Towards Realistic Light Field Experiences: Camera Animation, User Interaction, HDR Reconstruction, and Quality of Experience

Ph.D. Dissertation

Mary GUINDY

Thesis Advisors:

PhD Supervisor: Prof. Dr. Péter SZOLGAY, DSc

PhD co-supervisor: Dr. Vamsi Kiran ADHIKARLA, PhD

Pázmány Péter Catholic University

Faculty of Information Technology and Bionics

Roska Tamás Doctoral School of Sciences and Technology

2024

STATEMENT OF PhD SUPERVISOR

1. PUBLICATION OUTPUT

Signed below Dr. SZOLGAY Péter PhD Supervisor declare that the Mary Mohsen Messak Guindy doctoral candidate meet the requirements of the Multidisciplinary Doctoral Council of Sciences and Technology in the field of the Doctoral Studies as a prerequisite for the award of the doctorate (PhD).

2. PhD THESIS

I declare that the topic of the doctoral thesis submitted by Mary Mohsen Messak Guindy docandidate is scientifically meaningful and contains authentic data, that the scientific results contained therein are the candidate's own scientific results, and that the thesis complies with the requirements of the regulations of PPKE and the Roska Tamás Doctoral School of Sciences and Technology.

I support the submission of the PhD thesis for public discussion.

Date

15 November 2024

Prof. Dr. SZOLGAY Péter DSc

Full professor (PPKE)

Acknowledgements

I am profoundly grateful to the individuals and organizations who have supported and encouraged me throughout my journey, and without whom this thesis would not have been possible.

First and foremost, I extend my deepest gratitude to my family: my parents, my brother, and my uncle, who has been like a second father to me. Despite the physical distance between us, they have consistently offered unwavering support, helping me overcome every challenge and encouraging me through each step of this endeavor. Their encouragement and belief in me have been invaluable.

I would like to express my sincere appreciation to the Marie Curie Project Horizon 2020 EVOCATION for making it possible for me to pursue my PhD in Hungary. Their support allowed me not only to travel but also to engage in a range of invaluable academic opportunities, including workshops, conferences, and fairs across Europe, where I could collaborate with others and expand my horizons as a researcher. This project has been instrumental in broadening my perspective and advancing my academic career.

I am especially indebted to my two supervisors, Attila Barsi and Dr. Peter A. Kara, whose mentorship has been absolutely indispensable to my growth as a researcher and academic. Without their unwavering guidance, patience, and expertise, I would not have been able to accomplish my PhD.

Attila Barsi, my primary supervisor, has been the cornerstone of my understanding of light fields and light field displays. His thorough teaching, step-by-step guidance, and patient approach, especially in explaining complex coding challenges, have been invaluable. Since my arrival in Hungary, he has been not only a mentor but also my most steadfast supporter, for which I am deeply grateful.

Dr. Peter A. Kara has been equally essential, generously giving me his time to support my development as a scholar. His dedication to helping me refine my academic writing and ideas has enabled me to publish over twenty papers – an accomplishment I could not have achieved without his investment in my growth. His patient mentorship has taught me to think critically, generate new ideas, and write with clarity and precision. The commitment of Dr. Kara to my progress has been an invaluable gift, one for which I am profoundly thankful.

To both Attila Barsi and Dr. Peter A. Kara, I owe my sincerest thanks. Their mentorship, patience, and belief in my potential made this achievement possible, and their influence will continue to guide me in my future academic pursuits.

I would also like to extend my thanks to Dr. Peter Szolgay for his assistance in navigating university requirements and ensuring a smooth academic experience, to Dr. Adhikarla Vamsi for his guidance, and to Tibor Balogh, CEO of Holografika Kft., for his insights and contributions to the field.

To all of these remarkable mentors, supporters, and collaborators, I owe my heartfelt appreciation for their patience, guidance, and the invaluable knowledge they have shared with me. Their influence has shaped my path and will continue to inspire me in my future endeavors.

Abstract

This research investigates critical advancements in light field technology, which enables the capture and representation of 3D scenes through light rays, facilitating immersive, glasses-free visualization for multiple simultaneous viewers. While light field technology holds transformative potential across various domains, it encounters significant challenges that this study addresses, including the underexplored area of light field camera animation, complexities in user interface design, and the integration of high dynamic range (HDR) to enhance color fidelity and luminance depth in 3D scenes. This research introduces novel light field camera animation techniques to expand cinematic and interactive applications, along with innovative interaction models that improve light field display usability, making them more intuitive for diverse applications. The study further advances HDR light field imaging through an analysis of HDR applications within light field systems, the application of LDR-to-HDR reconstruction convolutional neural networks to light field images, and the development of a dedicated HDR light field dataset tailored to support precise HDR reconstructions and visualization. The thesis also presents a detailed quality of experience analysis for light field displays, examining essential factors such as optimal viewing distances and visualization quality for diverse user profiles. To sum up, key contributions include developments in light field camera animation, interactive user interfaces, and high dynamic range light field imaging, complemented by comprehensive assessments of user experience and quality of experience metrics. The findings aim to propel light field technology toward practical applications in various fields, enhancing visualization quality and accessibility for diverse user profiles.

Contents

CHAPTER 1

Introduction

Among novel capture and visualization technologies, Light Field (LF) has made significant advancements, edging closer to everyday applications. LF technology has emerged as a means of representing the 3D world – to which it acts as a window – by light rays filling up the 3D space under representation [42]. Light Field Display (LFD)s were then invented as a means of visualizing the captured LFs. Unlike many 3D display systems, LFDs deliver a complete 3D experience without requiring personal viewing devices. Due to the lack of such constraint, these displays may be viewed by any number of observers simultaneously, and the corresponding use case contexts may also involve a virtually unlimited numbers of users; any number that the Valid Viewing Area (VVA) of the display may accommodate [117, 120]. Chapter 2 provides a comprehensive review of LF, exploring their historical evolution, foundational properties, and representation models. It covers LF history, imaging and capture techniques, available datasets, and methods for visualization, compression, and super-resolution of LF data, establishing a solid foundation for the advancements discussed in subsequent chapters.

While many instances of the utilization of this technology operate with static contents, camera animation may also be relevant [120]. Regarding use cases, LF is applicable to both cinematic and interactive contents. Such contents often rely on camera animation, which is a frequent tool for the creation and presentation of 2D contents. However, while common 3D camera animation is often rather straightforward, LF visualization has certain constraints that must be considered before implementing any variation of such techniques. In Chapter 3, we introduce our work on camera animation for LF visualization. The different types of conventional camera animation were applied to LF contents, which produced an interactive simulation. The simulation was visualized and assessed on a real LFD, the results of which we present and discuss. Additionally, we tested different forms of realistic physical camera motion in our study and proposed multiple metrics for the quality evaluation of LF visualization in the investigated context and for the assessment of plausibility [117]. Subjective tests were also conducted to address user preferences regarding realistic physical camera motion on LFDs.

Despite the numerous advantages and attractive capabilities of such glasses-free 3D displays, their User Interface (UI) methods are quite complicated and they are currently underwhelming when compared to conventional 2D displays, due to the fact that visual feedback can only be rendered sharply on the emission surface of LFDs. Due to their overall value and usefulness, interaction techniques develop immediately as new types of displays arise. With the recent advancements in visualization technologies, UIs have been redesigned for use in Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR) visualization. This includes on-screen augmentation, which enables interaction with visual content on the screen. On the other hand, only basic UIs have been developed for LFDs so far. In order to test the different interaction methods on LFDs, a theater model depicting the virtual environment was implemented. Methods for rendering the theater model and monitor room, along with the results of the interactions are discussed in Chapter 4, illustrated by images of the actual visualization on LFDs. It is shown that producing plausible results with no noticeable visual artifacts is challenging, yet possible. The scientific contributions of the chapter also highlight the various novel UIs for future LF systems and services. Additionally, subjective assessment was conducted on a large-scale LF cinema system to evaluate the potential interaction techniques implemented via theater model. Multiple subjective quality metrics were used to decompose the visual experience of the observers, and additional attention was paid to the essential aspects of long-term utilization, such as dizziness [118, 123].

While multi-autostereoscopic systems, such as LFDs, offer immersive 3D experiences without the need for additional viewing gears, High Dynamic Range (HDR) technology enhances the realism of visual content. HDR images are created with more luminance levels compared to conventional Low Dynamic Range (LDR) images [323, 124]. Currently, there is a strive towards enhancing the capabilities of the capture and display devices to accommodate the dynamic range of HDR images, which, in turn, adds realism to visualization by being close to the capabilities of the Human Visual System (HVS). In addition to providing a wider color gamut compared to conventional Red Green Blue (RGB), HDR images succeed at recording extra information which is not visible to the eye otherwise. Combining both LF and HDR technologies is rather powerful, where 3D content is visualized with an added sense of realism, close to the HVS. However, this combination presents challenges due to the inherent limitations of both technologies [122]. Chapter 5 begins with an in-depth analysis of HDR LF imaging applications, highlighting the benefits and challenges of integrating HDR with LF technology. Following this, the chapter explores Convolutional Neural Network (CNN)-based methods for LDR-to-HDR reconstruction, extending their application to LF images, with results evaluated through objective quality metrics [124, 125]. Finally, the chapter introduces the "CLASSROOM" LF dataset – a custom-built resource rendered in both HOP and FP formats – to support CNN training and testing in the HDR LF imaging research [116].

As research in projection-based LF visualization advances, understanding the human observer experience remains a significant challenge, primarily due to the lack of standardized testing methodologies. This limitation complicates both experimental design and interpretation. Nevertheless, the limited introduction of LFDs in research institutions has expanded scientific possibilities, offering a solid foundation for studying visualization quality and Quality of Experience (QoE). The immersive, 3D perception capabilities of LF visualization present numerous applications across fields such as cinematography, medical imaging, digital signage, telepresence, and industrial and military uses. Ensuring that the QoE meets or exceeds user expectations is crucial, which is typically achieved through subjective tests focusing on either single or multiple variables. To enhance the user experience of LF visualization, we conducted a series of experiments across multiple LFDs to investigate the factors influencing the overall visual experience. Chapter 6 presents these subjective studies, which examine general aspects and specific use cases, involving participants with both normal and reduced visual abilities. A crucial factor in LF visualization is viewing distance, which, unlike for 2D screens, remains an open question for LFDs; thus, the first experiment explores both the perceptually-supported and subjectively-preferred viewing distances. The second experiment investigates angular resolution and 3D rendering effects on perceived quality in industrial contexts, as these factors are interconnected and warrant joint study. Finally, with rising vision impairment among younger generations, the last two experiments focus on participants with impaired visual acuity, including those with color blindness and one individual with over 90% vision loss

In summary, this thesis comprehensively investigates LF technology, covering essential topics from its foundations to advanced applications. Initially, Chapter 2 offers an extensive literature review on LFs, encompassing the historical background, core properties, capture and visualization methods, as well as other relevant developments. Following this, Chapters 3 to 5 detail novel contributions in LF camera animation, interactive UIs for LFDs, and HDR LF imaging, respectively. Chapter 6 presents subjective assessments of QoE across diverse LFDs, with a focus on enhancing LF visualization quality and accessibility across user profiles. Subsequently, Chapter 7 concludes the thesis, reflecting on the broader implications of these findings, while Chapter 8 highlights new scientific contributions introduced through this research. A detailed statement of contribution is provided in Chapter 8, outlining the specific work conducted in this study. This structure aims to advance LF technology towards practical, high-quality applications.

CHAPTER 2

Light fields

The technical term “light field” was first introduced by Gershun in 1936 [110]. Nonetheless, the original concept was specified more than a century ago as a means to encapsulate the visual information of the physical world, while evolving throughout the years along with the advancements of digital and optical technologies [75, 99, 149, 206]. LF describes the radiance at a point in a certain direction [196]. In other words, it is “the amount of light traveling in every direction through every point in space” [42]. One of the major advantages of LFs is their ability to improve the comprehension of how the HVS interprets the surrounding world. Accordingly, LF images provide tremendous amounts of visual information regarding the represented scenes, as they describe the light traversing in all directions for all the points of 3D space. Therefore, unlike conventional photography –which only captures a 2D image– LF imaging demands the acquisition of multi-dimensional data (i.e., spatial and angular information) [320, 322]. LF in general is meant to represent a portion of 3D space. In terms of visualization, this means that we can imagine a plane in front of and behind the visualized content. In order to characterize the light rays within such constrained portion of 3D space, one may take the two coordinate pairs on these parallel planes where the line representing the light ray intersects them [196]. This, however, poses a notable limitation on the scope of LF visualization, since the constrained portion of 3D space is indeed finite, which means that it is impossible to visualize portions of a content that are virtually infinitely far away.

This chapter examines the concept of LF and provides a comprehensive review of the related literature. It begins with a historical overview in Section 2.1, where the development and evolution of LF concepts are discussed. Section 2.2 offers a detailed explanation of the various representations of LFs and their evolution over time. Following this, Section 2.3 examines the key properties of LFs, highlighting their unique characteristics. Section 2.4 focuses on LF imaging techniques, covering the methods and technologies involved in capturing LFs. Section 2.5 then addresses LF super-resolution, detailing the distinct interpretations and applications of super-resolution techniques specific to

LF data. In Section 2.6, an introduction to LF visualization is provided, offering an overview of various 3D visualization techniques. This section explores different methods for visualizing 3D content, with LFDs discussed as one of the key techniques for achieving immersive, depth-rich visual experiences. Building on this, Section 2.7 delves into LFDs, addressing the challenges and solutions for rendering and displaying LF data. The section also provides a classification of LFDs and explores the various LFDs currently available on the market. Section 2.8 reviews the existing LF datasets, emphasizing their significance for research and development in the field. Finally, Section 2.9 explores LF compression techniques, which are essential for managing the large volumes of data generated by LF imaging. This chapter lays the groundwork for understanding the complex and multifaceted nature of LFs and sets the stage for further exploration in subsequent chapters.

2.1 History of light fields

The question examining the elements of vision has long been imposed, leading to the development of various models to describe how light interacts with objects and how visual information is captured. One of the earliest models is the pinhole camera model, based on the principle that light travels in straight lines. A small aperture (pinhole) allows light rays from an object to pass through and project an inverted image on the opposite surface. This principle laid the foundation for early imaging techniques and inspired the development of more advanced optical models [136].

As imaging technology advanced, stereo imaging became essential in simulating human binocular vision by capturing two slightly different perspectives of a scene. Stereo vision, which relies on stereoscopic cameras, computes depth information through disparity matching, where corresponding points in the stereo images are used to infer 3D structure. This process mimics human depth perception and has enabled applications such as 3D reconstruction and object recognition [135]. The exploration of depth representation led to further advancements, including a deeper understanding of light propagation in space, which ultimately contributed to the development of LFs.

LFs extend beyond traditional imaging by capturing the full radiance of light rays at every point in space, allowing for a richer representation of scene information. Throughout history, several key milestones have shaped our understanding of LFs, arranged in chronological order:

1. The mathematical term "pencil" is used to describe a set of light rays passing through a point in space. Leonardo da Vinci describes these light rays filling space as "radiant pyramids" that intersect and cross one another while having different intensities. Da Vinci added that a pinhole camera can be used to retrieve information on the image at any position [75].

2. Michael Faraday used the term "lines of force" to describe light rays, claiming that LFs are more or less analogous to magnetic fields [99].

3. Frederic E. Ives managed to record parallax stereograms in 1903 by means of a single lens apparatus, which uses one exposure to record two views with respect to the observer [149].

4. LF photography was first introduced by Gabriel Lippmann in 1908. He provided the theoretical foundations for LF photography under the name of "integral photography" [206], and proposed a setup where multiple crystalline lenses are placed hexagonally –similar to a beehive.

5. In 1936, Arun Gershun proposed the term "Light Field" to describe light rays that fill space by means of their radiometric properties [110]. In other words, he described LFs as "the amount of light that travels in every direction through every point in space" [42].

6. The first plenoptic camera was proposed by Edward Adelson and John Wang in 1992, consisting of a single lens and a sensor plane, in front of which a lenticular array was planted [18]. Section 2.4.3 elaborates more on plenoptic cameras.

2.2 Light field representation

Although the concept of LFs has been introduced by Gershun in 1936 [110], attempts to represent the LF function began by the end of the last century.

Figure 2.1 depicts the progression of the LF representation over the years. Each of these representations is discussed in detail in the following subsections.

2.2.1 Plenoptic 7D function

As a means to describe the capability of the HVS to extract geometric information from images, Adelson and Bergen introduced the plenoptic function in 1991 [17]. Plenoptic function describes everything that can be perceived within a given segment of space, hence, the name "plenoptic" (made up from "plenus" meaning complete, and "optic"). The idea behind the plenopic function lies in the fact that the 3D objects constituting the world communicate their properties indirectly to the observer, as the latter takes samples from the light rays filling the space (i.e. plenoptic function) around the 3D objects. In other words, the plenoptic function acts as an intermediary between the world –where physical objects reside– and the eye –where the retinal images of the objects are formed.

The plenoptic function is a 7D parameterized function formulated as $P = P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$, describing the light emitted from an object to the human eye. The main idea of the plenoptic function is to describe the intensity of light viewed from any position, for any wavelength λ, at any given time t. Rather than the source viewpoint, the plenoptic function is calculated in a way that describes all the light rays viewed from the observer viewpoint (V_x, V_y, V_z) with an angle (θ, ϕ) between the light rays and the center of the pupil. Although this function provides a rather accurate description for light rays within a scene, its high dimensionality introduces complexity in calculations.

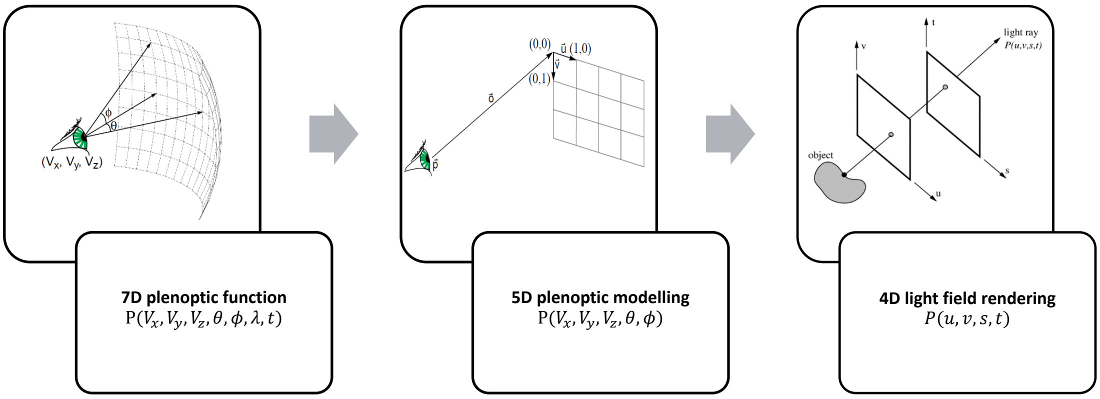


Figure 2.1: LF representation progression [226, 272, 120]

2.2.2 Plenoptic modelling 5D function

Due to the high dimensionality of the plenoptic function, calculations can be extremely complex and hard to process. In an effort to reduce the complexity, McMillan and Bishop [226] reduced the dimensionality of LF representation to 5D by means of plenoptic modelling. Plenoptic modelling is an image-based rendering system, where the plenoptic function undergoes sampling, reconstruction and then resampling. According to McMillan and Bishop, a sample set of the plenoptic function can be used to reconstruct the plenoptic function itself when using image-based rendering approaches. In order to represent a plenoptic sample, McMillan and Bishop suggested using cylindrical projections as they can be easily unrolled to planar maps, which furtherly simplifies the calculations. On the other hand, boundary conditions are introduced at the top and bottom of the cylinder due to its surface being finite, resulting in the limitation of the vertical Field Of View (FOV). For acquiring cylindrical projections, a simple setup consisting of a video camera and a tripod with continuous pan movement is used. This results in undesirable slight panning rotations, which can be approximated by further translating the pixels close to the center of the image. Next, the relative positions between the centers-of-projections across the different acquired cylindrical projections –having different locations in the static scene– are calculated, which can be used later to set the geometric constraints for all possible reprojections. Finally, given the panoramic reference images that are cylindrically projected, as well as, the scalar disparity images for each cylinder pair, image warps can be generated. These warps are used for mapping reference images to planar or cylindrical views, depicting occlusion and perspective effects.

In conclusion, plenoptic modelling considers only static scenes, therefore the time variable (t) from the plenoptic function is omitted, as well as the wavelength (λ), resulting in the 5D LF representation. The concept of plenoptic modelling is based on image-based rendering, where multiple panoramic images captured at different 3D positions are used to represent 5D LFs. Hence, using these pre-acquired images, different views of the environment can be generated. In other words, a complete spherical map is used to represent a full sample of the plenoptic function with respect to a specific viewpoint at a distinct time value, whereas a partial sample of the plenoptic function is represented by a solid angle subset of the calculated spherical map.

2.2.3 Light field rendering 4D function

Later in 1996, Levoy and Hanrahan [196] further reduced the LF representation to 4D in the case of free space (i.e., no occluders), since the radiance along a line remains unchanged unless intercepted. The 4D LF rendering representation is achieved via light slab, by means of the parametrization of light rays using their intersections with two planes while travelling in straight lines. In other words, the two intersection point pairs (u, v) and (s, t) on the planes are used to represent LFs. Another alternative is to parametrize the line by means of a point and direction, in case of placing one of the planes at infinity.

Accordingly, an LF scene is constructed by rendering multiple 2D images, where each image depicts a 2D slice inserted into the 4D LF representation. In the case of HOP visualization, the array of images is a 1D horizontal array; whereas for FP imaging, a 2D array of images is rendered to create an LF scene.

2.3 Light field properties

2.3.1 Field of view, valid viewing area

One of the important aspects of LFs is the FOV, which in turn determines the angle of the VVA, within which any number of spectators can fit to view the visualized content [159]. Based on the FOV, the baseline of the LF system is determined.

2.3.2 System baseline and parallax

As previously stated, the FOV determines the baseline of the LF system, which can either be a narrow- or wide-baseline system. System baseline denotes the distance between the extremes of the FOV [71]. In other words, considering the LF system, baseline is the maximum distance between perspective changes. In industrial practice, an FOV between 10° and 15° is considered a narrow-baseline system, while a wide-baseline device usually have an FOV greater than or equal to 30°. However, at the time of writing, there is no scientific-community-wide consensus regarding this classification [117, 120].

