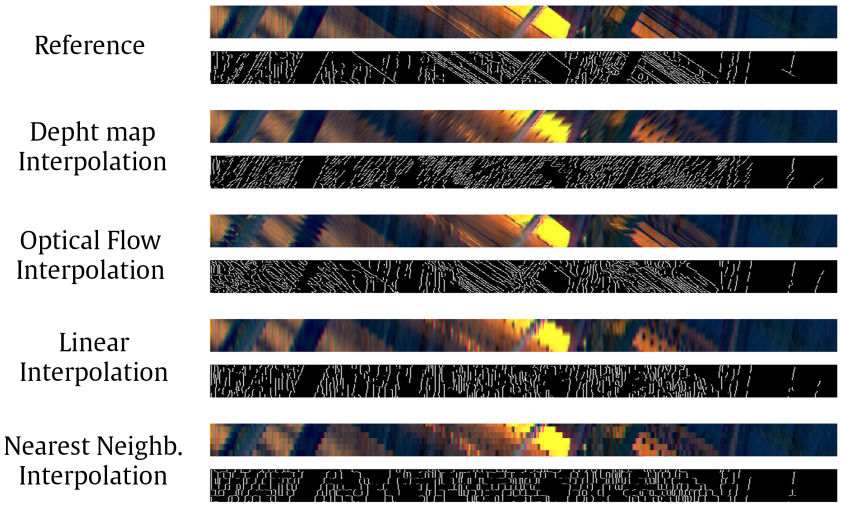

FIGURE 1.5

Sample EPI patches of 5 different LFs from MPI-LFA dataset[15]. Edge maps of the EPI patches are also presented below to demonstrate the visibility and characteristics of various distortions.

$$m_r(i) = \sqrt{(r \circledast h_x)^2(i) + (r \circledast h_y)^2(i)} \tag{1.2}$$

$$m_d(i) = \sqrt{(d \circledast h_x)^2(i) + (d \circledast h_y)^2(i)} \tag{1.3}$$

to be used to calculate gradient magnitude similarity (GMS) map with the following function

$$GMS(i) = \frac{2m_r(i)m_d(i)}{(m_r)^2(i) + (m_d)^2(i)} \tag{1.4}$$

Finally, GMS maps are pooled with standard deviation pooling to estimate the final image quality of the distorted image in comparison to the reference image.

Morphological Wavelet Peak Signal-to-Noise Ratio (MW-PSNR) [40] calculates the image quality based on morphological wavelet decomposition. PSNR is calculated between each band of decomposed reference and distorted image pairs.

1.4.1.2 Sub-aperture Views vs EPI

Structural information present in the EPI regarding the angular dimension of the LF makes it an important representation to understand the presence and the intensity

Table 1.2 PCC between metric predictions and MOS scores on FVV dataset[20]. First column corresponds to PCC values when EPI representations are used whereas second column corresponds to PCC values on sub-aperture views.

	EPI	View
MW-PSNR	0.7698	0.7921
GMSD	0.7410	0.6715
NICE	0.5122	0.4310

of distortions. With this motivation, we investigate the performance of the three structural image quality metrics that are introduced in the previous section on sub-aperture view images and EPIs. Interested readers are recommended to refer to the original work [41] for more information regarding the motivation and the experiment.

Experiment was conducted on FVV dataset [15] (see Section 1.3.3.1). Stimuli in FVV dataset contains 50 horizontally arranged sub-aperture views with no vertical parallax. For each

Table 1.2 presents the result of the experiment in terms of Pearson Correllation Coefficient (PCC) between the metric predictions and MOS values for the three metrics used in the experiments. Higher PCC values indicate a better performance for the metrics. Both sub-aperture views and EPI representations were used to test the metrics. It can be observed that the GMSD and NICE metrics performs better when used over EPI representations instead of sup-aperture views. Structural information such as directional gradients and edge maps reveals the distortions that causes inconsistencies in the angular domain. Consequently, GMSD and NICE that relies on filtering the image with Prewitt and Canny kernels performs better on EPI representations. On the other hand, MW-PSNR performance is lower when used over EPI representations. This can be explained by the low spatial resolution of each EPI slice. Multi-scale wavelet decomposition cannot be fully utilized due to the low resolution.