Table 2.1: Comparison between narrow-baseline and wide-baseline LF cameras [117]

	Narrow-baseline LF cameras	**Wide-baseline LF cameras**
Baseline length	Measured in centimeters (less than 1 meter)	More than 1 meter
Reconstruction accuracy	Limited and can lead to sub-pixel feature disparities	Better
Depth map estimation	Limited	Better
Spatial resolution	Deteriorated	Enhanced
Portability	Relatively portable	Not portable

The baseline for a camera system arranged as a linear array is the Euclidian distance between the leftmost and the rightmost camera. Considering HOP LF cameras, they can be furtherly differentiated into narrow-baseline (baseline shorter than 1 m) and wide-baseline. Due to having smaller baselines, narrow-baseline LF cameras are more portable compared to wide-baseline LF cameras [71]. On the other hand, reconstruction accuracy is limited in narrow-baseline LF cameras since accuracy is linearly proportional to the baseline. Reconstruction accuracy refers to how precisely depth can be estimated from captured images, with larger baselines generally providing more reliable depth information. Moreover, narrower baselines lead to "sub-pixel feature disparities", resulting in the deterioration of spatial resolution [28, 192]. Since both the angular and spatial information are captured by LF cameras for light rays, LF images provide easier methods for depth map estimation [76]. Analogous to the aspect of accuracy, wide-baseline LF cameras are better in depth map estimation since baseline is inversely proportional to the depth estimation error [28, 117]. Table 2.1 sums up the differences between narrow-baseline and wide-baseline LF cameras [117].

Regarding LFDs, a baseline describes the distance between the extremes of the FOV [71]. LFDs provide a naturally wide baseline due to their large screen size $S_{x,y}$, optimal observer distance $D_{observer}$ (usually 1 to 4 m, depending on screen size and the choice of vertical perspective [159]), and outward facing light emission angle $FOV_{xdisplay}$ (45° to 170°). For HOP systems with a planar screen –as seen in Figure 2.2– the baseline $B_{xdisplay}$ corresponding to an LFD can be calculated as [117]:

$$B_{xdisplay} = 2 * D_{observer} * tan(\frac{FOV_{xdisplay}}{2}) + S_x \tag{2.1}$$

Figure 2.3 illustrates baseline configurations for an LFD and a camera system arranged in a linear array. More on the different baseline LFDs is explained in Section 2.7.

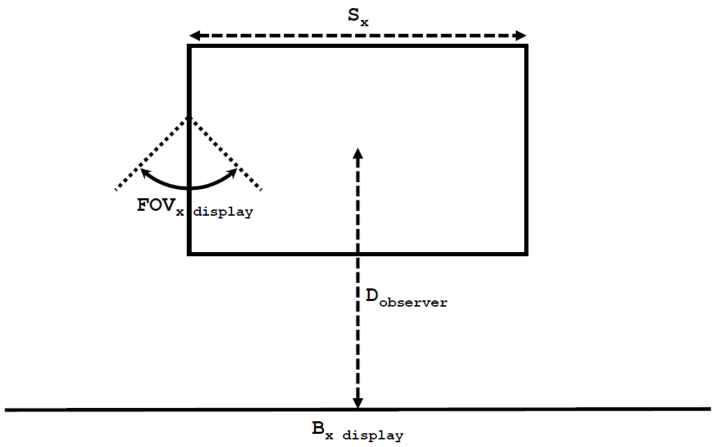


Figure 2.2: View of the display setup to calculate the baseline [117]

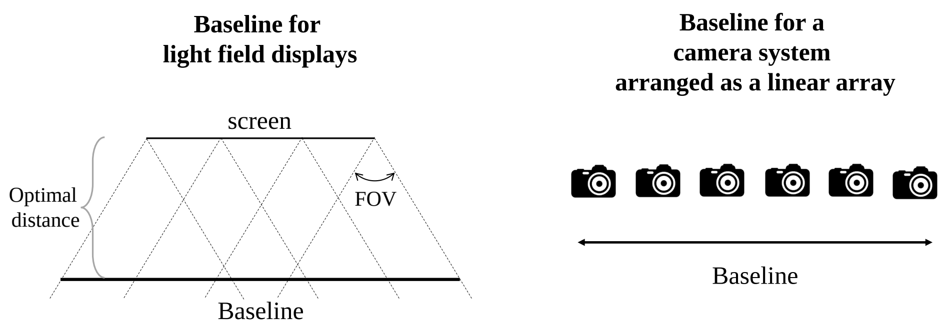


Figure 2.3: Baseline configurations for an LFD and a camera system arranged in a linear array

Parallax is one of the LF properties dependant on the baseline. It denotes the change in the perspective of LFs. Three different types of parallax exist, depending on how the baseline is extended: HOP, Vertical-Only-Parallax (VOP) and FP. HOP considers the change in the angular perspective horizontally (i.e. baseline extends horizontally), which is similar to the change in the human's perspective since our eyes are horizontally separated. On the contrary, VOP is not practical, as the baseline extends only in the vertical direction. Lastly, the FP LFs change the angular perspective both horizontally and vertically and thus, they are far more challenging in the respect of design and implementation.

2.3.3 Angular resolution

Angular resolution is technically the smallest measurable angle of change reproduced by light rays relative to a single point on the screen [181, 164]. If the angular resolution of the visualization is insufficient, adjacent distinct sections of the visualized model may experience crosstalk, significantly diminishing the perceived quality. Angular resolution can be expressed in two interchangeable formats. The first format adheres to the previously mentioned definition, where angular resolution is quantified as an angle, typically measured in degrees. A smaller angle corresponds to a higher angular density, leading to improved visualization quality. Within the context of LF research, a higher angular resolution is represented by a smaller angle. For instance, an angular resolution of 0.5 degrees is superior to that of 1 degree. The alternative format describes angular resolution based on the number of source views used to create and display the LF within a given FOV. This format is measured in views per degree. For instance, if a system with a 45-degree FOV is calibrated with 90 source views, the angular resolution would be 2 views per degree. Both formats are directly interchangeable: for example, 2 views per degree corresponds to an angular resolution of 0.5 degrees, whereas 1 view per degree equates to an angular resolution of 1 degree [164].

2.3.4 Region of interest

This feature is concerned with LFDs. Region Of Interest (ROI) defines a region of space that is box-shaped in the virtual scene, within which everything is visible on the LFD [81].

2.4 Light field imaging

LFs incorporate spatial and angular information of light rays. Although conventional cameras record most of the scene information, light distribution that is penetrated from the world is mostly not recorded. LF imaging techniques, however, have the capability to re-capture the aforementioned lost information by capturing 4D LFs through acquiring the 2D position on the image plane, along with the 2D incident direction. LF imaging process includes calibration, 3D depth estimation and resolution enhancement. Capture LF hardware allows the storage of almost all the information of the viewed scene from the camera's point of view, where the amount of light in each light ray arriving at the camera sensor is recorded [238, 112, 42]. This information can be useful in further applications requiring additional knowledge about the visualized scene. For more information about the scene, dynamic cameras can be used for navigation.

Nowadays, LF capture systems can be either HOP or FP systems, with the latter capturing the parallax in both directions. Regarding the baseline of the LF capture system, it can either be a narrow- and wide-baseline system.

According to Wetzstein [320], LF acquisition can be classified into three main categories: (i) multiple sensors, (ii) temporal multiplexing and (iii) spatial and frequency multiplexing. In this section, we discuss each of these acquisition methods in detail.

2.4.1 Multiple sensors

This set of methods sets up camera arrays for wide-baseline capture, where multiple cameras arranged in specific configurations are used to capture the same scene in a synchronized manner from multiple perspectives. Each captured image represents a 2D slice constituting the final 4D LF [196]. In the case of HOP wide-baseline capture, cameras are placed horizontally in a linear manner (e.g., the LF transmission of Balogh and Kovacs [40]) or an arc manner (e.g., the telepresence system of Cserkaszky *et al.* [68]). For FP wide-baseline systems, cameras are arranged in a 2D grid (e.g., a 64-camera setup arranged in an 8×8 grid [326]) or spherically (e.g., spherical LF camera using Gaussian blending method for vision reconstruction [249]). Camera arrays can be built in various configurations from any type of industrial camera that has a synchronization port. For example, LF camera arrays are offered by Fraunhofer IIS[1].

Due to their ability to capture multiple images with broad range of distances, this set of methods generate high spatial resolution LFs by enabling efficient depth reconstruction. On the other hand, portability issues are encountered due to the significant weight of the camera setup. The portability of camera arrays can be generally difficult, especially if they are used outdoors due to their reliance on power supply, as well as the usage of complicated transmission lines [244]. Constant synchronization and calibration should be maintained across the different cameras within the system, as well as managing the storage and processing of the huge amount of captured data. Many solutions were proposed for the calibration of camera arrays, such as simply calibrating each view point by means of a single camera calibration. An additional solution is the usage of plane plus parallax for calibrating the camera array used for LF acquisition [324]. A major limitation to using multiple cameras is the limited view resolution, where the physical dimensions (i.e., size of the camera) and limitations (e.g., constrained speed and degrees of freedom of the rig) restrict the gaps between the cameras when being placed beside one another, along with the possibility of self-capture [320, 112, 42, 71, 120].

[1]https://www.iis.fraunhofer.de/en/ff/amm/for/forschbewegtbildtechn/lichtfeld.html

2.4.2 Temporal multiplexing

An alternative to using multiple cameras for wide-baseline capture is temporal multiplexing, achieved by a single sensor. A multitude of possible solutions was devised to capture multiple images via single camera including placing the object of interest on a turntable, or moving the camera while reorienting it towards the object of interest over a spherical or a planar path [196, 114, 285]. Other solutions include programmable aperture photography [202], extension of integral photography [18, 240, 238], rotation of a planar mirror [148], lensless LF camera [333], and many others.

Unlike using multiple sensors, temporal multiplexing has the advantage of significant reduction in costs and complexity, where a single camera is used, requiring less calibration. On the other hand, the scene is required to be static, hence, the utilization of this method is rather limited, less universal, and thus, less practical [120].

2.4.3 Spatial and frequency multiplexing

This category for LF acquisition aims at solving the problems encountered in the previous two methods. Possible solutions include using a high-speed camera or multiplexing a 4D LF onto a 2D image. Unlike high-speed cameras, multiplexing (i.e., single-shot multiplexing) can be used for video recordings, where a single sensor is used to capture 4D LFs in a single shot. Types of multiplexing include spatial and frequency multiplexing associated with spatial and spectral characteristics, respectively [320, 112, 42].

According to Wetzstein [320],"spatial multiplexing produces an interlaced array of elemental images within the image formed on the sensor". In order to achieve spatial multiplexing, a multitude of solutions was proposed. Early efforts included parallax barriers and integral photography introduced by Ives in 1903 [149] and Lippmann in 1908 [206], respectively. Later, a single camera with a 2D MicroLens Array (MLA) was used. Whereas, the MLA can be placed at different locations in-between the main camera lens and the image sensor, the most common is the one in which the MLA and image sensor are placed at the focal planes of the main lens and the MLA, respectively. This approach captures $N \times N$ perspectives by means of the sensor elements located under each microlens. The resulting sub-images can then be processed to generate a collection of $N \times N$ images, each corresponding to a distinct viewpoint with a spatial resolution of $K \times K$ [42]. Figure 2.4 illustrates the architecture of the LF camera, while Figure 2.5 highlights the duality between the camera array and MLA methods. Other solutions included the utilization of external lens arrays [107, 297, 298], array of planar, tilted mirrors or mirrored spheres [299, 195, 185, 286], lens arrays and a single sensor in related compound imaging systems [239, 293, 294, 137], and combining a lens array and a flatbed scanner in a lenslet-based architecture [325].

Among the different equipment used to capture 4D LFs via spatial multiplexing are plenoptic cameras, where the MLA is positioned in front of the camera sensor [258]. In other words, a plenoptic camera is a specific type of camera that incorporates MLA as an integral part of its design. The plenoptic camera is trivially named after the plenoptic

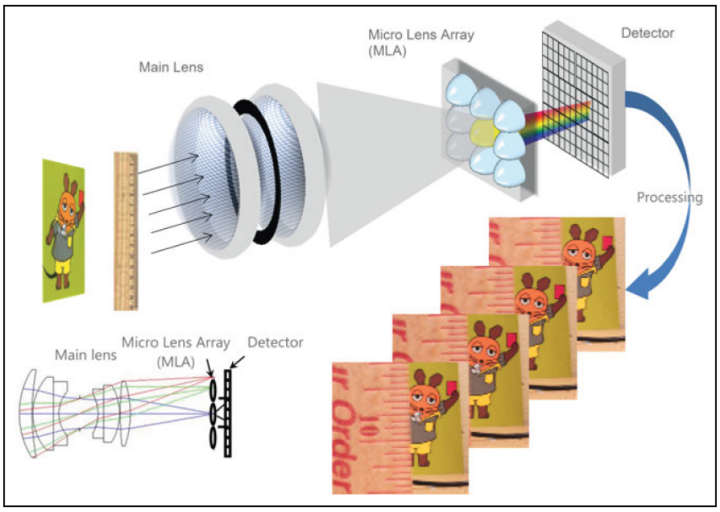


Figure 2.4: Architecture of LF camera. The raw image is processed to generate sub-images corresponding to different views [42].

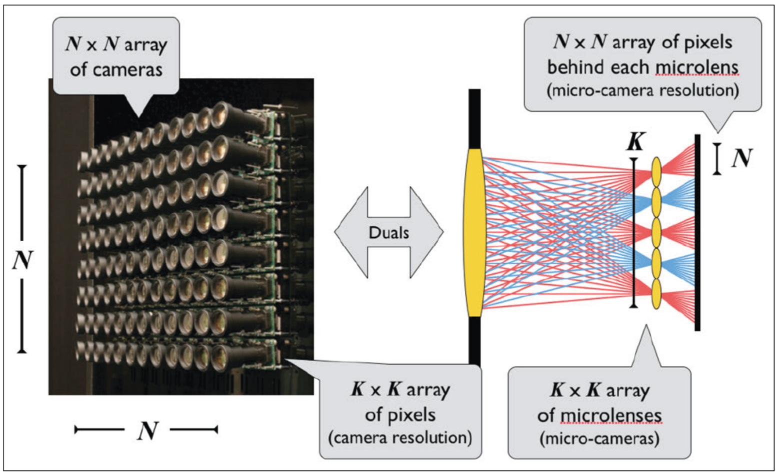


Figure 2.5: Duality between capture methods: a camera array (left) and a single sensor with an MLA (right) [42].

function itself introduced by Adelson and Bergen [17]. Unlike the case of conventional cameras where an object space point is projected onto a single pixel, a light ray emitted from a point is projected to many positions on the plenoptic camera sensor [232]. The first plenoptic camera was proposed by Adelson and Wang in 1992, consisting of a single lens and a sensor plane in front of which a lenticular array is planted [18]. Plenoptic cameras were made commercially available by Lytro (until 2018), and are still available for purchase from Raytrix[2].

Another means to capture LFs via single-sensor is using frequency multiplexing, introduced by Veeraraghavan *et al.* in 2007 [305]. This is accomplished via optical heterodyning by using light-attenuating, non-refractive masks, located at a slight distance in front of the conventional sensor. The 4D Fourier transform of LFs is encoded into multiple spatial-angular bands within the Fourier transform of the 2D sensor image. Methods to achieve optical heterodyning included sum of sinusoids pattern [305], tiled-broadband patterns [186], and adaptive mask patterns [304].

Unlike the techniques utilizing multiple sensors, this method overcomes the portability issue, while generating high dense views. Additionally, the generation of multi-spectral content is possible by using a multi-spectral filter in the capture process, located in front of the main lens. While dynamic scenes can be captured efficiently by means of spatial multiplexing, a trade-off occurs between the angular and spatial sampling rates. Accordingly, compared to the previous methods, reduction in the spatial resolution is noticeable [320, 112, 42]. Moreover, in most cases, the baseline between the captured views is small due to the small distance between the microlenses in the MLA setup. Thus, the majority of the methods within this category are classified as narrow-baseline [120].

2.4.4 Recap

In conclusion, capture LF hardware allows the storage of almost all the information of the viewed scene from the camera's point of view, where the amount of light in each light ray arriving at the camera sensor is recorded [238]. This information can be useful in further applications requiring additional knowledge about the visualized scene. Nowadays, LF capture systems can be either HOP or FP systems, with the latter capturing the parallax in both directions. Considering the baseline length of LF capture and display devices, they can be categorized as narrow- and wide-baseline systems. Table 2.2 summarizes the different types of LF acquisition [120].

2.5 Light field super resolution

A common research topic within the scientific community of LF technology is super resolution. However, as there are two distinct interpretations for the same terminology,

[2] https://raytrix.de/

Table 2.2: LF acquisition types and methods [120]

LF acquisition type	**Definition**	**Acquisition methods**	**Examples**
Multiple sensors	Camera arrays for wide-baseline capture	- Linear camera setup - Arc camera setup - 2D grid camera setup - Spherical camera setup	[40] [68] [326] [249]
Temporal multiplexing	Uses a single camera instead of multiple cameras for wide-baseline capture	- Camera on turntable or rotating camera while reorienting - Programmable aperture photography - Extension of integral photography - Rotation of a planar mirror - Lensless LF camera	[196, 114, 285] [202] [18, 240, 238] [148] [333]
Spatial and frequency multiplexing	Uses a single camera to create LF images by means of spatial or frequency multiplexing	- Parallax barriers - Integral photography - External lens arrays - Array of planar, tilted mirrors or mirrored spheres - Lens arrays and a single sensor in related compound imaging systems - Combining a lens array and a flatbed scanner in a lenslet-based architecture - Plenoptic cameras - Frequency multiplexing	[149] [206] [107, 297, 298] [299, 195, 185, 286] [239, 293, 294, 137] [325] [18, 13] [305]

it requires clarification via a prefix. Among the published works, the most frequent interpretation of super resolution is image resolution enhancement. As this may be considered the default interpretation, the terminology is often used without a prefix. An accurate prefix for such may be spatial super resolution, but image super resolution also describes the notion faithfully. Different methods were devised to achieve LF image super resolution, including projection-based methods [237, 204, 108, 203] and optimization-based methods [46, 230, 315, 261, 260, 27, 100]. A great number of novel attempts employ CNNs and deep networks, specifically targeted for data captured by LF cameras, since such devices have limited spatial and angular resolutions. These networks aiming to achieve spatial super resolution for LF images include a two-stage CNN, exploiting the correlations among the LF images both internally and externally [98]; a bidirectional recurrent network [314]; a deformable convolution network, taking into account the angular information among images while handling disparities [313]; residual networks, where the LF images are first grouped and then fed into different network branches from which the residual information along different directions is calculated [335]; and an algorithm applying optical flow to align LFs, after which the angular dimension is reduced by means of low-rank approximation, and then, a deep CNN is used for spatial super resolution [101]. Additionally, among the other methods of spatial super resolution are the LF-DFnet (deformable convolution network) [313]; the LF-IINet (intra-inter view interaction network), preserving the system parallax while exploiting the correlations among images [210]; dense dual-attention networks [231]; and end-to-end networks using epipolar geometry, in order to learn the details of sub-pixels per view image [334]. In efforts to reduce the dimensionality, the complexity, and the cost of 4D LF data, the work of Van Duong *et al.* [303] proposes a network that decomposes LF data into a lower data subspace while exploiting the information resulting from the possible 4D LF representations, including Epipolar Plane Image (EPI), as well as spatial and angular information. Regarding networks enhancing both spatial and angular resolutions, the work of Yoon *et al.* [329] proposes the LF CNN (LFCNN), composed of spatial and angular super resolution networks. Furthermore, LF-InterNet [312] enhances both the spatial and angular resolutions by extracting their features from LFs separately, with interactions occurring between them later, ending up by fusing the interacted features. Another method uses two super resolution networks, targeted for spatial and angular super resolutions separately, generating multiview features that are later remixed by an Adaptive Feature Remixing (AFR) module [179, 120].

The other interpretation of super resolution is angular super resolution [120]. This is furtherly elaborated in Section 2.7.

2.6 Introduction to light field visualization: An overview of 3D visualization techniques

To comprehend LF visualization, it is essential first to review the history of 3D visualization, tracing its evolution from glasses-based to glasses-free displays. This includes examining the various types of displays and the technologies underlying them. Figure 2.6 provides an overview of the 3D display technologies.

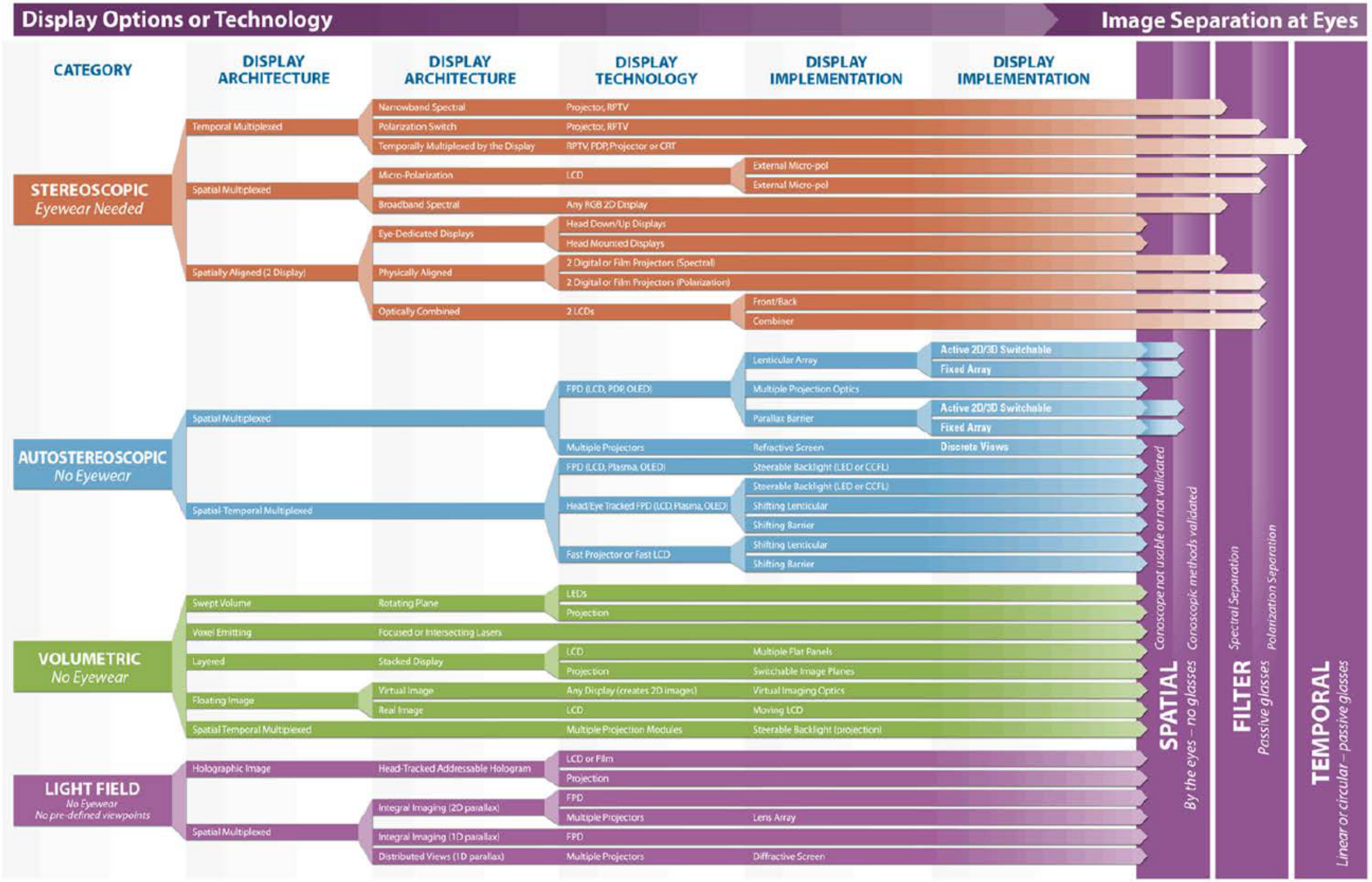


Figure 2.6: 3D display technologies [38]

2.6.1 Glasses-based 3D displays

The term "Stereopsis" is derived from the Greek "stereós" and "ópsis", meaning "solid" and "power of sight", respectively [248]. Stereopsis is based on binocular disparity and binocular cues. Stereoscopic 3D (S3D) displays are built on the concept of stereopsis–two 2D images represent the content simultaneously from two similar yet different perspectives (analogous to the human eyes), and a viewing apparatus (e.g., 3D glasses) allocates one image to each eye, generating the perception of depth [248, 225]. Based on the apparatus, one of the following approaches could be used to deliver these images to the eyes of the observer: (i) color multiplexing, (ii) polarization multiplexing, or (iii) time multiplexing. A common example for color multiplexing is the utilization of anaglyph images, where both the left and right images are combined by means of a complementary color coding method. Although plausible results may be achieved via anaglyph glasses, there is a possibility for losing color information, as well as to be affected by crosstalk (i.e., interference). To overcome the aforementioned problems, a multitude of solutions were suggested, including adjusting the depth map, aligning images, and blurring color component [145, 146]. An example for a color-multiplexed approach is the ColorCode 3D technique –commonly used in movies and video games–generating full-color images while working with standard hardware at lower costs [277]. In the case of polarization multiplexing, for a stereo image pair, each image has its light's State Of Polarization (SOP) mutually orthogonal. Such solutions rely on visual gears with polarizers, aiming to block the image unintended for the given eye.

VR is defined as immersive 3D environments that are virtually generated by computers. These are usually interactive environments, incorporating multiple sensory channels (e.g., position, touch, etc.). Navigation of the generated environments may also be possible, which is crucial for real-time simulations [129, 52, 259]. With such potentials and advantages, VR has been incorporated in various fields, such as tourism, medicine, military, sports, physical education, virtual stores, training and education, and many more [121]. The concept of VR dates back to 1968 when Ivan Sutherland developed the first VR system, featuring wire-frame graphics and a Head-Mounted Display (HMD) [283]. Examples for VR displays include HMD with a narrow FOV and a 3-Degrees Of Freedom (DOF) tracker, HMD with a wide FOV and a 6-DOF tracker [189], desktop-based VR systems [31], Cave Automatic Virtual Environment (CAVE) [66], and "Fish Tank" VR, which is characterized by a stereoscopic image of a 3D scene displayed on a monitor, with the perspective projection adjusted based on the observer's head position[318].

Unlike VR –which provides a complete 3D environment– AR generates only the overlays (i.e., virtual imagery information) over real environments, with a possibility of doing so in real time [35, 338, 144]. This capability of combining computer-generated visuals together with real environments made AR a viable tool for training and education. The work of Azuma *et al.* [35, 34] summarizes AR via three criteria: (i) it is a combination of both real and virtual components, (ii) real-time interactions are allowed, and (iii) registration in 3D [121]. The integration of virtual content into the real environment categorizes AR displays into many different types [284] including optical see-through displays [250, 242], video-based displays [104], projection-based displays [173, 158], eye-multiplexed displays [79], head-attached displays [180, 229], HMDs [103, 97, 256, 44], body-attached and handheld displays [177, 178].

2.6.2 Glasses-free 3D displays

Unlike S3D, VR and AR, glasses-free 3D displays do not require additional viewing gear. Among glasses-free 3D displays are the autostereoscopic displays, generating images with the needed disparity. They are autostereoscopic due to the fact that they provide different perspectives for the two eyes of the observer or user without relying on viewing devices. Such systems can be either two-view or multi-view displays. In the case of two-view autostereoscopic displays, a single stereo pair of parallax views is generated. The image pair can be generated either at a single location (for a single viewer) or in multiple points of space (for multiple viewers). In order to achieve stereoscopy, the viewer needs to be in the right position within the range of ideal distance from the screen. Two-view autostereoscopic display systems can be parallax-barrier-based or lenticular systems [300]. Regarding multi-view autostereoscopic displays, multiple stereo image pairs are generated for various locations (also known as "sweet spots") within the viewing area of the display. A major limitation for autostereoscopic devices is the location requirement of the spectators [262]. It should also be highlighted that the content is repeated over the different viewing locations; the same perspective is provided to each and every spectator. Other types of glasses-free 3D displays include volumetric and holographic displays. Volumetric displays generate volume-filling 3D visual representations, where light is emitted by voxels –located in 3D space– in the areas where they appear [102, 48]. Volumetric displays have proven their efficiency in many fields, including medicine, military, and engineering [102]. Considering holographic displays, holography was introduced in 1948 by Dennis Gabor [106, 105] (for which he was awarded the Nobel Prize in Physics in 1971); however, the generated holograms proved to be of poor quality. In order to enhance the holographic quality, various works were carried out [193, 86], with the idea of digital holograms introduced in 1967 [113]. Later in 1980, the fundamental theory for digital holography was proposed by Yaroslavskii and Merzlyakov [327, 121].

Glasses-free 3D visualization comes with two evident advantages. First of all, no viewing devices are necessary to view the visualized content –as the term "glasses-free" suggests. In contrast, the maximum number of simultaneous spectators in the case of S3D visualization is limited to the number of 3D glasses. Hence, the other major advantage is that multiple spectators may view the display in operation simultaneously. However, if we take for instance multi-view displays, the different perspectives that one individual may perceive is very limited. Similarly to multi-view displays, LFDs deliver a glasses-free 3D viewing experience to multiple viewers simultaneously [121]. Further details on LFDs are covered extensively in Section 2.7.

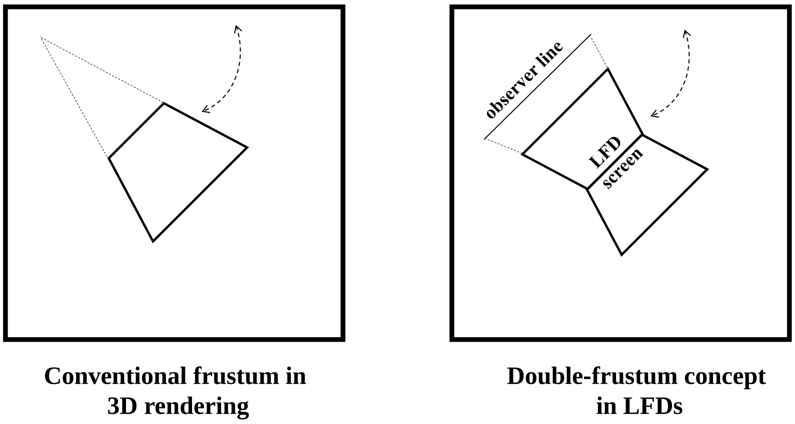


Figure 2.7: Comparison between conventional frustums in 3D rendering and the double-frustum concept in LFDs

2.7 Light field visualization: Light field displays

LFDs convey realistic visual experiences to spectators via the natural, glasses-free 3D perception of the content. In other words, an LFD acts as a window to the 3D world described by the corresponding light rays [42]. Enabled by parallax barrier and integral imaging [321], LFD technology represents a breakthrough in 3D visualization, requiring extensive data due to its need for capturing scenes from multiple angular perspectives. This corresponds to the 4D function representation of LFs in the case of free occluder space, where the spatial and angular information are both recorded representing the different perspectives of the scene from multiple viewing points [116].

Double viewing frustums - inside which the elements within the ROI are visualized –are employed in LFDs in a manner analogous to viewing frustums used in general 3D rendering [282]. Figure 2.7 illustrates the contrast between frustums used in conventional 3D rendering and the double-frustum concept in LFDs. The ROI defines a box-shaped volume in the virtual scene, within which all contents are visualized on the LFD [81, 82]. For projection-based LFDs, the characteristics of the ROI can have considerable effects on the perceived visualization quality, where poorly chosen ROI values can greatly deteriorate the quality of the displayed content [81]. The ROI plays a major role in generating LF camera animations virtually, the topic of which is discussed in detail in Chapter 3. In the concept of double viewing frustum, the screen is placed between the frustums. Contents rendered around the screen are considered to be in the sharp region of the LFD, and hence, rendered sharply. However, contents further away from the screen enter the blurry region of display, and thus, suffers blurriness, resulting in lower perceived quality [120].

Similarly to multi-view displays, LFDs provide a glasses-free 3D visual experience to multiple spectators simultaneously, with the important distinction that the display may use its entire FOV to provide a single continuous parallax effect. It needs to be emphasized that FOV in this context is measured from the screen of the display, and not from the user's perspective (e.g., as in the case of VR). Moreover, any number of simultaneous viewers may be accommodated, as long as they can fit inside the VVA, determined primarily by the FOV of the display. Among the most important Key Performance Indicator (KPI)s of LF visualization are spatial resolution (which, if the content is generated from a series of 2D images, corresponds to their resolution), angular resolution (which is technically the density of distinct light rays), screen dimensions, depth (which is "depth budget" as a display attribute), brightness, contrast, as well as FOV [171, 121]. These KPIs align with the objectives outlined by the Turing test [296] for LF visualization. Originally proposed for 3D displays within visualization technologies, the Turing test represents the ultimate goal of LF visualization –to become indistinguishable from reality [43]. Hamilton *et al.* [131] created a framework for an LFD-related Turing test, in which certain visual characteristics must be reached to satisfy human visual acuity [140, 73]. In essence, passing the visual Turing test requires both high image resolution and angular density [120].

Though available, LFDs are not yet widely adopted in the consumer market due to technical challenges, with advances continuously introduced by researchers and industry professionals. For instance, the recent study by Balogh *et al.* [39] showcases a prototype for a 3D LF Light-Emitting Diode (LED) wall, essentially representing a glasses-free 3D display that can be tailored to any size, aspect ratio, or shape [163]. Additional information on currently available LFDs is provided in Section 2.7.2.

LF visualization is an emerging 3D technology that does not rely on viewing devices. This capability enables numerous users to experience 3D content at the same time, as the continuous and smooth motion parallax facilitates a wide variety of viewing angles. However, despite this significant advantage in viewing flexibility, the accurate perception of the visualized material remains contingent on the observer's position [159]. Hence, the standardization of LF QoE is crucial, as outlined in the current IEEE recommendation [147], which establishes a viewing distance threshold. This viewing distance threshold (i.e, recommended maximum viewing distance) beyond which visualization becomes more 2D than 3D, is defined as

$$VDT = \frac{ID}{tan(AR)}, \tag{2.2}$$

where ID is the average interpupillary distance, and AR is the angular resolution of LF visualization. This threshold defines the maximum theoretical distance at which two distinct light rays with respect to a single point on the screen of the display may address the two pupils –hence, enabling proper 3D perception of the visualized content without the need for movement (e.g., sideways movement in the case of HOP visualization).

As stated in Section 2.5, the other interpretation of super resolution is the angular super resolution. It refers to an angular density so high that not only two distinct light rays with respect to a point on the LFD screen address the two pupils of the observer –which is essential to the 3D visual experience [169]– but also a single pupil. Based on the state-of-the-art LF visualization technology and its current usage, angular super resolution has not yet been achieved. The reason why the word “usage” is involved in this statement is that angular super resolution evidently depends on the viewing distance as well. After all, the farther the observer, the lower the perceivable light ray density –making LF visualization appear flat 2D beyond certain distances. The most significant benefit of reaching angular super resolution is that it allows observers to change their focal distance. While with lower angular resolution, one may only focus on the plane of the screen of the LFD, with angular super resolution, one may focus on closer and farther portions of the visualized content. Although achieving such a goal may greatly benefit LF use cases, it poses great challenges on multiple fronts [120].

Equation 2.2 accounts for the average interpupillary distance as 6.5 cm. However, in the context of angular super resolution, the interpupillary distance should be replaced by pupil size, since two distinct light rays with respect to a single point on the screen address only one pupil in such case. In subjective tests that aim to study angular super resolution, various lighting conditions and display brightness values should be investigated, as the size of the pupil typically varies between 2 mm and 8 mm, depending on the intensity of light. The novel results that are to be obtained by subjective studies may provide the foundations of new standards of LF QoE [120]. In light of this, Equation 2.2 can be adjusted by replacing the average interpupillary distance with the diameter of the pupil–which rages between 2 mm and 4 mm in bright light and between 4 mm and 8 mm in the dark [96, 308, 33, 50]– Equation 2.2 for viewing distance threshold can be reformulated for angular super resolution as follows:

$$VDT_{SR} = \frac{PD}{tan(AR)}, \tag{2.3}$$

where PD is the diameter of the pupil [120].

2.7.1 Classification of light field displays

LFDs can be either HOP or FP. Since the eyes are horizontally separated, HOP LFDs are more practical than the implementation of VOP displays, in addition to being less complex than FP solutions. FP displays can visualize contents recorded by FP cameras, while HOP displays evidently need to select a subset of the content. In practice, high-quality visualization demands that the capture device and the display device have LFs that match as much as possible [116, 117, 120].

Regarding HOP LFDs, farther viewing distances are only possible with sufficient corresponding angular resolution [169] –as insufficient light ray density makes the visualized content look flat 2D, as no two or more distinct rays with respect to a given point on the screen can reach the two pupils– which also extends the VVA, the angle of which is originally determined by the FOV. The FOV itself is determined by the baseline of the system. Theoretically, an LFD is considered to be a narrow-baseline system when the

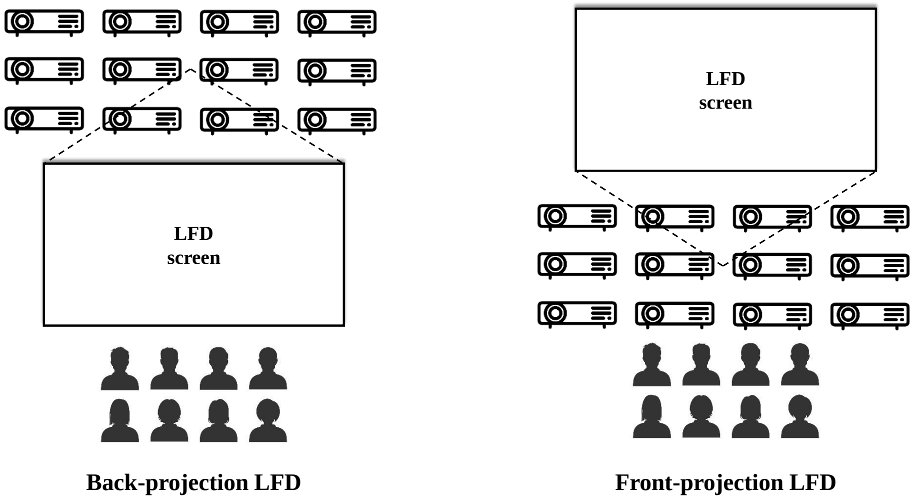


Figure 2.8: Back-projection and front-projection LFDs [120]

FOV ranges between 10° and 15°, whereas for FOV values greater than 30°, the LFD is counted as wide-baseline system [116]. However, at the time of writing, there is no scientific-community-wide consensus regarding this classification. A common example for narrow-baseline displays with small FOV, are HMDs, targeting a single user at a time. Whereas for wide-baseline LFDs, FOV is considered the most expensive part in their design, as it targets more spectators, thus, being more practical compared to narrow-baseline systems. Accordingly, wide-baseline LFDs have a better immersion experience, even though they are more challenging in their design. In their attempts to create wide-baseline LFDs, *Holografika Kft.* has managed to create the HoloVizio C80 cinema LFD[3] with an FOV of 40°, the HoloVizio 722RC LFD[4] with an FOV of 70° and the HoloVizio 80WLT LFD [5] with a full angle 3D display (i.e., an FOV of 180°) [117, 120].

Moreover, exploration into projection-based LF visualization is flourishing. In the case of projection-based LFDs, the location of the projector array with respect to the screen and the observers creates two categories as well. If the projectors are on the same side of the screen as the observers, then it is a front-projection LFD, and if they are on the other side, then it is a back-projection LFD. The two types of projections are illustrated in Figure 2.8. Note that in the case of front-projection solutions, the projectors are typically above the viewers, and they may also be located behind the viewers [120]. Projection-based LFDs have been effectively deployed, including notable examples such as the HoloVizio displays. These displays employ a holographic screen and a series of optical modules to emit light beams. The resulting 3D view is constructed by the holographic screen, where

[3] https://holografika.com/c80-glasses-free-3d-cinema/
[4] https://holografika.com/722rc/
[5] https://holografika.com/80wlt/

these beams converge [41].

2.7.2 Current light field displays

The vast majority of LFDs at the time of writing are HOP. Table 2.3 examines the characteristics of different LFDs with regards to angular resolution and the associated maximum viewing distance for angular super resolution. Additional properties (e.g. number of projectors) are missing, as this information is not disclosed by companies. For this analysis, we used 8 mm for the threshold calculation in Equation 2.3, as any viable viewing distance should be smaller than what is listed in the table. Note that the table does not contain LFDs that are glasses-based (e.g., AR LFD [275]) and those that do not have their angular resolution values precisely specified (e.g., those that simply state that "hundreds of views are supported" [2, 3]). The angular resolution values are expressed in the degree format, which means that lower values indicate higher density for distinct light rays. As explained earlier, the feasible viewing distance for the utilization of the LFD in a given context should be smaller than the maximum viewing distance for super resolution, as it assumes the darkest lighting conditions. If the diameter of the pupil was taken as 4 mm, then the values in the table would be halved. Even with 8 mm, the maximum viewing distance does not exceed 1 m. While such displays may be used in single-user scenarios, there are other factors that may impact feasibility. For example, in the case of the HoloVizio C80 [6], as the large-scale LFD is a front-projection display, viewing the screen from the distance specified in Table 2.3 would potentially result in invalid LF, as the observer's body would block the light rays coming from the projector array. Additionally, FOV and screen dimensions are listed in the table as well, since they are also crucial to feasibility. For instance, if the viewing distance threshold is low, than a low FOV can severely limit the VVA –and thus the maximum number of simultaneous observers and their mobility– which is already constrained. Regarding screen dimensions, large screens are not necessarily made to be observed from a close distance, and may come with other limitations, as exhibited previously. Based on the available information and the analysis above, we can conclude that LFD solutions at the time of writing are not feasible for use cases that incorporate angular super resolution [120].

2.8 Light field datasets

Given its growing significance in numerous applications, LF imaging has become crucial in various research domains. As stated at the beginning of this chapter, LF imaging captures more information about a scene compared to conventional imaging, since spatial and angular information are both recorded [116]. Hence, LF datasets are significantly larger than conventional datasets, as a single scene is represented by multiple images [95]. Such datasets vary a lot in their characteristics and provided information. They can contain real-world captured contents, synthetic ones (i.e., rendered), or a combination of

Table 2.3: Current LFDs along with their specifications [120]

LFD	Angular Resolution	FOV	Screen Dimensions	Maximum Viewing Distance for Super Resolution
Lume Pad 2 [12]	10.75°	86°	12.4”	4.21 cm
HoloVizio 80WLT [5]	1°	180°	30”	45.83 cm
HoloVizio 640RC [41]	0.8°	100°	72”	57.29 cm
Looking Glass Portrait [11]	0.58°	58°	7.9”	79.03 cm
Looking Glass Go [10]	0.58°	160° overall / 58° optimal	60”	79.03 cm
Looking Glass 65” [9]	0.53°	53°	65”	86.48 cm
Looking Glass 32” Spatial Display [8]	0.53°	53°	32”	86.48 cm
Looking Glass 16” Spatial Display [7]	0.53°	150° overall / 53° optimal	16”	86.48 cm
HoloVizio 722RC [4]	0.5°	70°	72”	91.67 cm
HoloVizio C80 [6]	0.5°	40°	140”	91.67 cm

both [120]. Numerous efforts have been made in creating LF datasets. Notable examples include the multiview HOP dataset, which features a single object in each scene [287], the SMART dataset comprising 15 LF images [243], the dense LF dataset, which contains 14 scenes, each captured with 5 synthetic images [21], the VALID dataset [307], and a dataset with 10 scenes offering 5 degrees of freedom [271].

Regarding LF QoE, it is evident that subjective quality assessment fundamentally relies on LF content, thereby requiring the availability of LF datasets. Datasets targeted for evaluating the quality of LFs usually incorporate high-quality contents, along with their impaired counterparts [120]. According to Shafiee and Martini [266], LF datasets can be categorized into three groups: (i) content-only datasets, (ii) task-based datasets, and (iii) QoE datasets. As the name implies, content-only datasets contain the LF contents only, and nothing more. Real-world captured contents may be acquired by a lenslet camera [301, 255, 269, 291, 95, 14], a single-lens camera [115, 80, 234, 332, 95, 127], or an array of cameras [95, 331]. In the case of rendered content-only datasets [254, 234, 127], the camera is virtual, which is, of course, applicable to the other dataset types as well, along with the classification of real cameras. Task-based datasets include additional information on the task for which the dataset was created. Similarly to content-only datasets, task-based datasets can also be captured by a lenslet camera [316, 235, 139, 263, 26], a single-lens camera [176, 142], an array of cameras [279], or by a virtual camera [176, 309, 26, 15]. Finally, QoE datasets contain subjective ratings that were acquired through extensive testing with numerous test participants. The currently available QoE datasets were

Table 2.4: LF dataset types [266, 120]

LF dataset type	**Definition**	**Data capture methods**	**Examples**
Content-only	Contains the LF contents only	- Lenslet camera - Single-lens camera - Array of cameras - Virtual camera	[301, 255, 269, 291, 95, 14] [115, 80, 234, 332, 95, 127] [95, 331] [254, 234, 127]
Task-based	Includes additional information on the task for which the dataset was created	- Lenslet camera - Single-lens camera - Array of cameras - Virtual camera	[316, 235, 139, 263, 26] [176, 142] [279] [176, 309, 26, 15]
QoE	Contains subjective ratings acquired through extensive testing with numerous test participants	-Lenslet camera - Single-lens camera - Virtual camera	[21, 307, 271, 264, 340] [287, 243] [267, 340]

captured by either a lenslet camera [21, 307, 271, 264, 340], a single-lens camera [287, 243], or a virtual camera [267, 340]. The dominant portion of the datasets covered in this section contain LF images. LF video datasets –such as the work of Guillo *et al.* [115]– are exceptionally rare at the time of writing. The different types of LF datasets [266] –along with relevant examples for each data capture method– are summarized in Table 2.4 [120].