1.4.2 LF image quality metrics

As it stands today, LF is an emerging technology and quality assessment of LF is an emerging research field. Consequently, there are no well established image quality metrics for LF content. Due to complex imaging pipeline, ill-defined nature of the LF related distortions and the higher dimensionality of the LF content developing objective quality metrics is even more challenging. However, many attempts have been made in the last decade and more than 30 objective image quality metrics exist in the literature.

Similar to the traditional image quality metrics, we can categorize LF image quality metrics based on the availability of the reference image as an input as full-reference (FR), reduced-reference (RR) or no-reference (NR). When the pristine reference image

is present as an input along with the distorted LF, the metric is called a FR metric. Conversely, in the absence of the pristine reference the metric is called NR. In cases where a set of feature but not the full reference LF is used, the metric is referred as a RR metric.

On another front, we can categorize the metrics based on the LF representations used as an input. Objective metrics can utilize a variety of LF representations such as lenslet image, sub-aperture view, EPI, refocused image stack, pseudo video sequence.

The density of the LF (i.e., the baseline distance) can be also used to categorize the objective quality metrics. Currently, due to majority of the publicly available datasets using dense LFs (i.e. low baseline distance) majority of objective quality metrics are also developed and evaluated on dense LF content. Therefore, at its current state, this categorization might be redundant despite its importance.

The distortions being targeted by the metrics also play an important role and therefore an informative way to categorize the metrics. Since the datasets used to develop the metrics often dictates the distortions that are targeted by the metric, this categorization can be done via the utilized datasets.

For LF images, metrics often contain two separate streams to evaluate the quality of the spatial and angular dimensions. Two streams are often used to extract features and typically a regression model or a pooling strategy is utilized to merge the two streams into predicting the final quality of the LF content. In addition to the two streams, a third stream can be utilized for LF video content for the temporal dimension. To this date, there are no LF video quality metrics in the literature. Tamboli et al. conducted a study [42] on the use of LF image quality metric [19] to assess LF video frames. Result of the study indicates that the distortions on the temporal space can be separated in the metric space. However, further confirmation with a subjective experiment is required to demonstrate the efficacy of the approach.

Non-exhaustive list of the LF image quality metrics is given in Table 1.3 along with the properties introduced above. At the input column of the table, SAV stands for sub-aperture views whereas SMV stands for super multi-view. Below we introduce some of these metrics. Note that, the following metrics do not consider any use-case. In other words, they evaluate the LF quality independently from the rendering methodology/trajectory. Therefore, the results are purely data driven and various use cases may introduce other complications which are ignored by the proposed metrics.

First metric in the table is proposed by Ak et al. [43] and it is a NR metric which takes EPIs from the distorted LF image as an input. It is built on the work illustrated in Section 1.4.1. The metric utilizes the visibility of the LF related distortions on the EPIs and relies on two set of features as the histogram of oriented gradients (HOG) based bag-of-words codebook and a convolutional sparse coding dictionary. The two set of features than used to predict the quality score with a support vector regression model. The model is trained and evaluated on the MPI-LFA dataset and the results indicate that the metric provides a high correlation to human judgement on LF content distorted with compression and super resolution distortions with passive display method.

Tamboli et al. proposed an objective quality metric [19] that targets the sparse LF

Table 1.3 A non-exhaustive list of LF image quality metrics from the literature.