2.9 Light field compression

Since LF rendering requires the storage of almost all the visual information related to the captured scene by means of storing multiple views of the given scene, compression techniques are needed to accommodate the huge amount of information. In order to achieve a smooth continuous motion parallax effect, the number of captured views needs to be sufficiently high. In efforts to solve the data storage and bandwidth problem, various compression techniques –taking into account the similarities between the LF images representing the scene– were suggested for LFs. This includes disparity compensation for compressing synthetic 4D LFs [215, 111, 150], which was already highly investigated prior to the emergence of the first modern projection-based LFDs. Approximation through factorization [59] and geometry estimation using Wyner-Ziv coding [339] were also relevant approaches of that scientific era. From the beginning of the 2010s, various compression methods for LF images captured by hand-held devices were proposed [200, 201, 245, 199, 63, 233, 209]. Subsequent efforts relied on Homography-based Low-Rank Approximation (HLRA) [152], disparity-guided sparse coding [58], deep-learning-based assessment of the intrinsic similarities between LF images [337], and Fourier disparity layer representation

–where the Fourier domain can effectively construct a set of layers for LF representation given very few views [87]. In the recent years, the contemporary solutions included low-bitrate LF compression based on structural consistency [143], disparity-based global representation prediction [60], compression by means of a Generative Adversarial Network (GAN) [208], Spatial-Angular-Decorrelated Network (SADN) [295], bit allocation based on a Coding Tree Unit (CTU) (which takes into account the HVS to remove perceptual redundancy) [153], compressed representation via MultiPlane Image (MPI)s comprised of semi-transparent stacked images [174], and neural-network-based compression by using the visual aspects of Sub-Aperture Image (SAI)s, incorporating descriptive and modulatory kernels [270]. Further lossy compression methods for LFs include transform coding [216, 22, 89, 56, 23], predictive coding [184, 62, 209, 154], pseudo-sequence coding methods [77, 306, 198], and utilizing a two-dimensional prediction coding structure [268] [120]. Table 2.5 summarizes the different LF compression techniques.

Table 2.5: LF compression techniques [120]

LF compression technique	**Citations**	**Type of compression**
Disparity compensation for compressing synthetic 4D LFs	[215, 111, 150]	lossy
Approximation through factorization	[59]	lossy
Geometry estimation using Wyner-Ziv coding	[339]	lossy
Compression methods for LF images captured by hand-held devices	[200, 201, 199, 63, 233, 209] [245]	lossy lossless
Homography-based low-rank approximation	[152]	lossy
Disparity-guided sparse coding	[58]	lossy
Deep-learning-based assessment of the intrinsic similarities between LF images	[337]	lossy
Fourier disparity layer representation	[87]	lossy
Low-bitrate LF compression based on structural consistency	[143]	lossy
Disparity-based global representation prediction	[60]	lossy
Compression by means of a generative adversarial network	[208]	lossy
Spatial-angular-decorrelated network	[295]	lossy
Bit allocation based on a coding tree unit	[153]	lossy
Compressed representation via multiplane images comprised of semi-transparent stacked images	[174]	lossy
Neural-network-based compression by using the visual aspects of sub-aperture images, incorporating descriptive and modulatory kernels	[270]	lossless
Transform coding	[216, 22, 89, 56, 23]	lossy
Predictive coding	[184, 62, 209, 154]	lossy
Pseudo-sequence coding methods	[77, 306, 198]	lossy
2D prediction coding framework	[268]	lossy

CHAPTER 3

Light field camera animation

As LF technology is rapidly advancing, its presence in the industry is continuously growing and researchers are addressing new applications of LF capture and visualization. While the large-scale penetration of the consumer market is still a moderately long-term goal, the availability of real LFDs already allows experts to investigate the relevant use cases, which may progressively evolve into common daily activities of future societies. Such use cases include medical applications (such as radiology [205, 70]), telepresence [336, 68], cinematography [168], digital signage (e.g., via LF LED wall panels [39]) and so many more.

A great number of the potential use cases requires camera animation, which has not been thoroughly investigated for LF technology yet. The aim of this scientific contribution is to investigate camera animation in the context of LF visualization, with a focus on both theoretical analysis and practical application. Among the use cases extensively relying on camera animation is cinematography.

Cinematography –also known as film-making– is the field defining a set of techniques and rules for the effective communication of actions. It encompasses ideas, words, motions, tones and so many more, and communicates them visually [53]. Over the years, cinematography has evolved from 2D to 3D, where viewers can watch 3D movies in cinemas by means of 3D glasses. With the aforementioned technological evolution, LFDs offer a huge leap in the field of cinematography by providing viewers with 3D glasses-free experience, adding more sense of immersion, as such displays act as a 3D window to the real world. Thus, they are the most suitable for cinematographic purposes as conventional 3D solutions evoke a sense of isolation for viewers. Among the LFDs designed to support cinematographic purposes, is the HoloVizio C80 cinema system[1].

One of the major aspects of cinematography is the camera control which takes into account camera motion and path planning to produce realistic motion paths, both of which have

[1] https://holografika.com/c80-glasses-free-3d-cinema/

long been investigated for conventional 2D cameras. Applying the same techniques to LF cameras can be challenging, especially for wide-baseline devices. While adhering to the cinematographic rules, displaying contents on LFDs requires taking into account their technology-specific and device-specific challenges and limitations. Accordingly, the produced results and motions are evidently different from those displayed on conventional 2D screens. In order to overcome the problems imposed by wide-baseline LF cameras, the usage of virtual cameras to simulate a set of different (physical) camera motions is suggested.

Our work fundamentally builds on the extensive literature on camera animation that is already available for conventional 2D displays. The efforts presented in this chapter take into account the limitations and challenges that apply to LF capture and visualization [71], which resulted an interactive simulation on a real LFD. Moreover, the implemented camera animations were extended to include realistic physical motions. Similar to the different camera animation techniques, the realistic camera motions were simulated and tested on a real LFD. The plausibility and effectiveness of these motions were evaluated via different objective metrics [117], in addition to conducting subjective tests.

This chapter provides a comprehensive overview of LF camera animation. It begins by exploring general cinematography and simulation camera animations in Sections 3.1.1 and 3.1.2, respectively. Next, it discusses the principles of camera animation design on 3D displays in Section 3.2. Following this, Section 3.3 delves into camera animations for LF visualization, addressing key challenges, setup considerations, and techniques that effectively utilize the unique advantages of LF technology. Central to this discussion is the framework simulation for LF camera animation, outlined in Section 3.4, which establishes the essential technical foundation for effective LF camera animation. Section 3.5 examines the visualization of LF camera animations, addressing the associated objective metrics, evaluation and results, and subjective assessments. Finally, the chapter concludes in Section 3.6, highlighting potential directions for further research on the investigated topic.

3.1 General camera animation

3.1.1 Cinematography camera animations

Among the main components of cinematography are camera movements and shots, which play important roles in storytelling. In this section, we discuss the most relevant types of cinematographic camera movements. Aside from the pan, tilt, zoom and rack focus, a change in the camera position is required.

Pan is short for panoramic. It describes the horizontal rotation of the camera left or right without changing its position. The strobing effect, however, arises when the camera is being moved too fast, which accounts as a limitation for the pan movement itself. To minimize this effect, a general guideline suggests that with a 180° shutter opening and a frame rate of 24 or 25 fps, an object should take no less than five seconds to move across the frame. Moving any faster increases the risk of strobing, with higher frame rates requiring slower panning speeds to maintain smooth motion [53].

Similarly to pan, tilt does not change the position of the camera. Yet unlike pan, tilt describes the vertical (up and down) rotation of the camera. Furthermore, it needs to be stated that tilt is not used as frequently as pan since the majority of events in everyday life (and thus in cinematic content) occur along the horizontal plane [53].

The camera movement known as zoom encompasses an optical change in focal length. In the world of cinematography, it is crucial that zoom is only used when such visual method is necessary (i.e., carries meaning for artistic and/or storytelling purposes). Moreover, hiding zoom is somewhat advisable to suppress in order to avoid drawing the attention of audience to the zoom effect, which may make the audience aware that they are merely spectators of the movie instead of experiencing a sense of immersion. This can be achieved by combining zoom with other camera movements, such as pan, dolly or tilt, or with certain movements of the actors and objects in the scene [53].

Dolly is often called “move in/move out” and “push in/push out”. The move in/out camera movement combines both the wide and the tighter shots of the scene. This movement is used to focus the attention of the audience efficiently rather than cutting the scene from a wider to a closer shot. There are also many other cinematic uses for this type of camera movement. For example, dolly is commonly used as a form of pulling back from a scene upon the entrance of an actor. During this type of dolly, the camera moves towards or away from the subject of interest. Unlike zooming, the camera is a wheeled cart (or mounted on a track/motorized vehicle), so the camera itself moves. This gives a sense of world movement around the subject. In other words, the background appears to be moving behind the subject, which further enhances the sense of motion [53, 126, 280].

Truck movement is rather similar to dolly. However, it moves the camera horizontally (left and right) instead of in and out. This type of camera motion is typical for the cinematic use of following a moving entity (e.g., a character in action) [126, 280].

In case of pedestal, similarly to the concept of dolly and truck, the camera moves, but this movement is vertical (up and down). It is frequently used to capture tall/high entities (e.g., the cinematic introduction of a tall character or a tall building) [126, 227].

Regarding the punch in movement, similar results to the zoom in effect are attained to the extent that they are sometimes mixed up, however, they are concurrently different. Whereas both of them achieve the same goal of moving the audience closer to the element of interest, their methods and hence, their resulting effects, are different. A punch in involves capturing multiple shots at progressively closer distances rather than simply adjusting the focal length, creating a sense of urgency and engagement. In other words, punch in “cuts straight to the chase” as it is more direct [53, 223]. Unlike the zoom movement where a single shot is sufficient, punch in requires a minimum of two shots. An example for punch in is depicted in one of the scenes in the “Casino Royale” movie (illustrated in Figure 3.1), where the fuel tanker is the item of interest to which the punch in effect is applied. This denotes the fuel tanker being the important element about

Figure 3.1: Punch in camera motion in "Casino Royale" movie.

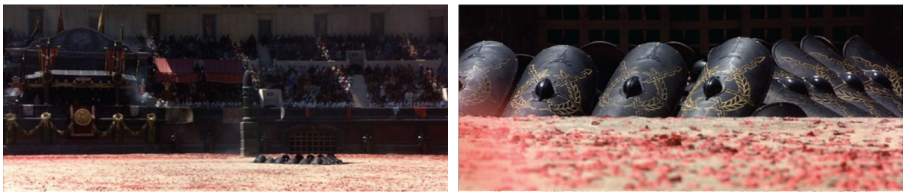


Figure 3.2: Punch in camera motion in "Gladiator" movie. Main characters rising to the arena.

which James Bond realizes was his main objective [223]. Another example for punch in is depicted in the classic "Gladiator" movie, where the main characters are driven into the arena from the underground area. The punch in effect is accomplished by capturing the scene primarily through a wide camera shot, switching right after to a long lens shot (illustrated in Figure 3.2) [53].

Last but not least, one of the widely used camera movements in filmmaking is rack focus, where the focus of camera lens is altered to shift the focus from one object another, spanning from minor to substantial alterations [53, 117].

3.1.2 Simulation camera animations

These types of camera animation are used extensively in video games, where the player interacts and perceives the surrounding environment by means of virtual cameras. For perceiving the virtual world from a certain perspective, the main components of a camera system have to be set (i.e., the position and the orientation of the camera) [265]. In this section, we discuss the most relevant types of simulation cameras.

Fly/Walk/Point-Of-View (POV)/First-person cameras are most commonly used in video games. The idea is to view the scene from the perspective of the character, the avatar of the player, or the player-controlled vehicle (e.g., first-person cockpit view or view from the front of the vehicle). This technique appears in a multitude of video game genres, among which first-person shooters and driving/flying simulations are very well known. Hence, the technique of first-person camera can provide a significant sense of immersion. Although first-person cameras may add reality to the game, its field of vision is rather limited. Furthermore, in addition to video games, first-person cameras are sometimes used in cinematic content to present the perspective of a given character. Such storytelling techniques are also referred to as the POV shot [55].

The idea of second-person camera animation is to view the entity of interest from the perspective of another entity [246]. For example, the main character is viewed from the perspective of a different character. This camera was incorporated in games such as Battletoads[2], where the fight is viewed from the POV of the opponent [117].

Unlike first-person and second-person cameras, third-person cameras are separated from the focus of the entity of interest. In this case, the context of the game is viewed from the perspective of an external position (i.e., a virtual camera) and not from the perspective of an actual entity [130, 117]. Some of the disadvantages introduced by first-person cameras are overcome by third-person cameras due to their wide FOVs, allowing the visualization of the 3D environment from multiple perspectives, in addition to the main character. In other words, third-person cameras offer a better presentation of the relation between the environment and the main entity. Accordingly, games emphasizing the importance of the game world as a major part of the gameplay rely on third-person cameras, where intense interactions exist. An example for such games are fighting games, requiring players to visualize themselves in the gameplay along with their opponents and the overall environment. On the other hand, third-person cameras are more difficult to implement compared to first-person cameras, since the latter requires less animations on account of the main character's invisibility. Whereas for third-person cameras, more character animations need to be considered including walking, jumping, running, crawling, etc [54].

An essential consideration regarding third-person cameras involves their placement within 3D scenes, which in turn has a significant impact on the story narrative. Unlike first-person cameras portraying the character's vision, third-person cameras are located externally with respect to the main entity, hence, their position should be adjusted in a way to capture the main entity's viewing direction. When placing third-person cameras in 3D scenes, it is crucial to consider the objective of attaining a clear unobstructed view of the surroundings. Several locations are proposed, with the most common ones being the following: (i) centered camera located behind and above the main entity accomplishing a central meaning, (ii) camera positioned at the same height of the main character or object, with a slight deviation to the left or right [130, 276, 54].

While employing third-person cameras, numerous challenges may arise, including the following:

- Non-centered camera: These cameras exhibit the perspective of the camera's center of vision rather than that of the primary subject, resulting in various issues, such as the main character becoming trapped in certain scenes and the camera bouncing off objects within the environment, among other challenges. A robust hybrid solution was proposed, alternating the camera's center of vision. This approach was implemented within the gameplay of "Batman Arkham Asylum", where the camera adjusts its position based on the current objective. In instances such as world exploration and regular movement, the camera veers to the right, enhancing visibility of objects. Conversely, during combat and sprint sequences, the camera re-centers itself and zooms out to aid the player's offensive maneuvers and provide

[2]Battletoads (©Masaya, 1991)

(a) Camera shifted to the right (b) Camera centered in battles

Figure 3.3: Different camera positions in “Batman Arkham Asylum” gameplay [54]

a broader view for escape tactics [54]. Figure 3.3 depicts the various camera placements utilized in the gameplay of “Batman Arkham Asylum”.

- Obstructed camera: One of the primary challenges encountered with third-person cameras is the potential for the main entity being hidden from view by obstructive objects. To address this issue, numerous solutions have been proposed, such as enabling the transparency for obstructive elements, implementing camera whiskers allowing the camera to navigate around nearby objects, ensuring an uninterrupted view of the main character, and implementing the silhouette solution, as seen in “Super Mario Sunshine” gameplay. This involves projecting a dark silhouette of the main character –in that case Mario– onto obstructive elements such as walls. This allows the player to maintain visibility of the main character.

- Camera bouncing off objects: Due to collisions with 3D elements, the third-person camera rebounds within the scene. This is frequently a result of the virtual sphere surrounding it, referred to as the “Sphere Collider”. This issue is prevalent in various gameplays like “Assassin Creed Syndicate” and “Dark Souls 3”. To address the previously mentioned issue, the “cut-out view” technique was implemented in the game “For Honor”. With this approach, the camera is allowed to pass through objects, ensuring the player maintains an uninterrupted view of the game world. Furthermore, an indicator for this view is displayed on the screen.

Orbiter cameras [257] always have their “lookat” point at the center of the bounding volume of the object of interest. The camera can rotate around this fixed point on a sphere with a fixed radius. In some implementations, it is possible to change the length of the radius or to scale the scene to achieve close-up or zoom-like effects. Such cameras are often used in industrial and medical applications [117]. The utilization of an orbiter camera in medical use cases is illustrated in Figure 3.4.

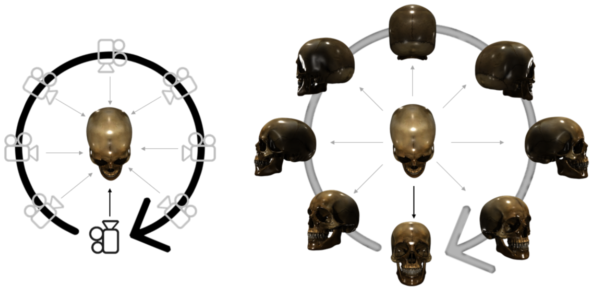


Figure 3.4: Orbiter camera for medical use cases [120]

3.2 Camera animation design on 3D displays

Camera animation design on 3D displays varies on a case-by-case basis, but for most solutions, they stick to a single-interaction type. Head-mounted AR and VR devices almost exclusively use the first-person camera model. Volumetric displays usually opt for orbiter camera interactions. The only exception to this rule is the case of S3D cinema, which retains its richness of expression and uses all camera movements that do not change the focal length. Changing the focal length would require a change in baseline (or lenses) and a possible calibration of the system. Recalculating the stereo base is usually calculated with the Bercovitz formula [45] as

$$B = P\frac{LN}{L-N}(\frac{1}{F} - \frac{L+N}{2LN}), \tag{3.1}$$

where B is the stereo baseline; P is the parallax aimed for; L is the far clipping plane; N is the near clipping plane; F is the focal length of the lens. The only exception to this rule are animated movies, where calibration is not required and the frame-by-frame changes in baseline or lens parameters are not an issue. Also, the cameras with asymmetric perspective that converge on a virtual screen can be used to provide a higher-quality stereoscopic image pair. The same stereo camera rigs are equipped with apparatus to change the baseline and consequently the focal length; however, most directors would prefer to cut due to the fact that this operation changes the "flatness" of on-screen objects [117].

3.3 Camera animations for light field visualization

3.3.1 Challenges and obstacles

LFs are characterized by the baseline: they are either narrow-baseline or wide-baseline LFs. For each category, there are corresponding LF cameras and displays. For LF cameras, the captured LFs should map to those of the LFDs. Unlike narrow-baseline devices, wide-baseline LF cameras (i.e., camera arrays) impose many challenges. Using camera arrays is challenging due to their weight, physical size, possibility of self-capture and high price [71]. Furthermore, it is hard to generate uniform lighting in a scene captured similarly by all cameras in the region. In addition to the aforementioned issues, using dynamic camera arrays introduce additional challenges, among which is the camera tracking and calibration [119]. More on the challenges introduced by various camera systems is discussed in Section 2.4.

3.3.2 Camera setup on light field display

Let us now define the capture surface of an LF camera. First, we determine a set of points by taking the individual spatial positions for each sensor per pixel. Then, we can tessellate a piece-wise flat spanning surface between the neighboring points to obtain the capture surface.

The normal of the capture plane is the average of the camera direction vectors. A camera direction vector is defined as $D_{camera} = \frac{P_{lookAt} - P_{camera}}{\|P_{lookAt} - P_{camera}\|}$, where P_{lookAt} represents the look-at point and P_{camera} denotes the camera position. This scientific discussion excludes camera systems with any two rays that have an angle larger than $\pm 90°$, as such systems should always be treated as 2 or more separate systems from this perspective. The plane contains the point of the capture surface for which the dot product of the point with the normal vector of the capture plane is minimal.

We define the capture rectangle by evaluating the intersection points of all light rays measured by the LF camera with the capture plane, and by calculating the axis-aligned bounding rectangle around them. We call this bounding rectangle the capture rectangle. Please note that in the 1D linear case, the capture rectangle and the baseline are one and the same.

In general, the more a camera system covers the whole baseline during the measurement of the LF, the better match it is going to be for the LF of the display –assuming the same camera count and camera parameters, such as resolution and FOV. As baselines are typically in the range of 3 to 24 meters for practical display sizes, LFDs are optimally matched by wide-baseline cameras.

To determine how well a camera system performs on an LFD, we need to establish an error metric. First, we have to define the observer rectangle (observer line for HOP) for LFDs. This rectangle lies on the observer plane, which is parallel to the display plane. The observer rectangle is the minimum axis-aligned bounding rectangle of all intersection points of emitted rays and the observer plane.

Then, we convert all camera rays into a Cartesian coordinate system, where we have defined the mathematical representation of the display rays and the observer rectangle, using a 4×4 affine transformation matrix, also known as the ROI matrix [90]. The coordinate system shall place the display plane on an xy plane at $z = 0$. We shall further restrict the parameters of the ROI matrix to contain uniform scaling, and we want to set the matrix in such a manner that after the transformation, the observer plane and the capture plane are equivalent. We recalculate the capture rectangle in the new coordinate system. It is easy to see that the only valid display rays –for which we can reliably render from the captured camera rays– lie in the intersection of the observer rectangle and the capture rectangle.

The closest camera ray to a display ray can be found by finding the minimum of the following sum for each camera ray: sum of the distance of the camera ray intersection with the display surface to the display ray's emission point and the distance of the display ray's intersection point with the observer plane and the camera ray's eye position.

An error metric for a set of camera rays, an ROI, and an LFD with a planar surface can be determined as

$$E_{d\ ray_n} = \frac{1}{4}\left(\frac{abs(O_{d_{nx}} - I_{c_{nx}})}{S_x} + \frac{abs(O_{d_{ny}} - I_{c_{ny}})}{S_y} + \frac{abs(I_{d_{nx}} - O_{c_{nx}})}{S_{int_x}} + \frac{abs(I_{d_{ny}} - O_{c_{ny}})}{S_{int_y}}\right) \tag{3.2}$$

for all $n \in N_i$ and

$$E_{capture} = \frac{\sum_{n \in N_o} 1 + \sum_{n \in N_i} E_{d\ ray_n}}{N}, \tag{3.3}$$

where N is the total number of display rays; N_i is the set of display rays inside and N_o is the set of display rays outside the intersection of the observer rectangle and the capture rectangle; S_{int} is the (2D) size of the intersection rectangle; S is the (2D) size of the display surface; O_{d_n} is the origin of the nth display ray; I_{d_n} is the intersection point of the display ray and the observer rectangle; O_{c_n} is the closest camera ray origin to the nth display ray, I_{c_n} is the closest intersection point to O_{d_n} on the display plane of all camera rays with origin O_{c_n}; x and y denote the x and y components of the points and sizes. Figure 3.5 illustrates the camera space and display space.

To extend this metric to LFDs with non-planar surfaces, Euclidian points and distances measured on the display plane and divided by the display size need to be replaced with u, v surface-normalized parametric points and distances. Distances inside the projected area of the pixel on the observer plane and the emission surface on the display surface, respectively, can be treated as zero to improve the metric. In case of additional color mixing from multiple camera rays, the metric can be extended to include all selected camera rays for a display ray and $E_{d\ ray_n}$ needs to be weighted by the weights used for mixing color from the chosen camera rays.

From this metric, it is easy to see that it would be extremely difficult to build LF capture systems for most LFDs where $E_{capture\ ROI} = 0$ holds true. However, it is easy to define new virtual cameras (sets of capture rays that match the display rays exactly) for any given ROI transform of a virtual scene that matches the criteria for the ROI transforms listed above. Therefore, using virtual cameras is a superior option to test camera-related problems, as they are both easy to place and move using only the ROI matrix and they are free from capture error by definition [117].