First Author	Ref	Input	Training Dataset	Evaluation Dataset	LF Density
Ak [43]	NR	EPI	MPI-LFA [15]	MPI-LFA [15]	Dense
Guo [44]	NR	SAV	HCI [45] EPFL [34]	WIN5-LID [1]	Dense
Shi [46]	NR	EPI Cyclopean Img	MPI-LFA [15] SMART [17] WIN5-LID [1]	MPI-LFA [15] SMART [17] WIN5-LID [1]	Dense
Tamboli [19]	FR	SMV	Turntable [19]	Turntable [19]	Sparse
Tian [47]	FR	SAV EPI	MPI-LFA [15]	MPI-LFA [15]	Dense
Zhou [48]	NR	SAV	MPI-LFA [15] SMART [17] VALID [16] WIN5-LID [1]	MPI-LFA [15] SMART [17] VALID [16] WIN5-LID [1]	Dense

content and currently it is the only publicly available objective quality metric which is developed for sparse LF content. Steerable pyramid decomposition over the 3D view of the reference and distorted LF content is used to extract features and quantify the distortions. To assess the quality on the angular dimension, the metric calculates the structural similarity between optical flow arrays obtained from the reference and distorted LFs. The metric were evaluated on the same dataset [19] proposed in the paper. A cross-validation is used to remove the bias towards the training/test split.

Guo et al. proposed the NR metric [44] which utilizes sub-aperture views with two parallel streams named sub-aperture view fusion and global context perception. The metric is unique in the sense of its training. Pre-training of the metric is done via synthetically generated labels, called ranking-MOS over a large collection of dense LF content. Finally, metric is fine-tuned on WIN5-LID dataset [1].

Shi et al. proposed another NR metric [46] which utilizes the cyclopean image array to measure the spatial quality of the LF images and EPIs to measure the angular quality. Similar to [43], the proposed metric relies on the distribution of gradient directions on the EPIs. A support vector regression model is used to combine extracted features to predict final LF image quality.

A FR metric [47] was proposed by Tian et al. relying on two sets of features (symmetry and depth) for quality estimation of LF image content. Symmetry features were calculated over the sub-aperture views of reference and distorted LF in 2 sub-streams as the magnitude-based and orientation-based. While orientation-based symmetry features reveals the distortions on the image details, magnitude-based features emphasises the contour information. For the angular consistency, the proposed

metric relies on depth features extracted from the EPIs. Angular quality is quantified by comparing the variance of the lines in a given set of reference and distorted EPI. A weighted average of symmetry and depth scores are then used to predict the final quality.

Zhou et al. proposed a NR quality metric [48] where the LF image is treated as a four dimensional tensor. The metric operates in CIELAB color space and Tucker decomposition is used to extract principal components. Principal components spatial characteristics is used to measure the spatial quality whereas tensor angular variation index is used to quantify the angular consistency. After extracting features from the two streams, a support vector regression model is equipped to predict final quality score.

1.5 Conclusion

Industrial and academical adoption of light field content widens thanks to unique opportunities it enables. Last decade brought a great amount of attention to the light field research. Despite the increasing interest, we are still far from light field content reaching out to the consumer market. In the commercial market, we have seen various level of adoption for different use-cases such as medical, cultural, communicational, industrial [13]. For a given use-case, the requirements for a greater user experience may be different than others [14]. This makes the quality assessment of light fields more challenging.

Conducting a subjective study is the ultimate way to assess the quality of experience. There is no exception for the use-cases concerning the light field content. To this end, various subjective studies were conducted and shared with public. Among these, there are studies assessing the quality under varying rendering trajectories, independent of any rendering trajectory, with different focus planes, eye-tracking experiments. A number of testing methodology has been used in the literature including DSIS and pairwise comparison. Consequently, subjective preferences are shared as mean opinion scores, pairwise comparison matrices or just observable differences along the tested stimuli as part of the dataset. Currently number of subjective studies with light field displays are low in number in the literature. Majority of the studies uses traditional 2D displays and treats the light field as a pseudo video sequence. Note that, the subjective studies that are not conducted with light field displays may not reflect the effect of all possible influencing factors for all use-cases, they still provide a good ground to improve upon.