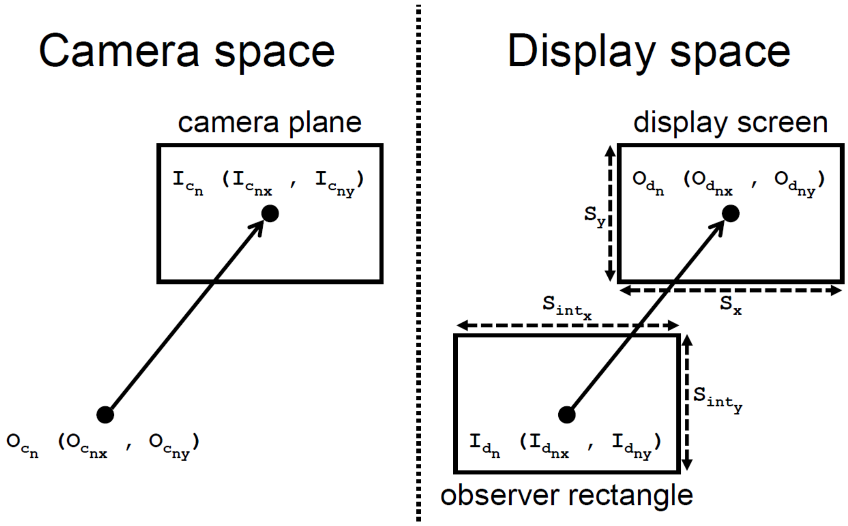


Figure 3.5: Camera space and display space [117]

3.3.3 Light field camera animation

As stated in Section 2.4, LF cameras are used to capture information about light distribution. In other words, for each ray arriving at the sensor, its amount of light is captured [238]. In our case, the LF of a virtual scene is captured by an error-free virtual LF camera to overcome the challenges introduced in Section 3.3.1. Camera movement is facilitated through the ROI matrix. In practice, display rays are evaluated once and are transformed with the inverse of the ROI matrix to be in world space. As all other virtual objects and lights are also in the same coordinate system, we can easily render the individual rays.

Previous works on LF virtual camera animation for LFDs involved orbiter cameras or cameras using scene-centered rotations with dolly and truck without camera-scene interactions [36, 24]. By implementing the various camera animation types, we can evaluate their usefulness for LF visualization. We generated animations for some typical scenarios used in cinematography, where we included an object of interest for the film, which is especially important for the first-person and third-person cases.

The following criteria were used to evaluate usefulness:

- General visibility of the scene along the observer line during animations.
- Frequency of immersion-breaking occluders.
- Frequency of collisions and course corrections with the scene.
- Frequency of depth-related artefacts.
- Expected depth of field changes are not occurring.

The implementation is flexible enough to work across a whole range of LFDs, specifically lenticular and projection-based ones. It was built using Holografika's clustered rendering modules. It also used the Bullet Physics library [65] to provide a level of realism for the scene. The application is implemented as a testing framework, where any combination of existing scenarios –namely camera motion– scene and scene-dependent interactions can be rendered in real time to aid the evaluation. Our findings can be directly applied to the motion and operation of physical LF cameras with comparable baselines, observing the scaling factor of the ROI transform, when capturing for scenes with comparable aspect ratios [117].

3.4 Framework simulation for light field camera animation

3.4.1 Physical properties for the simulation

Due to the issues elaborated in Section 3.3.1, it is advised to use virtual LF cameras in order to simulate realistic physical environments. Nonetheless, the physical properties for realistic LF cameras should be taken into consideration when using virtual cameras in order to accurately simulate a realistic camera path.

LF cameras are characterized by their sensor properties (frame rate, focal range, aperture size, intrinsic camera parameters), optical properties (baseline, arrangement of the cameras, resolution), physical dimensions (shape of the cameras and the mount) and weight (weight of the camera and the mount). Whereas the optical and sensor properties for LF cameras cannot be changed, physical properties can be easily altered.

Regarding the baseline for the LFD (and thus for the LF camera simulation), it cannot be altered. However, changing the ROI achieves the desired effect, since it is the same as changing the baseline and the arrangement of the cameras. ROI is used to transform between the display's physical coordinate system and the world coordinate system. In other words, changing the object of interest is applicable by means of changing the ROI [119].

3.4.2 Simulation setup

Our work aims at fusing path planning for wide-baseline LF cameras and the simulation of physical cameras for film making. The goal is to introduce an application used to build physical environments in which we can evaluate camera motions that typically come up during film production. Physical parameters used in path planning can be entered in the application, such as the speed (acceleration/deceleration), mass, etc. of the camera rig. Accordingly, path planning for LFDs is tackled by defining the scene, path and physical properties. However, the remaining properties (optical and sensor properties) are determined by the LFD, thus ensuring the consistency between the properties of both the LF cameras and displays, and hence, avoiding any unnecessary conversions.

In our work, camera animations by means of virtual cameras were implemented and tested on a real LFD, namely the HoloVizio C80[3]. This LFD has an aspect ratio of 16:9 with a screen size of 3 m×1.8 m, hence being the perfect candidate for testing camera animations due to its big size –simulating a cinema screen. The viewing angle of this screen is 40° with a brightness of 1000 cd/m^2. The tested physical scenes were built up by means of basic shapes, including a generic ground, boxes, cylinders, planes, cars, and suspension elements. In order to simulate the physical properties of the modelled shapes, the "Bullet Physics Library" [65] was used. In our application, we allow users to change the weight of the camera within given limits, as well as the size, speed and suspension properties of the camera mount platform. For the objects in the scene, change is possible for their sizes, positions, orientations and weights [119, 117].

3.5 Visualization of light field camera animation

3.5.1 Cinematography and simulation camera animations on light field displays

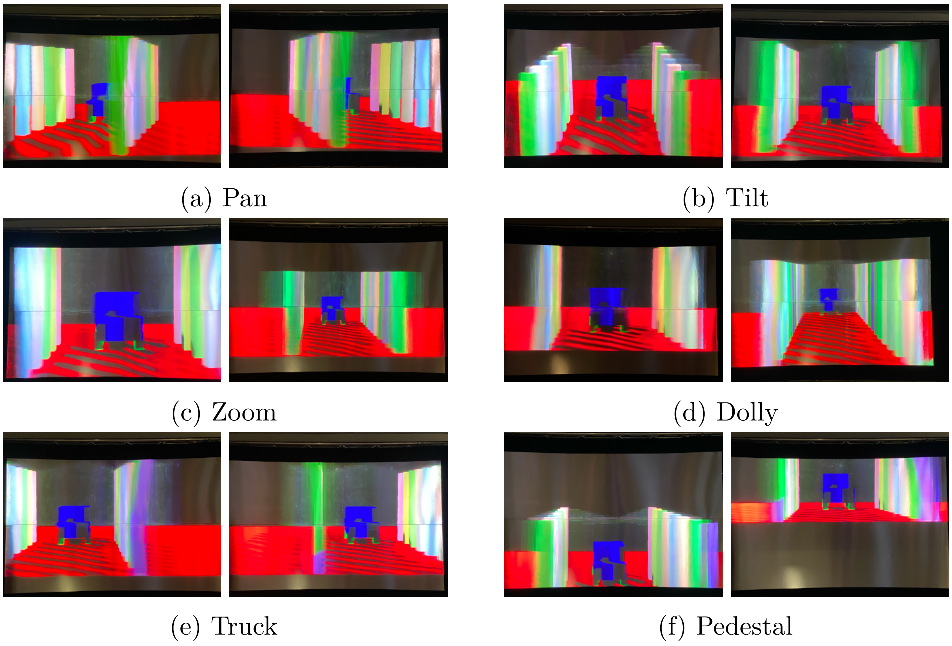
First, we simulated and tested the different camera animation techniques mentioned in Section 3.1.1. As a means of testing the different camera animations, a scene composed of an aisle of columns was implemented.

The investigated camera animations were pan, tilt, zoom, dolly, truck and pedestal. Figure 3.6 depicts the visualization of camera animations on the LFD. In order to get a better overview of the scene from multiple perspectives, orthographic views were added. Figure 3.7 shows the orthographic camera views for the scene. In addition to testing cinematographic camera animations, simulation camera animations were tested as well. Figure 3.8 shows the simulation camera animations (first-person and third-person cameras).

As a result of visualizing the different camera animations on the LFD, a series of inferences could be made. As discussed earlier, a set of criteria was used to evaluate camera animations. Those include general visibility, frequency of immersion, collision frequency, depth-related artifacts frequency and the expected depth of field changes not occurring.

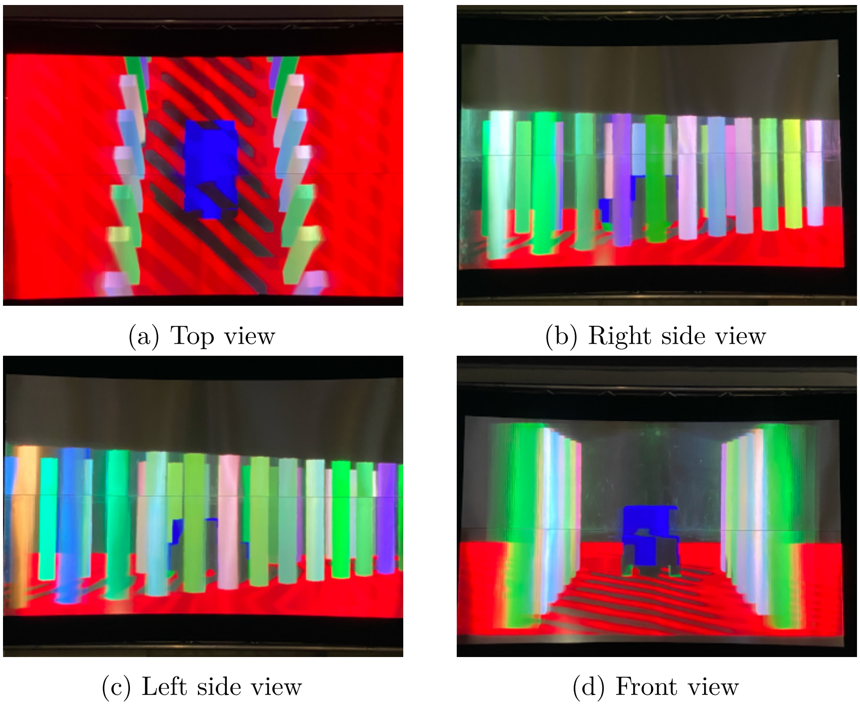
Perceptual assessment was carried out via expert evaluation. In this context, this means that various LF experts of the institution rated the investigated aspects of the different camera animations, choosing from a set of descriptive, subjective options for each aspect (e.g., collision frequency was either none, low, medium or high). The evaluations were based on the plausibility of the visualized content on the LFDs, as well as prior expert

[3]https://holografika.com/c80-glasses-free-3d-cinema/



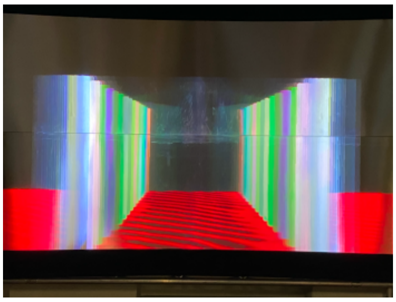
(a) Pan
(b) Tilt
(c) Zoom
(d) Dolly
(e) Truck
(f) Pedestal

Figure 3.6: Cinematography camera animations on LFD [117]

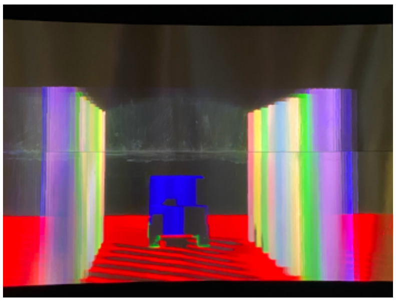


(a) Top view
(b) Right side view
(c) Left side view
(d) Front view

Figure 3.7: Orthographic views [117]



(a) First-person camera



(b) Third-person camera

Figure 3.8: Simulation camera animations [117]

Table 3.1: Results of camera animations visualized on the LFD [117]

Camera animation	General visibility	Occluder frequency	Collision frequency	Depth-related artifacts frequency	Expected depth of field changes not occuring
Pan	Good	Low	None	Low	N/A
Tilt	Mediocre	None	Medium	High	N/A
Zoom in	Mediocre	None	High	High	Yes
Zoom out	Mediocre	Low	Low	Low	Yes
Dolly in	Mediocre	None	High	High	N/A
Dolly out	Mediocre	None	Low	Low	N/A
Truck	Good	Low	None	Low	N/A
Pedestal	Mediocre	High	Medium	Medium	N/A
First-person	Bad	None	High	High	N/A
Third-person	Mediocre	None	High	High	N/A

knowledge of the optical limitations and challenges of LFDs. The results of the expert evaluation are presented in Table 3.1.

Starting off with the cinematographic camera animations, the pan and truck movements turned out to have the best general visibility, followed by tilt, zoom, dolly and pedestal.

Occluder frequency is the rate by which the camera is occluded throughout its animation. Pan, zoom out, dolly out and truck camera animations had the lowest occluder frequency, followed by tilt and pedestal. However, the highest occluder frequency was noticed in case of zoom in and dolly in motions.

Collision frequency is the rate by which the camera collides with objects from the scene when being animated and would need to stop, land or change trajectory. Collision is not expected for this scene, only for the pedestal case, as the one and only collider in the scene is the ground. Camera collision can be implemented for LF visualization in several different ways. It can be evaluated in world space against the bounding volume of the LF, the Axis-Aligned Bounding Box (AABB) of the bounding volume, the ROI box, the center of the ROI box, the observer line or the axis-aligned bounding rectangle of the intersection points of display rays and the maximum addressable depth plane towards the observers. The current implementation used the observer line for collision, as this behavior matches that of a physical LF camera system the best.

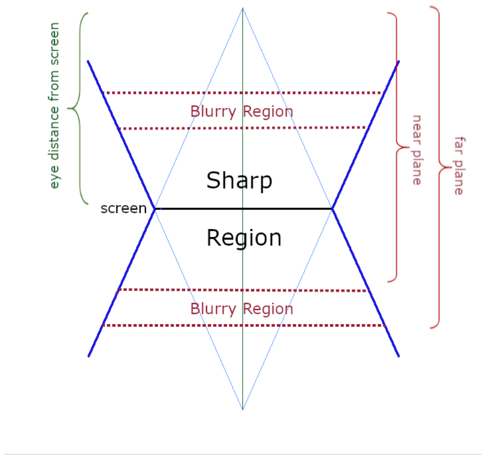
Depth-related artefacts arose when objects that were previously in the right range of depth for sharp visualization got close or over the range for the sharp region of the depth of field. Due to the arrangement of objects in this scene, this metric follows the occluder frequency quite closely. For some cases, such as tilt, the amount of ground that is visible changed significantly, resulting in more artefacts, while keeping the number of occluders similar throughout the motion sequence. For dolly out, this occurrence of depth-related artefacts became smaller in the back, and more frequent in the front. For dolly in, the opposite applied.

As for 2D camera animations, zooming in and out results in change of the camera's focal lens. Although the same effects are expected to occur when utilizing zoom on LF displays, change in the focal lens is, of course, not possible when using LF. Accordingly, the expected changes in the depth of field when zooming did not occur. In order to produce something similar to the zoom effect, the extents of the ROI were scaled.

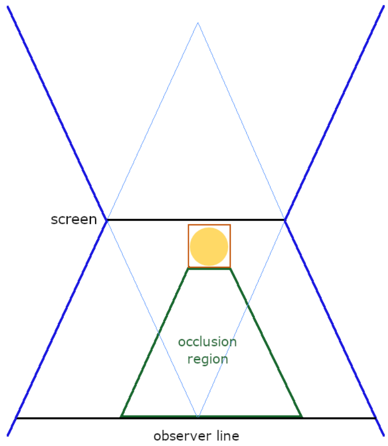
Moving on to the simulation camera animations, the first-person and third-person cameras were implemented and tested. The general visibility for the first-person camera was poor; however, it was better for the third-person camera. Both first-person and third-person cameras resulted in high rates for occluders, collision and depth-related artefacts. As illustrated in the figures, some camera animations led to plausible results on the LFD, while others were lacking. Among the cinematographic camera animations, pan, tilt, truck and pedestal camera movements resulted in satisfactory outputs. However, blurriness artefacts were present for dolly and zoom towards the scene. The same applied to first-person camera when testing simulation camera animations. On the other hand, third-person cameras resulted in plausible results as well. Table 3.1 summarizes the results for the camera animations visualized on the LFD [117].

3.5.2 Realistic physical simulation for light field cameras

As an extension to our work, various realistic physical camera animations were simulated and tested on the HoloVizio C80 HOP LFD. The primary motivation was to simulate some of the realistic motions that are common in cinematography by means of virtual LF cameras.



(a) Blurry regions



(b) Occlusion region

Figure 3.9: Metrics for light field visualization [119]

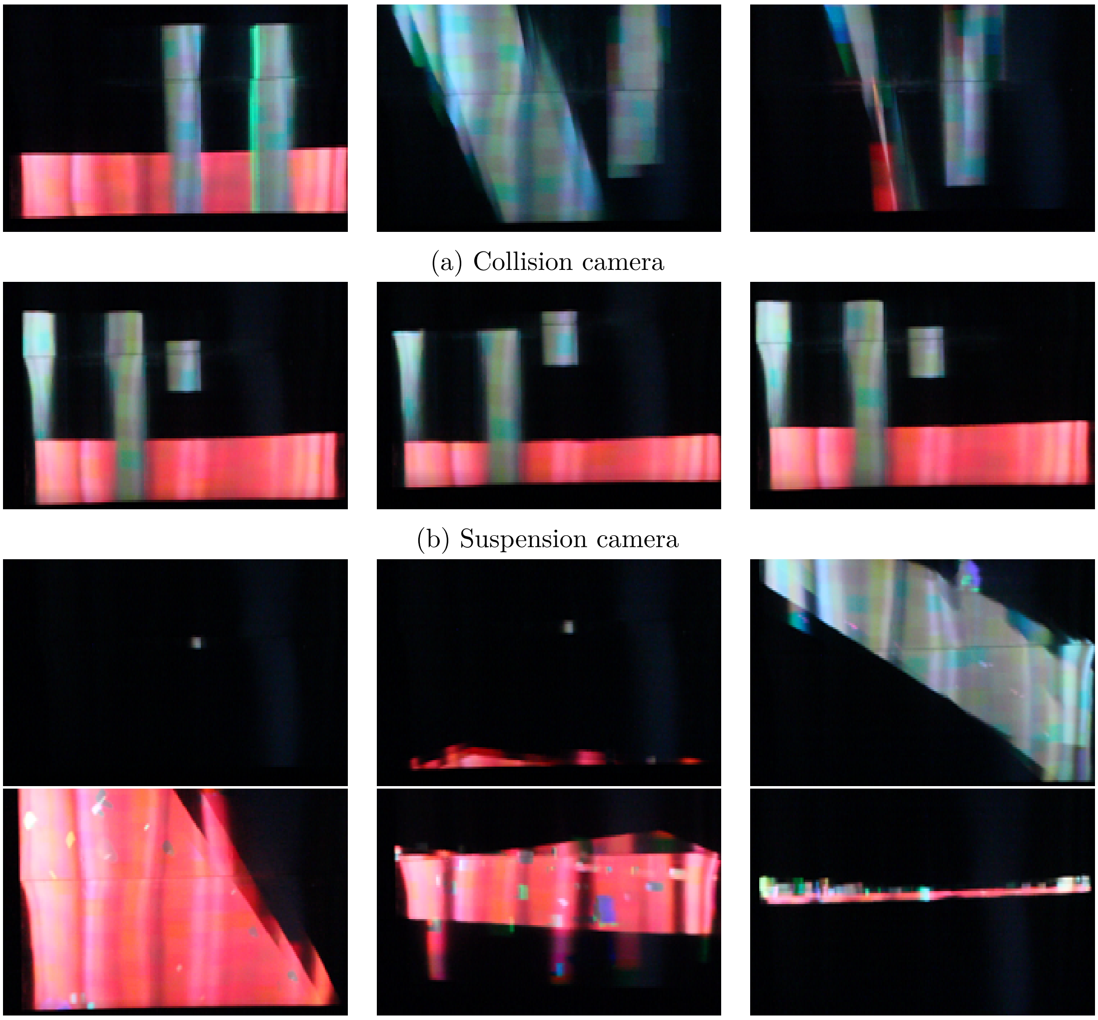
Metrics

Testing the different realistic camera motions in the physical environments require some metrics to define the plausibility of the achieved results. In our work, we specify some of the metrics to be measured, including the collisions, objects entering the blurry region of LFD, occlusions, and hence, image stability is based on these aforementioned metrics. It is important to take into account that these measurements are performed for HOP LFDs. Therefore, the observer line is considered.

Measuring camera collisions is achieved by means of AABBs. The number of objects colliding with the camera is done by counting the number of intersections between the AABBs of the objects in the scene and the AABB of camera.

For LFDs, double viewing frustums (in front of the screen and behind the screen) are constructed instead of one as in case of conventional displays. Figure 3.9a shows a top view for the setup of LFDs. The black line depicts the screen, whereas the blue lines show the viewing angles. The dotted lines encompass the areas where objects are rendered blurry. The number of objects in the blurry region is calculated by counting the number of intersections between the AABBs of the objects and the frustum encompassing the blurry region.

For third-person cameras, camera tracks and follows an object of interest. Occlusions indicate the existence of other objects in the path between the camera and the object of interest. In order to measure occlusions on LFDs, a frustum in front of the object of interest in the direction of camera is considered. Figure 3.9b shows the top view for the frustum with respect to the object of interest, where the front plane of the frustum is similar to that of the display. The back plane of frustum is the same as the front plane of the AABB of the object. The right and left planes are parallel to those of the display, but



(a) Collision camera

(b) Suspension camera

(c) Falling camera

Figure 3.10: Physical simulation of cameras on LFD [119]

bounding the object of interest instead. As for the top and bottom plane constituting the frustum, they are constructed from the top and bottom lines on the AABB of the object passing through the observer line. The number of objects existing in the occluded region is measured by calculating the number of intersections between the AABBs of the objects and the constructed frustum.

Evaluation and results

To evaluate the physical simulations on the LFD, three test cases (scenarios) were implemented (depicted in Figure 3.10) and assessed using the metrics discussed earlier in Section 3.5.2.

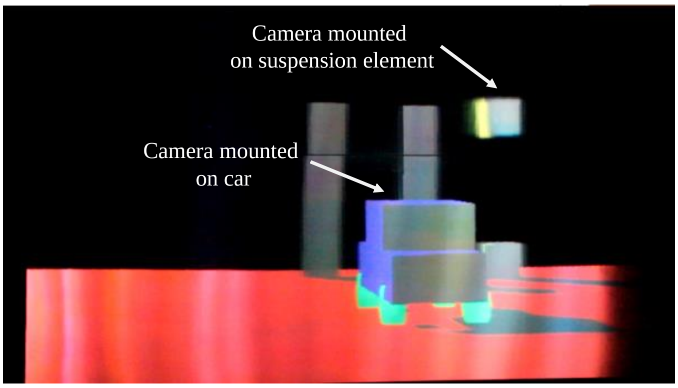


Figure 3.11: Suspension camera scenario

The first scene includes a car moving towards a set of columns. The car accelerates to collide with one of the columns, resulting in its fall. The camera is mounted twice on the car as a first-person and as a third-person camera, and once on the collided column.

The second scene contains a suspension element and a car (illustrated in Figure 3.11), where the camera is mounted once on the suspension object with the car placed in front of the suspension element and once on the car itself, looking towards the suspension element. It depicts the effect of a camera mounted on a suspended platform.

The third scene depicts a camera falling from an altitude towards the ground until it collides with the latter. There is a total of 50 objects (boxes and cylinders) on the ground [119, 117].

Applying the metrics outlined in Section 3.5.2 for each of the three physical scenarios depicted in Figure 3.10 yields the findings presented in Table 3.2 [117].

Subjective evaluation

Subjective tests were conducted on the HoloVizio C80 HOP LFD in a laboratory environment, isolated from audiovisual distractions. The lighting conditions were approximately 20 lux. The test participants could freely observe the visualized contents within a well-defined VVA. The viewing distance (i.e., the screen-perpendicular dimension of the VVA) ranged from 4 m to 8 m. The selection of the minimum value originated from the constraint that test participants should not be located between the screen and the optical engine array of the LFD, as it may risk invalid LF through ray occlusion. The maximum value was based on the findings of Kara *et al.* [169], which define the threshold at which the visual experience becomes closer to 2D due to the lack of disparity between the rays addressing the eyes of the viewer, with respect to a single point on the screen. As for the

Table 3.2: Metrics tested for realistic physical camera simulations [117]

Scenario	**Number of objects colliding**	**Number of objects in blurry region**	**Number of objects in occlusion region**
Collision camera scenario (First-person camera on car)	2	4	3
Collision camera scenario (Third-person camera on car)	0	3	3
Collision camera scenario (First-person camera on column)	2	3	3
Suspension camera scenario (First-person camera on suspension)	0	5	0
Suspension camera scenario (Third-person camera on car)	0	2	0
Falling camera scenario	0	17	51 (All)

distance for sideways mobility (i.e., the screen-parallel dimension of the VVA), the width of the screen was used. The top-down view of the test scenario is shown on Figure 3.12.

The subjective tests were completed by 21 participants. 9 (42.85%) of the test participants were female and 12 (57.15%) were male. The test participants were pooled from an age range between 20 and 65. The average age was 29.

Physical simulation tests were carried out in order to test the plausibility of these camera movements on the LFDs. Test participants were asked to view these scenarios on the LFD (illustrated in Figure 3.10) and evaluate the visualized contents through a series of questions. However, before displaying the videos on the LFD, participants were asked about their preferred simulation camera type ("first-person camera", "third-person camera" or "equal"). The same question was then re-asked at the end of the experiment –after watching the videos– based on what was visualized on the LFD. Regarding each scenario –where a different camera type is shown on the LFD– participants were asked several questions to evaluate the visualized contents based on the camera type. The questions (i.e., assessment tasks) were the following:

- Rate the dizziness and the loss of focus resulting from the camera motion (5-point Absolute Category Rating (ACR) scale).
- Choose your preference between standing still and moving around the LFD while observing the visualized content ("moving", "standing" or "equal").
- Rate occlusions with respect to the camera, the blurriness of objects and the camera collisions (5-point ACR scale).

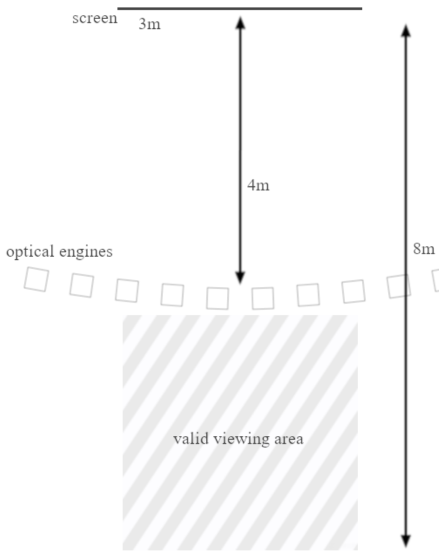


Figure 3.12: Top-down view of the test scenario [123]

- Indicate your personal preference whether or not it is better to visualize the contents on the LFD rather than on a 2D display ("better", "worse" or "same").

Regarding all the visualized scenarios, participants were asked to choose the type of simulation camera they preferred –first-person versus third-person– with respect to the LFD. Most participants (76.2%) preferred the third-person camera for visualizing the contents on LFDs, compared to conventional 2D displays. In general, for such conventional 2D displays, participants had a variety of choice between the first-person and third-person cameras, however, for LFDs, almost all of them considered using the third-person camera as the best option. This is expected, considering that the first-person camera has a closer look at the object of interest, leading to an increase in the blurriness of objects due to the optical limitations of the LFDs. Therefore, using third-person cameras on LFDs or considering a "visual hack" to imitate the first-person camera effect is preferable. As a means of doing so is to increase the scaling of the ROI with respect to the main element. The ROI is a box-shaped region within which all objects are visible. That is to say, for LFDs, the ROI is a virtual region within which all elements are visible on the LFD [81].

Considering all physical scenarios, most of the test participants (66.6%) preferred standing in front of the screen instead of walking around. Particularly, the preference of the participants for moving or walking depended on the type of physical camera used. Most users preferred standing for the collision camera (76.2%), followed by the falling camera (71.4%), whereas the suspension camera (52.4%) had the least amount of votes. This is due to the fact that for the collision camera, extreme motions were visualized on the LFD because of the camera collision in the scene. Hence, it could be inferred that

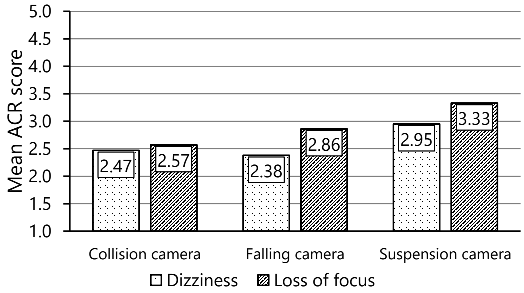


Figure 3.13: Diagram to illustrate the ratings for dizziness and loss of focus for the different camera types [123]

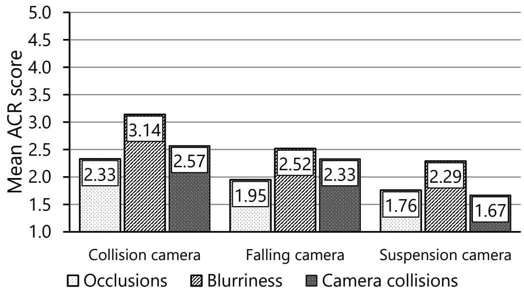


Figure 3.14: Diagram illustrating the ratings for the metrics proposed for the camera simulation types [123]

the movement of participants and the camera motions are indirectly proportional. Test participants were also asked to rate the dizziness and loss of focus (higher values indicate higher levels of dizziness and loss of focus). Participants rated the dizziness and loss of focus with an average of 2.60 and 2.92, respectively. Details for the mean scores for the different camera types are illustrated on Figure 3.13. Poor ratings for both the dizziness and loss of focus emphasize that irritating visuals may arise when using first-person cameras on LFDs. Consequently, 71.4% of the participants preferred conventional 2D displays to LFDs when watching physical camera simulations. It is noted that among the different camera types used, suspension camera had the lowest percentage (38.1%) for promoting the usage of conventional 2D displays compared to LFDs. This is due to the fact that the suspension camera had the least amount of movement in the scene.

Although the metrics discussed in Section 3.5.2 are objective –since they mostly count the number of elements in the scene that are occluded, collided or blurred– they could still be used in a way for subjective testing as well. In other words, spectators were asked to rate these metrics from their perspective instead of objectively applying these metrics. Figure 3.14 summarizes the results for this task. Starting with the occlusions with respect to the camera, test participants were asked to rate them from 1 to 5 (higher values indicate more perceived occlusions). Averages of 2.33, 1.95 and 1.76 were given to the collision, falling and suspension cameras, respectively. For the blurriness of objects,

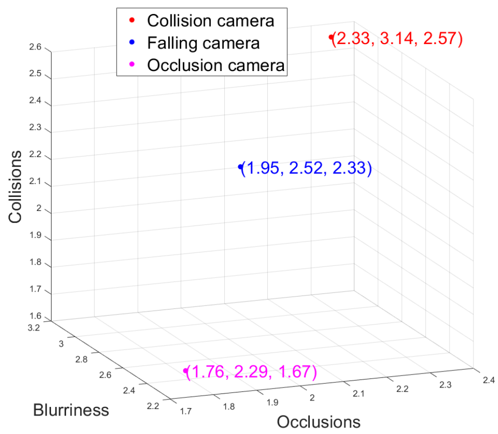


Figure 3.15: Graph showing the values of the metrics with respect to each camera type [123]

the same rating scale was applied (higher values indicate a higher extent of perceived blur). For the collision, falling and suspension cameras, average ratings of 3.14, 2.52 and 2.29 were given, respectively. Finally, the camera collisions were also rated from 1 to 5 (higher values indicate more collisions) with averages of 2.57, 2.33 and 1.67 for collision, falling and suspension cameras, respectively.

It is quite evident how the intensity of camera movements affects the metric measures. Figure 3.15 depicts the relation between camera motion and the different metrics (occlusions, blurriness and collisions). From the figure, the directly proportional relation between the different metrics measures and the intensity of camera motion is deduced. In other words, as the camera motion increases –becomes more "vigorous"– occlusions, blurriness and collisions increase as well. This, however, leads to poor perceived visual quality, hence the initiative to consider slight camera motions for LFDs [123].

Discussion

Since LFDs provide observers with an immersive 3D experience without the need of additional viewing gears, they evidently earned their place within the world of cinematography. As seen in Figure 3.10, the possibility of creating realistic physical contents on LFDs exist; however, not all physical camera motions produce plausible results. This is due to the optical limitations of LFDs, resulting in a degraded quality of visualized content when using a first-person camera. This was furtherly proved by the test participants, preferring third-person camera to first-person camera on LFDs. Hence, the simulation of the first-person camera effect on LFDs remains open to further research. Additional deterioration occurs with the speeding up of camera motions. In other words, the less camera motion, the more plausible visualization for the contents on LFDs. Accordingly, more research efforts and investigations are required to furtherly assess and improve realistic physical camera animations on LFDs [123].

3.6 Conclusion and future work

LFDs present the viewers with an immersive, glasses-free 3D environment, and according to the state-of-the-art technologies, at the time of this work, it can be stated that they are the most suitable for cinematographic purposes. LFs are captured by means of LF cameras (narrow-baseline and wide-baseline). Applying the conventional rules of cinematography and techniques for LF cameras can be challenging, especially for wide-baseline systems.

Hence, we presented a robust framework built for evaluating various camera animations –and typical scenarios used in simulation and cinematography– in the context of LF visualization. Realistic physical motion formats were included and investigated in our study, and they were assessed on a real LFD, using various metrics. The results indicate that the visualization of some of the motions are not adequate for LFDs due to optical limitations. Hence, these limitations should be taken into account when designing camera motions for LFDs. Moreover, we presented empirical studies on the user preference regarding the numerous realistic physical camera motions, where controversial opinions were given as the camera motions involving lots of oscillations and collisions resulted in loss of focus on the LFD.

As for potential future continuations of our scientific contribution, possible extensions of the test scenes to multiple commonly used cases can be carried out, with the aim of capturing numerous problems when it comes to camera path planning and interaction. Hard-to-navigate scenes –such as interiors or prop rooms with open sides– can be explored and a set of recommendations is to be compiled for all relevant scenarios. Additional important parameters for such scenes shall also be explored, including optimal camera placement, angular limits, camera speeds and many more. Moreover, considering the realistic physical camera simulations, additional physical properties could be furtherly added, including physical stress on the frame of the mount, oscillations when accelerating/decelerating, camera arrays with re-configurable frames and numerous others. Furthermore, the planning of lighting from different angles should also be considered. In other words, making sure that the necessary lighting for simulation should be exactly the same from all angles.

Addressing camera motions on LFDs remains an open research issue, including the choice of the best camera motions suitable for the LFDs and the means of implementing the other controversial movements resulting in less visual issues. Overall, more investigation is needed when it comes to the usage of first-person camera on LFDs. A number of alternatives can be proposed to imitate the first-person camera on LFDs without artefacts arising due to the limitations of the LFDs [119, 117, 123].

CHAPTER 4

Interaction techniques for light field displays

UIs have become more diverse with the rapid proliferation of new nontraditional interface components and devices such as 3D trackers. Accordingly, for an efficient interaction design, a thorough understanding is needed for each device's advantages and limitations –in addition to ergonomics, in order to have an intuitive mapping between the different methods of interaction and the corresponding device.

Among the factors affecting the choice of interaction techniques used in the UI design are the visual display under consideration and the user's perceptual and evaluative processes, rather than direct input interactions. As new types of displays emerge, interaction techniques evolve promptly. In general, visual displays are classified as either fully immersive or semi-immersive devices. Fully-immersive displays encompass environments that obscure the real world, exemplified by HMDs, whereas semi-immersive displays afford visibility of both the virtual and physical worlds [49]. An example to semi-immersive displays are LFDs.

LFDs immerse the users without the need for additional viewing devices. Despite the numerous advantages and attractive capabilities of such glasses-free 3D displays, their UI methods are quite complicated and they are currently underwhelming when compared to conventional 2D displays, due to the fact that visual feedback can only be rendered sharply on the emission surface of LFDs. The sharp rendering of UIs is a necessity, as blur may hinder their fundamental functions.

When it comes to 2D displays, many user interaction techniques and interfaces have been devised. Rendering a UI on a 2D display could be done in various ways, such as rendering overlays on top of the rendered scene, or by using billboards. These are extensively used in modern video games. Meanwhile, UI design for immersive virtual environments (e.g., AR, VR and MR) have been extensively investigated and redesigned, whereas the same cannot be stated for interactive 3D environments. These refer to simulated environments or depictions of actual surroundings wherein individuals can engage with various elements of the setting and navigate in real time [151]. Hence, UI design is considered to be a major component in an interactive 3D environment.

Although LFDs contain immense potentials, only basic UIs have been devised thus far, including FOX (Focus Sliding surface) [222, 221], which grants users the option to scale and to rotate 3D objects. Conversely, our work shifts the focus from direct user-device interactions to passive observation. We do not require any input from the users during the interaction; rather, users are asked to passively observe the content on the LFD and provide subjective feedback based on their visual experience.

In this chapter, we visualize the theater model on a real LFD, then test the different interactions by means of a monitor room. The theatre model is analogous to real-life theaters, where viewers may observe the theatrical presentation on the stage from various angles. The motivation to choose the theater model was the fact that LFDs similarly allow multiple simultaneous viewers within their FOV, in which the content can be observed in an anglularly-dependent manner. Moreover, from the users' perspective, the theater model is thus familiar and it provides high-quality visual feedback. Furthermore, theater stages encompass a lot of interactions, including rigging and flying systems, pulleys, rotating stages, lights, curtains, etc [118].

The chapter is structured as follows: Section 4.1.1 offers an overview of 3D interaction techniques, while Section 4.1.2 examines the 3D interaction techniques specific to LFDs. Section 4.2 outlines the various presentation models considered for implementing LF interactions. The methods for rendering the theater and monitor room, along with the results of these interactions are discussed in Section 4.3, illustrated by images of the actual visualization on LFDs. It is demonstrated that achieving plausible results without noticeable visual artifacts is challenging, yet possible. The scientific contributions of this chapter also highlight the various novel UIs for future LF systems and services. The focus of this work is on how users perceive and subjectively evaluate the content presented on the LFD, rather than how they interact with the system through input devices. Subsequently, subjective evaluations were conducted to assess the viability of 3D interactions on LFDs, with findings from empirical studies concerning perceptual preferences regarding potential interaction techniques. Finally, Section 4.4 concludes the chapter, pointing out potential directions for future work.

4.1 3D interaction techniques

Interactive 3D environments are those artificial environments or representations of real environments where users can interact with elements of the scene and navigate in real time. Interaction in such environments can be summarized in three tasks: (i) navigation, (ii) selection and manipulation and (iii) application/system control [132, 49, 151, 118].

4.1.1 Overview

In this section, we elaborate extensively on each of the sub-tasks used in the creation of UIs for 3D interactive environments.

Navigation

As is customary in the majority of 3D interactive environments, the area observed from a singular viewpoint commonly constitutes a portion of the complete scene, wherein scenes are typically observed through one or more viewports representing the respective (virtual) camera(s). Thus, navigation is carried out by modifying the viewing configurations within the environment [151]. Categorized according to the objective of navigation, the latter can be divided into three primary classifications: (i) exploration, (ii) search, and (iii) maneuvering [49, 118].

Selection and manipulation

This task endeavors to carry out at least one of the subsequent functions: (i) selecting objects, (ii) translating objects (repositioning), (iii) rotating objects (reorienting), and (iv) scaling objects. In essence, this task involves both the selection and alteration of an object within the scene.

Several methods exist for accomplishing the selection and manipulation task, none of which can be deemed as the "optimal" choice, as their effectiveness relies on the specific scene and task at hand. However, when designing any method, it is imperative to consider constraints and limitations related to DOF [49, 151, 118].

Application/system control

This task lies outside the scope of the virtual environment and pertains to the communication between the user and the system [132]. Furthermore, it offers visual feedback whereby user-issued commands are employed to alter the system's state or the interaction method [49]. This commonly entails UIs such as overlays and menus [118].

4.1.2 3D interaction techniques for light field displays

The application of these 3D interaction tasks to LFDs imposes many challenges that need to be dealt with. In this section, we discuss these limitations for HOP as well as FP LFDs [118].

Navigation

Unlike conventional 3D applications, where the view settings can be easily altered based on the position of the viewer, changing the view settings for LFDs is not feasible. For HOP LFDs, it requires changing either the observer line or the set of camera positions (1D array/arc of cameras) within the scene. On the other hand, changing the view settings for FP displays requires either moving the observer rectangle or the 2D camera array setup. In addition to changing the observer line or the camera arrays, modifying some view settings for LFDs is not feasible either. View settings include the perspective/projection parameters, aspect ratio, resolution, near/far clipping planes and focal length. Accordingly, changing the horizontal FOV for the horizontal parallax is not possible, whereas for FP systems, changing the FOV is not possible. LFDs have an angularly selective nature that allows multiple viewers to view the same scene from different angles on the observer line/rectangle. Since navigation in 3D environments is mostly concerned with the view settings, the view matrix should be investigated as well. Similarly to 2D visualization, HOP and FP LFDs have 4×4 viewing matrices. Yet unlike the 2D scenario, where the view matrix transforms the world coordinates to camera coordinates, the view matrix of LFDs does not perform the same task. There is a 1D or 2D array of cameras in case of HOP and FP displays, respectively. Accordingly, the view matrix cannot convert world coordinates to camera coordinates due to the fact of having multiple cameras. Instead, the view matrices of LFDs convert from the world space into the ROI within the world space. The ROI is the area in which the objects are viewed, whereas anything outside the ROI is clipped [118].

Selection and manipulation

In order to select or manipulate an object inside the environment, the object needs to be visible. For LFDs, the object under consideration must be visible from all points on the observer line or on the observer rectangle in case of HOP and FP displays, respectively. Full visibility does not necessarily mean that the entire object is visible. Instead, it requires that a substantial portion of the object, such as 50%, is observable from all points along the observer's line/rectangle. Regarding the dependability of results on the viewing angle, selection becomes unsuitable for LFDs due to their angularly selective nature. In general, an LF system is composed of multiple optical modules placed behind a semi-transparent screen. Hence, for LFDs, image space is defined for every optical module as the coordinates of its texture [90]. Due to the fact that such displays have many optical modules, the selection of visualization based on image space for LFDs is impossible. Summa summarum, selection for LFDs is unattainable in image space or with dependency on the viewing angle [118].

Application/system control

LFDs act as a viewing window to the 3D world, providing 3D depth perception for the users. As a result, rendering to overlays on LFDs is not feasible, as it breaches this concept of perception. Additionally, rendering into overlays depends on the image space, and therefore, it cannot be applied to LFDs due to the aforementioned reasons. Possible alternatives to rendering into overlays include rendering to the environment or the sharp plane. The latter is preferred from the perspective of the viewers [118].

4.2 3D presentation models

A presentation model is basically a combination of three interaction methods: (i) navigation, (ii) selection and manipulation and (iii) application/system control. Practically speaking, presentation models are used to view and arrange objects within a scene. Furthermore, they include a set of techniques for interaction and manipulation with the items present in the scene. In this section, we investigate the different 3D presentation models. This is followed by our proposal of possible presentation models for LFDs.

4.2.1 Overview

Unlike 2D interaction techniques, interaction in 3D environments is more challenging, since mapping between the 2D controls and the corresponding 3D functions is not straightforward at all. The following list contains the most relevant interaction techniques [151, 118]:

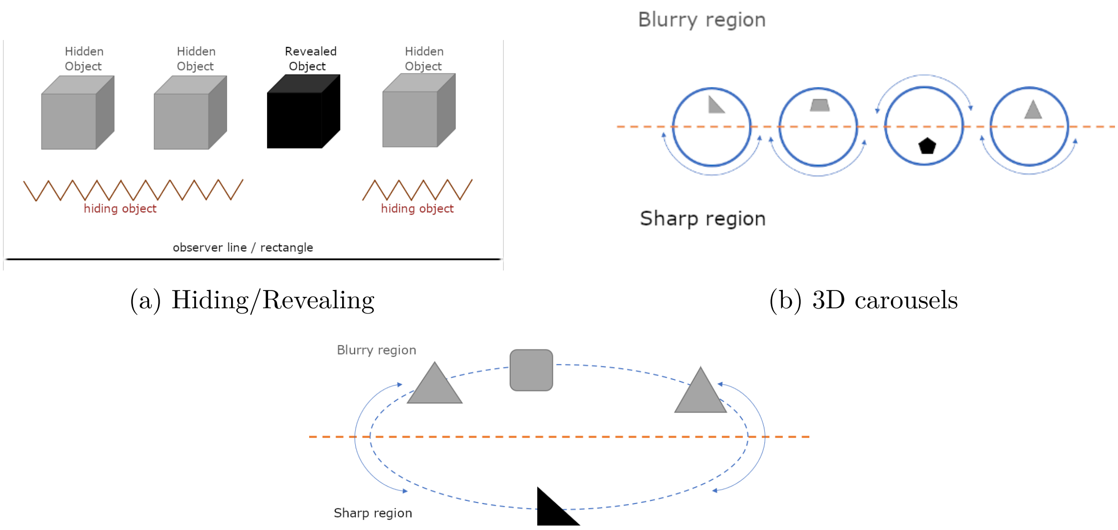
- Line-up and light: All objects in the scene are lined up. A spotlight is used to focus on the main object under selection and manipulation.
- Change focus: This technique is usually used in cinematography. It shifts the attention of spectators by changing the focus from one object/character to another. It is also known as rack focus [53].
- Animation/Freezing of selected object: This technique is adopted in many video games where the selected character is being frozen/animated to indicate its selection.
- Selection halo/circle/arrow: This is one of the most common techniques in 3D video games (particularly FIFA video games [1]) where a halo/selection circle/arrow is drawn on/above the character/object under selection.
- Decals: They are used in video games where additional textures are applied over the underlying textures.
- 3D text: Self-explanatory.
- Overlays: They are used in video games to present the background graphics with rich colors. Game controllers are usually rendered into overlays in order to be visible throughout the game [128].
- 3D carousels: Carousels were used extensively in the video games of the 1980s, where players were asked to enter their initials in order to record their high scores. Selecting items by means of carousels is easily understood by users, in addition to enriching to context with a sense of engagement by means of rotation [47].