Objective quality assessment of light field is still developing and the number of publicly available light field image quality metrics is increasing. To this day, there is no light field video quality metrics as well as an objective quality metric for high dynamic range light field images. The metrics currently do not consider rendering trajectory while assessing the light field quality and majority of the metrics are targeted for dense light fields. The common scheme for the majority of metrics contains two streams of features assessing the spatial and the angular quality separately. Often a module such as support vector regression model is used to pool the extracted features

from the two streams and predict the final light field quality.

With wider adoption of the light field in commercial uses, the key performance indicators will continue to evolve. A wide variety of application will likely to adopt light fields bringing new use-cases and challenges. This creates a need for constant development and improvement of methods and tools to assess light field quality.

Bibliography

[1] L. Shi, S. Zhao, W. Zhou, Z. Chen, Perceptual evaluation of light field image, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 41–45.

[2] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, Y. Liu, Light field image processing: An overview, IEEE Journal of Selected Topics in Signal Processing 11 (7) (2017) 926–954. doi:10.1109/JSTSP.2017.2747126.

[3] Raytrix 3d light-field camera, https://raytrix.de/, accessed: Jan 2022. [Online].

[4] D. Yue, M. S. Khan Gul, M. Bätz, J. Keinert, R. Mantiuk, A benchmark of light field view interpolation methods, in: 2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2020, pp. 1–6. doi:10.1109/ICMEW46912.2020.9106041.

[5] S. J. Koppal, Lambertian Reflectance, Springer US, Boston, MA, 2014, pp. 441–443. doi:10.1007/978-0-387-31439-6_534.
URL https://doi.org/10.1007/978-0-387-31439-6_534

[6] Lytro light-field camera, https://www.lytro.com/, currently the website redirects to Raytrix LF camera website.

[7] T. Wang, J. Zhu, N. K. Kalantari, A. A. Efros, R. Ramamoorthi, Light field video capture using a learning-based hybrid imaging system, CoRR abs/1705.02997 (2017). arXiv:1705.02997.
URL http://arxiv.org/abs/1705.02997

[8] O. Johannsen, K. Honauer, B. Goldluecke, A. Alperovich, F. Battisti, Y. Bok, M. Brizzi, M. Carli, G. Choe, M. Diebold, M. Gutsche, H.-G. Jeon, I. S. Kweon, J. Park, J. Park, H. Schilling, H. Sheng, L. Si, M. Strecke, A. Sulc, Y.-W. Tai, Q. Wang, T.-C. Wang, S. Wanner, Z. Xiong, J. Yu, S. Zhang, H. Zhu, A taxonomy and evaluation of dense light field depth estimation algorithms, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1795–1812. doi:10.1109/CVPRW.2017.226.

[9] W. Ellahi, T. Vigier, P. L. Callet, Analysis of public light field datasets for visual quality assessment and new challenges, 2019.

[10] .-. ISO/IEC, High efficiency video coding, apr 2013.

[11] .-. ISO/IEC, Versatile video coding, jul 2020.

[12] C. Conti, L. D. Soares, P. Nunes, Dense light field coding: A survey, IEEE Access 8 (2020) 49244–49284. doi:10.1109/ACCESS.2020.2977767.

[13] P. A. Kara, R. R. Tamboli, T. Balogh, B. Appina, A. Simon, On the use-case-specific quality degradations of light field visualization, in: C. F. Hahlweg, J. R. Mulley (Eds.), Novel Optical Systems, Methods, and Applications XXIV, Vol. 11815, International Society for Optics and Photonics, SPIE, 2021, pp. 81 – 94. doi:10.1117/12.2597363.
URL https://doi.org/10.1117/12.2597363

[14] P. A. Kara, R. R. Tamboli, O. Doronin, A. Cserkaszky, A. Barsi, Z. Nagy, M. G. Martini, A. Simon, The key performance indicators of projection-based light field visualization, Journal of Information Display 20 (2019) 81 – 93.

[15] V. Kiran Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, P. Didyk, Towards a quality metric for dense light fields, in: Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition, 2017, pp. 58–67.