[1] https://www.ea.com/en-gb/games/fifa

4.2.2 Presentation models for light fields

Unlike conventional 3D visualization, presentation models for LFs have not been investigated yet. One key point when dealing with LF presentation models is scene arrangement. Basically, arranging objects in a single row is less challenging than arranging objects along an arc or in multiple rows. Another major point in LF presentation models is the state of camera motion; whether the camera is static or dynamic. In this part of the section, we propose presentation models for LFDs [118].

- Navigation: Due to the various issues and challenges imposed by LFDs, it is preferable to use static cameras for scene navigation. Otherwise, objects would move back and forth between the sharp and blurry regions of visualization. In addition to static cameras, using free cameras (analogous to virtual on-the-fly cameras) is also possible.
- Selection and Manipulation: As stated in Section 4.1.2, rendering to overlays is quite difficult for LFDs. Typically, overlays are rendered on the closest plane to the observer. However, in case of LFDs, choosing the closest plane may result in blurriness due to the display optics. An alternative solution is to render on 2D area(s) on the plane of the screen in order to view the overlay sharply. However, any object along the way between the 2D area(s) and the viewers would block the overlay. Therefore, a possible solution is to cull or to set the transparency of the objects in the occluder region in order to avoid overdraw. In addition to rendering to 2D area(s), rendering to 3D regions can actually be effective. It can be performed by using the following proposed techniques:
  - Bounding box outlines: Using the AABBs of the objects to do the selection as drawing 2D shapes around the selected object would not work in 3D.
  - Color change: Changing the color/material (e.g., emission or light) of the selected object.
  - Decals: Changing the texture of the selected object.
  - Selection tube/halo/circle/arrow: see previous subsection.
  - Animations: For objects being manipulated or selected by means of animation, spatial bounds should be considered.
  - Hiding/Revealing: Objects in the scene are aligned in one row in the sharp region of the screen of the LFD. An extra object is used to hide all objects in the scene except for the object under selection/manipulation. Figure 4.1a illustrates this technique.
  - Change of object arrangement/spatial position: An example for this technique is using the line-up method, where all items of the scene are placed in one row in the blurry region of the screen. Whenever an item is selected, it moves forward/backward into the sharp region, whereas the remaining items retain their blurry states. Hence, the selected item is sharper in comparison and



(a) Hiding/Revealing

(b) 3D carousels

(c) Single carousel with elliptical path

Figure 4.1: Selection and manipulation for LFDs [118]

shall attract the attention of the viewer(s). Movement can be performed in a straight line or by means of 3D carousels. This could be applied by placing half of the carousels in the sharp region while placing the other half in the blurry region. In this case, all objects are placed in the blurry part of the carousel and rotation is applied only to the carousel holding the object under selection in order to position it in the sharp region. Figures 4.1c and 4.1b illustrate two ways for using 3D carousels on LFDs. The first figure depicts the placement of all items on a single elliptical carousel where the items under selection are placed on the front part of the carousel (i.e., in the sharp region). The second figure places each item on an individual carousel. Items are positioned on the carousel in a way that they inhabit the blurry region, whereas the carousel holding the item under selection is rotated in order to place the item in the sharp region of the screen.

- Application/system control: For LFDs, system control can be achieved by rendering the UI into 2D area(s), in a way similar to that stated earlier regarding selection and manipulation. As an alternative, the separation of the main scene and the 3D controls could be performed spatially while providing feedback of the main scene on the 3D control geometry. In all the techniques used for application/system control, widget design needs to take into account visibility along the observer line/rectangle.

Table 4.1: Presentation models for LFDs [118]

Presentation model	Navigation	Selection and manipulation	Application/system control
Line-up	Static Camera	- Bounding box outlines - Color change - Decals - Selection tube/halo/ circle/arrow - Animation - Change of object arrangement/spatial position	Switch 2 scenes
Carousel	Static camera	Change of object arrangement/spatial position	Switch 2 scenes
3D sphere	Static camera	Change of object arrangement/spatial position	Switch 2 scenes
CAD/CAM	Free camera	AABB	2D areas on screen + spatial separation for navigation feedback
Medical	Orbiter camera	Select on 2D area(s)	2D areas on screen + spatial separation for navigation feedback
Theater	Static camera	- Change colors - Change of object arrangement/spatial position - Hiding/Revealing	Switch 2 scenes

Table 4.1 introduces the possible 3D presentation models for LFDs by combining some of the aforementioned techniques to constitute plausible yet effective presentation models [118].

4.3 Theater presentation model for light field visualization

So far, we have proposed and investigated different presentation models that could be used for LF visualization. Among these suggested models, the theater model is potentially the most efficient, and thus may provide the best visual experience. Similarly to theaters, multi-user LFDs have the same viewing experience, as they allow numerous simultaneous viewers within their FOV, in which the content can be observed in an angularly-dependent manner. In addition to allowing the effective presence of simultaneous observers, high-quality visual feedback is provided by means of a monitor control room. Furthermore, theaters encompass lots of interactions and animations for their presentation elements [118].

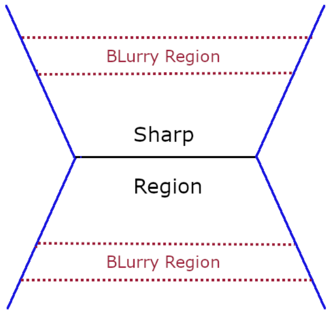


Figure 4.2: Top view of LFD setup [118]

4.3.1 Technical considerations

In order to test the different interaction methods for LFDs, a proscenium theater model and a monitor room were modelled using MAYA [2], and they were visualized on the HoloVizio C80 LFD [3]. The C80 has an aspect ratio of 16:9 with a screen size of $3\,\mathrm{m} \times 1.8\,\mathrm{m}$, and a 40° horizontal viewing angle. Although animations and interaction methods are easily implemented for and viewed on conventional 2D displays, 3D LFDs impose some challenges and limitations [71]. One of these challenges is the fact that only a certain portion of the visualization area supports sharp rendering, and thus the focus of the content is limited to that specific area.

Figure 4.2 shows the top view for a typical setup of LFDs. The black line depicts the screen, whereas the blue lines show the viewing angles. The dotted lines encompass the blurry regions. The area surrounding the screen is the one where objects are rendered sharply. Hence, if an object is animated on a line that is perpendicular to the observer line/rectangle, the object moves into and out of focus as it crosses the blurry and sharp areas. Accordingly, it is better to consider animations along any plane perpendicular to the screen (i.e., animations that include right/left or top/down motions). However, if animations along the lines perpendicular to the observer line/rectangle are to be considered, then they should be done within a small range in order to avoid the potential crossing. Therefore, the theater model fulfils these requirements by the animations of rigging/flying system and curtains [118].

4.3.2 Utilization of the theater model

As stated earlier, both the theater model and the monitor room were modelled in MAYA. The models of *bunny*, *buddha* and *teapot* were imported from the Computer Graphics Archive [224]. Figure 4.3 shows the full view theater model and the monitor room as displayed on the LFD. The monitor (control) room depicts the application/system control in the presentation model of the theater. Switching back and forth between these

[2] https://www.autodesk.com/products/maya
[3] https://holografika.com/c80-glasses-free-3d-cinema/

(a) Theater model

(b) Monitor control model

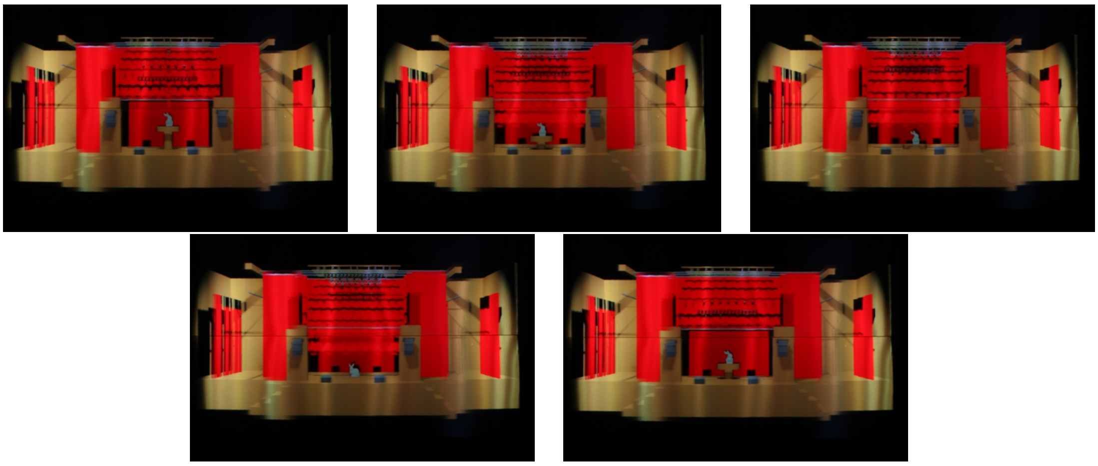
Figure 4.3: Theater and monitor room models [118]

views is achieved by pressing buttons. The corresponding animation/lighting is activated within the theater model and the monitor room is viewed at that time with a display screen showing the current theater view. Once the corresponding animation is activated, the view switches back to the theater model. Navigation within the theater model is performed via a static camera [118].

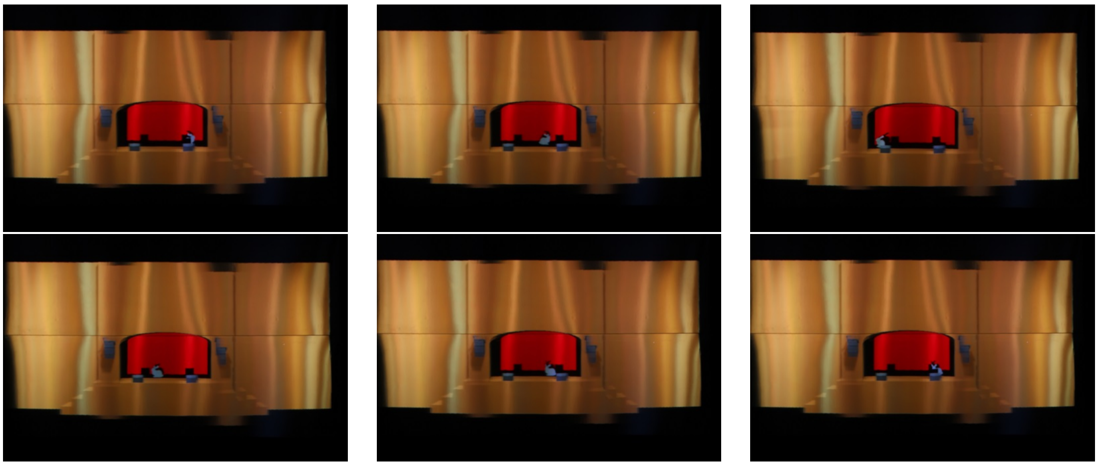
4.3.3 Evaluation and Results

Figure 4.4 presents the complete view of the various theater model scenarios displayed on the LFD. In our work, we tested different ideas for the selection/manipulation of objects [118]:

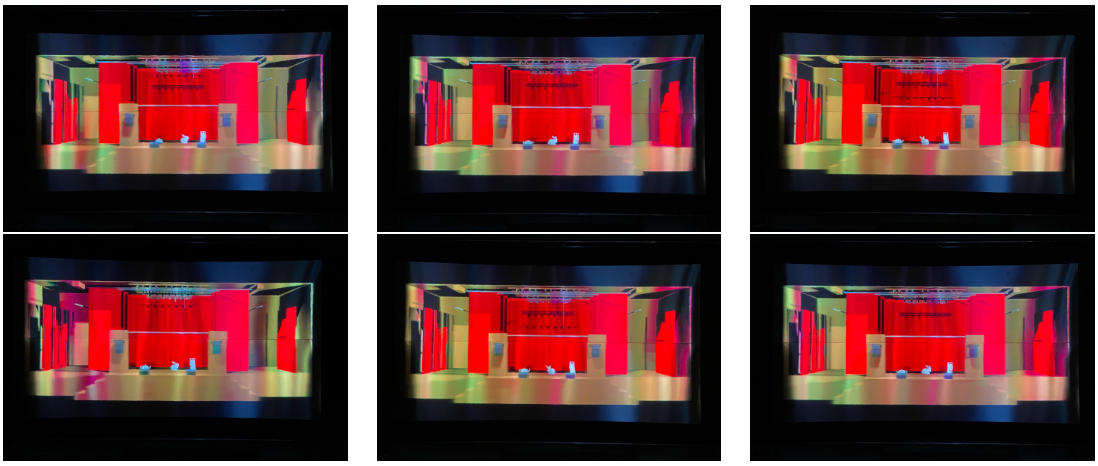
- A theater model with a rotating stage, where the rotating stage is placed in the sharp region of the LFD. Hence, the movement of the theater stage in the up/down direction and rotation do not cause any blurring effects.
- A theater model with an object animated along a path to change its position.
- Using curtains to hide some elements while displaying others under selection to apply the hiding/revealing technique.
- Spatial positioning of presentation elements is done within the sharp region in a plane parallel to the screen, thus avoiding the problem of moving in and out of the sharp region (e.g., animation of curtains and rigging/flying system). Animation of curtains and flying systems is done within their plane (right/left and up/down motion), hence avoiding the problem of moving in and out of the blurry region of the LFD.
- Usage of rotating stages where half of the stage is placed in the blurry region and the other half in the sharp region. Spatial positioning of objects that are selected is done within the sharp region.



(a) Rotating stage while moving up/down



(b) Moving the object along a path



(c) Usage of 3D carousels

Figure 4.4: Theater model simulation on LFD [118]

- Animating the spotlights and spotlight reflectors by rotating them within a very small range and thus they do not cause an issue on the LFD.

4.3.4 Subjective evaluation

Apparatus and test participants

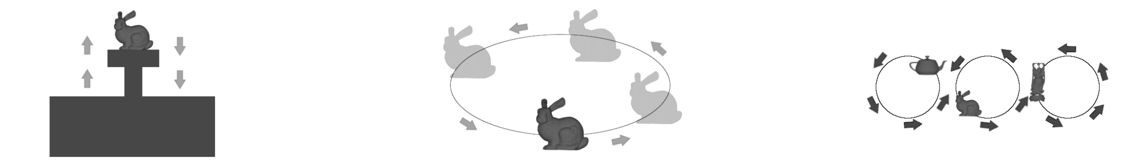
Subjective tests were conducted on the HoloVizio C80 HOP LFD, with brightness calibrated to 1000 cd/m^2. The tests were carried out in a laboratory environment, isolated from audiovisual distractions. The lighting conditions were approximately 20 lux. The test participants could freely observe the visualized contents within a well-defined VVA. The viewing distance (i.e., the screen-perpendicular dimension of the VVA) ranged from 4 m to 8 m. The selection of the minimum value originated from the constraint that test participants should not be located between the screen and the optical engine array of the LFD, as it may risk invalid LF through ray occlusion. The maximum value was based on the findings of Kara *et al.* [169], which define the threshold at which the visual experience becomes closer to 2D due to the lack of disparity between the rays addressing the eyes of the viewer, with respect to a single point on the screen. As for the distance for sideways mobility (i.e., the screen-parallel dimension of the VVA), the width of the screen was used. The top-down view of the test scenario is shown on Figure 3.12 (Found in Chapter 3).

The three distinct scenarios concerning the theater model (depicted in Figure 4.4) were rendered on the HoloVizio C80 LFD and presented to several test participants to assess their visual credibility. The subjective tests were completed by 21 participants: 9 (42.85%) of the test participants were female and 12 (57.15%) were male. The test participants were pooled from an age range between 20 and 65. The average age was 29 [123].

Subjective tests on 3D interactions

Test participants were asked to rate the test stimuli in terms of visual plausibility. Regarding the navigation task, users were asked whether they prefer the static camera or the moving camera on a 3-point scale ("static", "moving" or "unable to decide"). As for the second task, participants were asked to choose the preferred manipulation and selection model from their personal visual perspective ("moving stage", "single elliptical carousel" or "multiple carousels"). In addition to the selection of the model, users were asked to rate the motion of the curtains and the flying system on a 5-point ACR scale ("bad", "poor", "fair", "good" and "excellent")[4]. Lastly, regarding the application/system control task, users were asked whether they prefer the swapping process between the two scenes over the Graphical User Interface (GUI) buttons displayed over the main scene ("yes", "no" or "unable to decide"). For every investigated scenario, participants were asked whether they prefer standing still or walking around ("moving", "standing" or "equal"), less or more interactions displayed ("less interactions", "more interactions" or

[4] ITU-T Rec. P.910 : Subjective video quality assessment methods for multimedia applications



(a) Stage movement (upward / downward)

(b) Rotating the main character on an ellipse

(c) Rotation of different models on separate carousels

Figure 4.5: Selection and manipulation for the theater model [123]

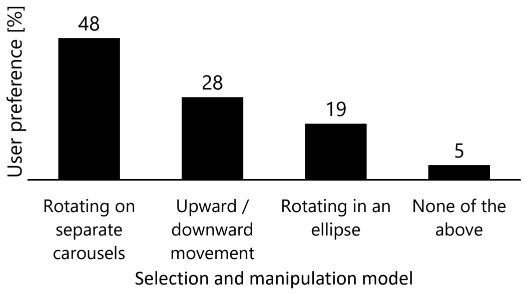


Figure 4.6: Users' preferences for the different selection and manipulation techniques used [123]

"equal") and whether or not they personally deem it better to visualize the content on the LFD compared to conventional 2D displays ("better", "worse" or "same") [123].

Results and discussion

As mentioned in Section 4.3.2, in efforts to imitate the theater environment, a static camera was used for visualizing the theater stage viewed by various spectators. When asking participants about their camera preference, 61.9% agreed that the static camera was the best since LFDs already create a perceptual 3D effect. Hence, in case a boost is needed in the achieved 3D effect, walking around the screen will suffice. Accordingly, a moving camera may be visually disturbing due to the extra added motion, leading to dizziness and loss of focus.

Considering the selection and manipulation task, a variety of models were implemented on the LFD (see Figure 4.5). They include the upward/downward stage movement, the rotation of the main model on an elliptical carousel and the rotation of different models on separate carousels.

Participants were asked to choose their personally preferred selection/manipulation model. As shown on Figure 4.6, almost half of the users preferred the rotation of different models on separate carousels. The reason is that the rotation of models incorporates more interactions on the LFDs, hence, enhancing the perceived 3D effect. This is followed by the upward/downward stage movement selection model with more than a quarter of the votes. Finally, 19% of the test participants preferred the rotation of a single model on an elliptical carousel, as it includes the least amount of motion on the LFD. In addition to applying animation to the main element on the stage –as a means of selection and manipulation– animations were applied as well to the curtains and the theater's flying system. For curtains, side-to-side movements were carried out, in addition to the conventional up/down curtain motion. However, for the flying system, conventional vertical motion was achieved while alternating between the different light bars. In addition to the aforementioned motion, rotations of the different traditional and ellipsoidal spotlights were performed. Test participants were asked to rate the movement of the curtains and the flying system on a scale from 1 to 5, where higher scores indicate higher user satisfaction. This resulted in scores of 2.76 and 3.81, respectively, further emphasizing the additional 3D effect achieved by increasing the visualized interactions.

As shown on Figures 4.4a and 4.4c, the theater scene is viewed along with its backstage, unlike on Figure 4.4b. Participants were asked to indicate their visual preference between these two alternatives. Most users (76.2%) decided upon the theater scenes showing the backstage as they included more interactions, hence, increasing the perceived 3D effect. In addition to increasing the 3D effect by means of interactions, walking around the LFD adds to the achieved sense of immersion. Accordingly, 57.14% of the participants preferred walking around compared to remaining still. Regarding the last task of 3D interactions, application/system control illustrated some differences in the participants' visual preference. Users were asked to choose between two scenarios: (i) control buttons displayed on the main theater scene and (ii) switching between the theater scene and a monitor room with the control buttons while providing a visual feedback for the current theater state. A total of 57.14% preferred to have the control buttons on the same screen as the main scene, while 33.33% favored the second scenario. Although, at first sight it seems that incorporating the control buttons on the main scene is the best option, the problems of breaking the 3D immersion on the LFDs may arise. The reason is that it is quite obvious on the LFD that the GUI is a 2D overlay visualized on a 3D scene. It remains an ongoing research question how the feedback for 3D scenes should be conveyed to the users of LFDs. Overall, 76.19% preferred the interaction techniques on the LFDs compared to the conventional 2D displays [123].

4.4 Conclusion and future work

Interaction techniques for wide-baseline LFDs is a new, yet promising research topic. In our work, we investigated the possible presentation models for LFDs and used the theater model for illustration and testing. Furthermore, our work presented empirical studies on the user preference regarding the different interaction techniques. As a result of these studies, some notable inferences can be deduced. Generally, test participants agreed that LFDs provide a sense of 3D immersion, hence they are better than conventional 2D displays for visualizing specific contents. Participants not only concurred on the feasibility of the 3D interactions but also advocated for the inclusion of additional interactions on the LFDs, as it contributes to the perceived 3D perspective. Finally, the subjective evaluation indicates an inverse relationship between the interactions on the LFD and the mobility of participants. In other words, the less interactions and movements on the LFD, the better for users to walk around, and vice versa.

As future continuation of this work, other presentation models are encouraged to be visualized on LFDs to further validate the plausibility of the investigated 3D interactions. Addressing 3D interactions on LFDs remains an open research question, specifically for the application/system control task. Different approaches for an application/system control feedback on LFDs can be still investigated in order to come up with an efficient solution to display the GUI while giving feedback to the current scene, without breaking the 3D immersion [118, 123].

CHAPTER 5

Towards HDR light field imaging

The term “dynamic range” is utilized across various domains, with its specific definition depending on the particular context. In the context of displays, dynamic range signifies the ratio between the highest and lowest levels of luminance emitted by the screen. Similarly, for cameras, dynamic range refers to the ratio between sensor saturation and the noise threshold [88]. Whereas, for images, dynamic range describes the proportion between the brightest and darkest pixels. For more robust calculations, dynamic range can be computed by excluding a fraction of the extreme pixels –both the brightest and the darkest– thereby reducing the influence of outliers [88, 253].