[16] I. Viola, T. Ebrahimi, VALID: Visual quality assessment for light field images dataset, in: 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2018, pp. 1–3.

[17] P. Paudyal, R. Olsson, M. Sjöström, F. Battisti, M. Carli, SMART: A light field image quality dataset, in: Proceedings of the 7th International Conference on Multimedia systems, 2016, pp. 1–6.

[18] A. Zizien, K. Fliegel, LFDD: Light field image dataset for performance evaluation of objective quality metrics, in: Applications of Digital Image Processing XLIII, Vol. 11510, International Society for Optics and Photonics, 2020, p. 115102U.

[19] R. R. Tamboli, B. Appina, S. Channappayya, S. Jana, Super-multiview content with high angular resolution: 3d quality assessment on horizontal-parallax lightfield display, Signal Processing: Image Communication 47 (2016) 42–55. doi:https://doi.org/10.1016/j.image.2016.05.010.
URL https://www.sciencedirect.com/science/article/pii/S0923596516300674

[20] E. Bosc, P. Hanhart, P. Le Callet, T. Ebrahimi, A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data, in: Fifth International Workshop on Quality of Multimedia Experience (QoMEX), Klagenfurt, Germany, 2014, pp. pp.100–105. doi:10.1109/QoMEX.2013.6603218.

[21] A. Gill, E. Zerman, C. Ozcinar, A. Smolic, A study on visual perception of light field content, in: The Irish Machine Vision and Image Processing Conference (IMVIP), 2020.

[22] S. Darukumalli, P. A. Kara, A. Barsi, M. G. Martini, T. Balogh, A. Chehaibi, Performance comparison of subjective assessment methodologies for light field displays, in: 2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2016, pp. 28–33. doi:10.1109/ISSPIT.2016.7886004.

[23] P. A. Kara, R. R. Tamboli, A. Cserkaszky, M. G. Martini, A. Barsi, L. Bokor, The viewing conditions of light-field video for subjective quality assessment, in: 2018 International Conference on 3D Immersion (IC3D), 2018, pp. 1–8. doi:10.1109/IC3D.2018.8657881.

[24] P. Paudyal, J. Gutiérrez, P. Le Callet, M. Carli, F. Battisti, Characterization and selection of light field content for perceptual assessment, in: 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), 2017, pp. 1–6. doi:10.1109/QoMEX.2017.7965635.

[25] ITU-P, Subjective video quality assessment methods for multimedia applications, ITU-P.910 (Jan 2022).

[26] N. Kanopoulos, N. Vasanthavada, R. L. Baker, Design of an image edge detection filter using the sobel operator, IEEE Journal of solid-state circuits 23 (2) (1988) 358–367.

[27] C. Amati, N. J. Mitra, T. Weyrich, A study of image colourfulness, in: Proceedings of the Workshop on Computational Aesthetics, CAe '14, Association for Computing Machinery, New York, NY, USA, 2014, p. 23–31. doi:10.1145/2630099.2630801.
URL https://doi.org/10.1145/2630099.2630801

[28] R. A. Frazor, W. S. Geisler, Local luminance and contrast in natural images, Vision Research 46 (10) (2006) 1585–1598. doi:https://doi.org/10.1016/j.visres.2005.06.038.
URL https://www.sciencedirect.com/science/article/pii/S0042698905005559

[29] S. Wanner, B. Goldluecke, Variational light field analysis for disparity estimation and super-resolution, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (3) (2014) 606–619. doi:10.1109/TPAMI.2013.147.

[30] D. G. Dansereau, O. Pizarro, S. B. Williams, Decoding, calibration and rectification for lenselet-based plenoptic cameras, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1027–1034. doi:10.1109/CVPR.2013.137.

[31] Holovizio c80 glasses-free 3d cinema system, https://holografika.com/c80-glasses-free-3d-cinema/, accessed: Jan 2022. [Online].

[32] M. Le Pendu, C. Guillemot, A. Smolic, A Fourier disparity layer representation for light fields, IEEE

Transactions on Image Processing 28 (11) (2019) 5740–5753.

[33] F. Battisti, M. Carli, P. Le Callet, A study on the impact of visualization techniques on light field perception, in: 2018 26th European Signal Processing Conference (EUSIPCO), 2018, pp. 2155–2159. doi:10.23919/EUSIPCO.2018.8553558.

[34] M. Rerabek, T. Ebrahimi, New light field image dataset, 2016.

[35] S. Ling, J. Gutiérrez, K. Gu, P. Le Callet, Prediction of the influence of navigation scan-path on perceived quality of free-viewpoint videos, IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9 (1) (2019) 204–216. doi:10.1109/JETCAS.2019.2893484.

[36] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8 (6) (1986) 679–698. doi:10.1109/TPAMI.1986.4767851.

[37] D. M. Rouse, S. S. Hemami, Natural image utility assessment using image contours, in: 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 2217–2220. doi:10.1109/ICIP.2009.5413882.

[38] W. Xue, L. Zhang, X. Mou, A. C. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE Transactions on Image Processing 23 (2) (2014) 684–695. doi:10.1109/TIP.2013.2293423.

[39] J. M. S. PREWITT, Object enhancement and extraction, Picture Processing and. Psychopictorics (1970).
URL https://ci.nii.ac.jp/naid/10017095478/en/

[40] D. Sandić-Stanković, D. Kukolj, P. Le Callet, Dibr-synthesized image quality assessment based on morphological multi-scale approach, EURASIP Journal on Image and Video Processing 2017 (07 2016). doi:10.1186/s13640-016-0124-7.

[41] A. Ak, P. Le-Callet, Investigating epipolar plane image representations for objective quality evaluation of light field images, in: 2019 8th European Workshop on Visual Information Processing (EUVIP), 2019, pp. 135–139. doi:10.1109/EUVIP47703.2019.8946194.

[42] R. R. Tamboli, P. A. Kara, A. Cserkaszky, A. Barsi, M. G. Martini, B. Appina, S. S. Channappayya, S. Jana, 3d objective quality assessment of light field video frames, in: 2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2018, pp. 1–4. doi:10.1109/3DTV.2018.8478557.

[43] A. Ak, S. Ling, P. Le Callet, NO-REFERENCE QUALITY EVALUATION OF LIGHT FIELD CONTENT BASED ON STRUCTURAL REPRESENTATION OF THE EPIPOLAR PLANE IMAGE, in: The 1st ICME Workshop on Hyper-Realistic Multimedia for Enhanced Quality of Experience, London, United Kingdom, 2020.

[44] Z. Guo, W. Gao, H. Wang, J. Wang, S. Fan, No-reference deep quality assessment of compressed light field images, in: 2021 IEEE International Conference on Multimedia and Expo (ICME), 2021, pp. 1–6. doi:10.1109/ICME51207.2021.9428383.

[45] K. Honauer, O. Johannsen, D. Kondermann, B. Goldlücke, A dataset and evaluation methodology for depth estimation on 4d light fields, in: ACCV, 2016.

[46] L. Shi, W. Zhou, Z. Chen, J. Zhang, No-reference light field image quality assessment based on spatial-angular measurement, IEEE Transactions on Circuits and Systems for Video Technology 30 (11) (2020) 4114–4128. doi:10.1109/TCSVT.2019.2955011.

[47] Y. Tian, H. Zeng, J. Hou, J. Chen, J. Zhu, K.-K. Ma, A light field image quality assessment model based on symmetry and depth features, IEEE Transactions on Circuits and Systems for Video Technology 31 (5) (2021) 2046–2050. doi:10.1109/TCSVT.2020.2971256.

[48] W. Zhou, L. Shi, Z. Chen, J. Zhang, Tensor oriented no-reference light field image quality assessment, IEEE Transactions on Image Processing 29 (2020) 4070–4084. doi:10.1109/TIP.2020.2969777.