Accordingly, HDR denotes a high ratio value, representing a wider range of colors and brightness compared to LDR and Standard Dynamic Range (SDR) images. HDR images are also called “radiance maps” due to their ability to encompass a wide spectrum of intensities corresponding to the visual characteristics of the depicted scene. Thus, HDR images have higher fidelity and realism compared to LDR and SDR images. Moreover, HDR images capture additional information that may not be visible to the naked eye, making them valuable for various applications, including satellite imagery, physically based rendering, medical visualization, and many others [253, 219, 323, 124, 125, 122].

One of the major aspects to consider when processing and working with digital images is their storage method. This consideration is especially pertinent for HDR images, which possess the capability to encode a broader range of colors and light intensities in contrast to the standard 24-bit image formats. Section 5.1 provides a detailed discussion regarding the existing HDR formats.

HDR imaging can be classified into two main categories: single-camera and multi-camera techniques. For single-camera techniques, capturing is done by means of one camera as the name implies. Sequential capturing is carried out for the exposure stack, which is the main disadvantage of the single-camera technique. This drawback is significant for time-critical tasks. Following the capture process, the inverse camera response function is then reconstructed in order to approximately evaluate the radiance mapping to the scene. Finally, a tone mapper is applied. On the other hand, the exposure stack is captured simultaneously, using the multi-camera technique. Although this technique is efficient in terms of capture time, there still remains the issue of stereo view correspondence since the images resulting from the different cameras have different luminance values [323, 124].

For both methods, however, multiple LDR images are required, from which the HDR image is recovered. Accordingly, for single-camera approaches, a tripod is required to capture the same static scene multiple times with different exposures. Hence, these methods will not work with hand-held cameras. Neither will they work with dynamic scenes, and therefore, ghosting effects may appear in the final HDR image [156, 124].

Since using multiple cameras to capture a single HDR image is more expensive and requires more calibration, therefore, single-camera techniques appear to be better. Moreover, to overcome the drawback concerning the time limitation previously introduced, as well as fixing the camera position, single-shot capturing can be applied instead of taking multiple shots. However, this comes with an adjustment of multiplexing different exposure patterns on the sensor. In order to reconstruct the HDR image from the LDR single-shot images, many methods have been devised, including reverse tone mapping, computational photography and CNNs. Among those approaches, CNNs have proven to be the best [124].

In addition to HDR image reconstruction, HDR video reconstruction has also emerged. Since videos are composed of multiple frames, challenges arise in their process of HDR reconstruction. However, due to the temporal coherence between the consecutive frames, common information can be used for reconstructing HDR videos to boost the results compared to single HDR images. In Section 5.2, we address the different techniques used to reconstruct HDR images from LDR images, as well as HDR video reconstruction [228, 124].

While the HVS encompasses a dynamic range of approximately 120 dB [32], it typically operates within four orders of magnitude concurrently, with the potential to adapt to an additional six orders of magnitude, both upwards and downwards. This adaptability is crucial for perceiving a wide variety of lighting conditions. Currently, there is a concerted effort to enhance the capabilities of capture and display devices to accommodate the dynamic range of HDR images, thereby adding realism to visualizations that closely align with the capabilities of the HVS.

In this regard, LF technology plays a pivotal role, offering the ability to capture and reproduce scenes with enhanced depth and realism, aligning with our understanding of human visual perception. Combining both LF and HDR technologies is rather powerful, where 3D content is visualized with an added sense of realism, close to the HVS. This, however, shall be rather challenging, as the limitations of both HDR and LF imaging need to be taken into account. Section 5.3 elaborates on the various applications in which HDR imaging is employed. Furthermore, a comprehensive analysis of HDR LF imaging is provided, in which we address the future use cases with the highest practical potentials.

The investigated types of LF utilization include physically-based rendering, digital photography, image editing, cinematography, healthcare, cultural heritage, education, digital signage, and telepresence [122].

Given the importance of potential HDR LF applications, it stands to reason that the techniques used for LDR to HDR reconstruction could also be effectively applied to LF images. However, reconstructing HDR LF content from LDR LF images can be challenging, and at the same time, generates better outputs since the information of the scene is encoded in multiple images [125]. In Section 5.4, we test some of the existing HDR CNNs (*ExpandNet* [220], *HDR-DeepCNN* [94], and *DeepHDRVideo* [57]) on the *Teddy* LF image dataset [127] and evaluate their performance using Peak-Signal-to-Noise-Ratio (PSNR), Structural Similarity Index Measure (SSIM) and HDR-Visible Difference Predictor (VDP) 2.2.1. Our work addresses both image and video reconstruction techniques in the context of LF imaging. The results indicate that further modifications to the state-of-the-art reconstruction techniques are required to efficiently accommodate the spatial coherence in LF images [124].

LF images provide vast amounts of visual information by capturing light traveling in all directions for every point in 3D space, making them invaluable for detailed scene representation. As previously noted, CNNs have been tested on LF images, highlighting the pressing need for more datasets to support both the training and testing of these models, further emphasizing the requirement for a greater number of HDR LF image datasets. While various capture methods for LF images exist (as discussed in Section 2.4), these baseline-specific setups can be extremely expensive and often require substantial computing resources for accurate calibration. Furthermore, the resulting LF is commonly limited with regard to angular resolution. A suitable alternative to produce an LF dataset is to do it synthetically by rendering LF images, which may easily overcome the aforementioned issues. In Section 5.5, we discuss our work on creating the "CLASSROOM" HDR LF image dataset, depicting a classroom scene. The content is rendered in HOP and FP as well. The scene contains a high variety of light distribution, particularly involving under-exposed and over-exposed regions, which are essential to HDR image applications [124, 116].

To summarize, the chapter opens with an introduction to various HDR image formats and encoding techniques in Section 5.1, followed by a comprehensive overview of LDR to HDR reconstruction methods in Section 5.2. Section 5.3 delves into the potential applications of HDR LF images, providing an in-depth exploration. In Section 5.4, the previously discussed LDR to HDR reconstruction methods are evaluated on LF images, underscoring the necessity for additional HDR LF datasets. Consequently, our approach to generating synthetic LF datasets is detailed in Section 5.5. The chapter concludes in Section 5.6, where future directions and areas for further research are outlined.

5.1 HDR image formats and encoding

As a means of storing HDR images, HDR image formats have emerged, recording wider color gamuts compared to RGB images. These formats take into consideration several aspects, including file size, total dynamic range and the size of the smallest step between the consecutive values. Among the different HDR image formats are HDR, Tagged Image File Format (TIFF) and EXtended Range format (EXR). The HDR format (*.hdr* and *.pic*) was first introduced in 1989, covering more than 76 orders of magnitude, with files as large as uncompressed 24-bit RGB images, since the used run-length encoding achieves 25% compression rate. Compared to HDR encoding, the TIFF float format takes almost three times the storage space, since floating numbers are not well compressed. On the other hand, it is best suited for writing and reading float-point frame buffers. Since users always favor compressed files for easier usage and storage, the *LogLuv* encoding was introduced for a more compact TIFF representation [253]. Later in 2002, EXR was introduced as an open source C++ library used for reading and writing EXR images. In EXR, both the 16-bit and 32-bit floating point numbers are used for storing pixel data. The EXR format supports mipmapping, tiling, as well as lossless compression. For compression, either ZIP deflate library or Industrial Light and Magics (ILM) are used, with the latter being more compression efficient, resulting in a 60% compression. Moreover, EXR supports random channels such as user-defined ones, alpha, depth, etc. [187, 155, 138, 253, 116].

5.2 Overview of LDR to HDR reconstruction methods

5.2.1 HDR image reconstruction

Much work has been carried out concerning HDR image reconstruction via CNNs. In this section, we discuss some of those works. For single-shot single-camera approaches, multiplexing the exposure for the sensor is carried out in order to hold more information about the HDR image. However, this is not always the case. The work of Eilertsen *et al.* [94] provides a method to reconstruct HDR images from single-exposure LDRs. However, the approach is to recover information concerning the saturated pixels, and not those pixels in the lower part of the dynamic range. The idea is based on a CNN acting as an autoencoder with a hybrid dynamic range. The primary concept is to convert the input LDR image by means of an LDR encoder into a set of spatial feature representations to be used later by the HDR decoder in log domain, resulting a recovered HDR image. In addition to the encoder-decoder pipeline and the aim to efficiently exploit the high-resolution details of the HDR image, skip connections are available all the way between both the LDR encoder and the HDR decoder.

Other HDR reconstruction techniques are based on reversing the camera pipeline used to create LDR images [211]. The camera pipeline for LDR image formation is composed of three main steps:

1. Clipping of the dynamic range: HDR image values are clipped, causing information loss in over-exposed (i.e., bright) areas.

2. Non-linear mapping: Adjusts contrast using the Camera Response Function (CRF) to align with HVS.

3. Quantization: Limits pixel values to 8-bit RGB, leading to information loss in under-exposed (i.e., dark) and gradient-smooth regions.

The HDR image reconstruction pipeline reverses the LDR formation process in three steps. Each step is handled by a specifically trained CNN network, aimed at restoring HDR details from the corresponding LDR processing stages:

1. Dequantization: This step eliminates contouring artifacts and noise in smooth regions caused by quantization in the LDR image generation pipeline. The dequantization CNN is trained such that the loss is minimized between the dequantized image ($\hat{I}_{deq}$) and its respective ground truth image (I_n). The loss –required to be minimized– can is defined as follows: $L_{deq} = \|\hat{I}_{deq} - I_n\|_2^2$.

2. Linearization: Here, the goal is to convert the non-linear LDR image into linear radiance using the CRF. The CNN is designed with constraints that ensure a monotonically increasing function and correct mapping of the output's minimum and maximum values.

3. Hallucination: This step recovers lost information from over-exposed regions due to dynamic range clipping. The CNN minimizes the log loss: $L_{hal} = \|log(\hat{H}) - log(H)\|_2^2$ where $\hat{H}$ is the output from the hallucination step and H is the HDR ground truth image. This log domain measurement is crucial for reducing errors in high-value regions.

Figure 5.1 shows both the LDR and the HDR pipelines with elaboration on each step and their corresponding ones in the other pipeline.

Another suggested method for HDR image reconstruction is the *ExpandNet* CNN [220]. This CNN takes an LDR image and propagates it through three branches simultaneously: (i) local branch, (ii) dilation branch and (iii) global branch. Each of the branches handles a respective level of detail (low, medium and high details, respectively). For the first two branches, the LDR image is passed without any sampling –unlike the global branch, where the image is down-sampled. Finally, the outputs from all branches are convoluted, resulting an estimated HDR image [124].

5.2.2 HDR video reconstruction

Unlike the previous methods, where the HDR image reconstruction approaches are applied for single images, the works of Kalantari and Ramamoorthi [156, 157] tackle the problem

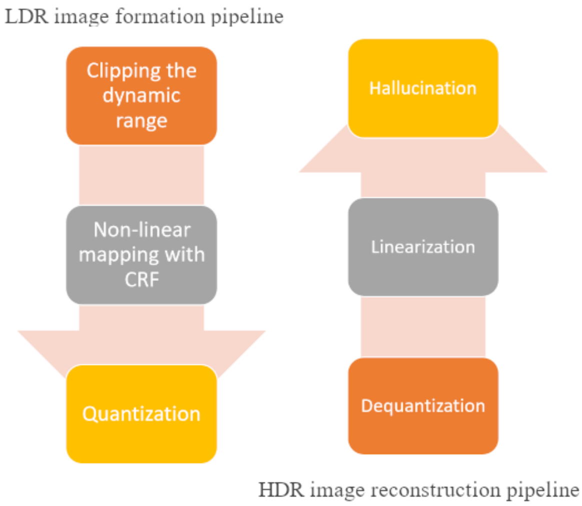


Figure 5.1: Camera pipelines for LDR and HDR image reconstruction [124]

of HDR video reconstruction from multi-exposure frame sequences. Considering videos, HDR image reconstruction is usually carried out in the following two steps:

1. The first step is to align consecutive frames with various exposures to the current frame. Frames need to be temporally coherent. Therefore, reconstructing frame Z_i is done via its neighboring frames Z_{i-1} and Z_{i+1}. The optical flow method proposed by Liu *et al.* [207] is used for optical flow prediction [156]. This method is carried out by aligning the images with extreme exposures (low and high) to that with medium exposure. A later work proposes a CNN to estimate the optical flow in order to minimize the resulting error between the estimated HDR image and the ground truth image [157].

2. The second step is to fuse the aligned frames for HDR image generation. The proposed CNN [156, 157] is utilized to estimate the fusion weights used in the merging process, hence improving the quality of the resulting images.

Although this HDR video reconstruction method [156, 157] achieves success, ghosting artefacts arise. This is because of the noise and the missing information in the under- and over-exposed regions, respectively. Accordingly, accurate image alignment and fusion is not feasible, leading to ghosting [57]. In order to overcome the aforementioned issue, a coarse-to-fine CNN was proposed for a more accurate image alignment and HDR fusion. The proposed algorithm [57] consists of two main steps:

1. The first step is to align and blend images. This is done via *CoarseNet* CNN. This CNN has the same structure as the CNN of Kalantari and Ramamoorthi [157], as

it estimates the optical flow (using the *flow network*) and the blending weights (using the *weight network*). As the name implies, this CNN results in coarse HDR reconstruction, since it uses a smaller number of feature channels compared to the CNN of Kalantari and Ramamoorthi [157]. Calculation of the loss function used in network training is done via the computation of the tonemapping loss in HDR space, using the μ-law function:

$$T_i^c = \frac{log(1 + \mu H_i^c)}{log(1 + \mu)} \tag{5.1}$$

where μ is the parameter used in controlling the level of compression and it is set to 5000. T_i^c is the HDR image resulting from the tonemapping process. Accordingly, the loss in the *CoarseNet* is calculated against the ground truth HDR image ($\tilde{T}_i$) as $\|T_i^c - \tilde{T}_i\|_1$. This CNN succeeds at recovering some of the missing information in the over-exposed regions, as well as removing some noise from the under-exposed regions.

2. The second step is the alignment and fusion in feature space. This is done via *RefineNet* CNN, which is applied in feature space while performing frame alignment and fusion. *RefineNet* starts by taking as input three coarse HDR images, denoted as $H_{i-1}^c, H_i^c, H_{i+1}^c$ and producing the corresponding 64-channel feature outputs $F_{i-1}^c, F_i^c, F_{i+1}^c$. Deformable convolution [78] is then applied to perform feature alignment, resulting in $F_{i-1}^{\tilde{c}}, \tilde{F_i^c}, F_{i+1}^{\tilde{c}}$. These features are then convoluted into the center frame.

Finally, at the end of the pipeline, the reconstruction branch applies regression to the input-fused feature image, resulting in H_i^r, which is used to compute the final estimated HDR image (H_i) as follows:

$$H_i = M_i \bigodot H_i^c + (1 - M_i) \bigodot H_i^r \tag{5.2}$$

where the element-wise product is denoted by $\bigodot$ and M_i is a mask used to define the well-exposed areas for reference frame i. The following equations show how M_i is defined for low- and high-exposure reference images L_i, respectively [124].

$$M_i\,(low) = \begin{cases} 1, & \text{if } L_i \geq 0.15 \\ (L_i/0.15)^2, & \text{if } L_i < 0.15 \end{cases} \tag{5.3}$$

$$M_i\,(high) = \begin{cases} 1, & \text{if } L_i \leq 0.9 \\ (\frac{-L_i}{0.1} + 10)^2, & \text{if } L_i > 0.9 \end{cases} \tag{5.4}$$

5.3 Analysis of HDR light field images in practical utilization contexts

HDR technology enhances the realism of visual content, while multi-autostereoscopic systems, such as LFDs, offer immersive 3D experiences without the need for additional viewing gears. The integration of HDR and LF imaging has the potential to yield powerful and engaging results across various applications; however, this combination presents challenges due to the inherent limitations of both technologies. This section analyzes the applications of HDR LF imaging and explores future use cases with significant practical potential. We will begin each use case by outlining its specific application in HDR imaging, followed by a discussion of the potential integration of HDR and LF imaging [122].

5.3.1 Physically-based rendering

Not only do we need to store the absolute radiometric values by the lighting and physically-based rendering programs, but also other quantities that are not visible to the human eye, as they could be used in further processing, reduction of accumulated errors, alpha and depth channels [253]. Since HDR adds more realism and is closer to the HVS, it is necessarily used in Image-Based Lighting (IBL), especially for scenes with daylight. IBL describes the process of using real-world light images to illuminate real and synthetic scenes [84]. As an application to IBL by means of HDR images, the "Radiance" software was developed by Ward [317] as a physically-based rendering system. It is used in the context of architectural design as a means of predicting light levels and not-yet-built elements. Another attempt was carried out by Larson *et al.* [188], where the authors created an operator that performs tone reproduction while maintaining visibility in HDR scenes.

Considering LF imaging, the same concept of IBL can be applied. In other words, a single HDR real-world light image is used for illuminating the scene. However, unlike the conventional methods, for a single LF scene, multiple images need to be rendered. This evidently takes more time, and thus, can be inefficient for real-time applications. Moreover, the baseline of the system needs to be considered, as wide-baseline systems may consider more than one image to be used in the IBL, since such systems span more space compared to narrow-baseline ones [122].

5.3.2 Digital photography

In order to create digital images with higher color fidelity, HDR digital photography was introduced. Although nowadays HDR cameras are already being introduced and novel ones are being developed, previous attempts for HDR capture included the use of exposure bracketing. The principle of such solution is that multiple LDR images with various exposures are captured and then combined together. The different exposures ensure having some pixels to be properly exposed unlike others. Due to the multi-capture method, static scenes are preferred, hence results are expected to be better when using a tripod or other types of stabilization in comparison to hand-held approaches [92].

Generally speaking, HDR is incorporated in the current professional (video) cameras. In order to create a wide dynamic range, these cameras may either have the most sensitive light sensors or have the ability to combine multiple frames with various exposures. Most of these cameras have high-quality optics in order to capture as much light energy as possible, specifically in the dark areas [64].

Capturing HDR images for LFDs depends on the baseline of the required system. For narrow-baseline systems, multiple attempts have been performed, including the focused plenoptic camera using a lenslet array, which was introduced in 2009 [213]. Later, this camera was upgraded further for rich image capture [109]. In 2016, a two-camera hybrid system was proposed for HDR LF image capture by Wang *et al.* [311]. Similar to the HDR image capture by means of multiple cameras, HDR LF image capture is achieved by using multiple plenoptic cameras [197, 190]. Although these attempts rendered plausible results, they are expensive and custom-designed for specific applications. Moreover, they are specifically targeted for narrow-baseline LF systems. On the other hand, modifying hardware for wide-baseline systems is extremely expensive and infeasible. Accordingly, an alternative is to reconstruct HDR images from the captured LDR LF images by means of CNNs [125, 122].

5.3.3 Image editing

Nowadays, many applications support HDR image editing including –but not limited to– *Photoshop* since the CS2 release, *Photogenics*, *Photomatix*, *Fotor*, *dpBestflow* and *Cinepaint*. Certain image editing operations for LDR cannot be used for HDR, such as the addition and subtraction of pixels. Whereas the operations for LDR and HDR images are the same in terms of algorithm, running under or over range is possible when applying LDR operations to HDR images. Moreover, extreme colors, changes in contrast and white balancing for HDR images is different from that of LDR [253]. An example to the operations used in HDR (32-bpc) to LDR (8- or 16-bpc) image conversion by *Photoshop* includes either automatic operations (histogram equalization and highlight compression) or manual operations (local adaptation, edge glow, tone and detail, color, toning curve, exposure, and gamma adjustment) [247].

Since an LF scene is composed of multiple LF images, editing is more challenging. Depending on the case, editing for one image may need to be carried out for others as well to ensure consistency. In that case, editors for HDR LF imaging should include an option of either editing a single image –which may depend on the angular perspective– or editing all images in correspondence with the edited one at hand. Accordingly, keypoints between images need to be detected to further carry on the changes [122].

5.3.4 Cinematography

In order to use HDR in cinematography, it must be cost-efficiently fit for every digital cinema system. Compared to televisions, digital cinemas have relatively better image characteristics and increased sense of immersion for the audience, since they give film-makers the ability to overcome the artistic limitations imposed by conventional displays. Accordingly, for the upcoming HDR generation of cinemas, they are required to provide a premium experience that exceeds that of HDR televisions. For the current conventional projectors, the vast majority has a contrast ratio of 6000:1. As for the Dolby Cinema, the same value is described as 1,000,000:1. For laser technologies, a wider gamut of colors is used [194]. Another attempt for using HDR in cinemas was performed by the *EclairColor* project in 2017, which aimed at deploying HDR for theaters around the world [194]. The project reported a contrast ratio of 8000:1.

For cinemas, projection technology has always been deployed by default. With the technological advancements in cinematography, modern digital projectors operate by emitting a uniform quantity of light onto a Spatial Light Modulator (SLM) to create the output images. These projectors use a subtractive-approach for creating colors. In other words, they block the light on a pixel-by-pixel basis to produce the required colors and shades in the image. Accordingly, by means of the subtractive approach, a pixel can never exceed the luminance of the Full-Screen White (FSW). Moreover, for most contents of the movie images, they have a Frame Average Light Level (FALL) of 10%, which means that almost 90% of the generated light is thrown away without reaching the screen. For HDR systems, the FALL shall include a smaller fragment of the peak luminance, resulting in more waste. In addition to the huge waste, due to the black level being linearly related to the FSW and the inability of SLM to fully block the light, the contrast ratio is almost fixed. This is even worse in the case of HDR systems, as they inherently rely on high-contrast images.

Hence, the main issue in HDR cinematography is the projection image-formation model. The following are the problems arising from this model in HDR, along with their unfeasible solutions [37]:

1. The illumination of a screen with high levels of light is demanding. A possible solution is to reach higher FSW by increasing the power, which requires higher energy consumption. It gets expensive exponentially, as light sources are not linearly scaled. Furthermore, it exerts very high capabilities on heat management, as well as system stability.

2. The dynamic range is the same even when the power is increased due to the light subtractive approach. A possible solution is to increase the contrast ratio capabilities of the SLM, as well as the black levels of the projector by using better optical architectures. First of all, it leads to the reduction of power efficiency. Moreover, due to the projector screen being white or silver, light pollution from anywhere in the auditorium is added to the projected image, which, in turn, increases the effective black level of the projection system. Additionally, this solution also adds more expenses, and thus, it is economically unfeasible.

Ballestad *et al.* [37] proposed the idea of a light-steering projector in 2019. The work suggests to steer the light rather than to block it. In other words, the light that was blocked before in the dark regions is steered to increase the illumination of the bright regions instead. The authors displayed prototypes in the 2018 CinemaCon [37] –Advanced Imaging Society's HDR Summit and Hollywood studios. Even though light-steering projectors are capable of producing high peak luminances, they are limited by the light source supplies. Moreover, due to the light-steering techniques, these systems are adaptive in the sense that their peak luminance is affected by the image content statistics, where the peak luminance is inversely proportional to the FALL level of image. In addition to being affected by the image content and light source supplies, peak luminance is dependant on the maximum current that powers every diode. This steering technique affects the peak luminance, as well as the deep black levels within the image, depending on its content.

In addition to the problems raised by the projection techniques for HDR, additional challenges arise from the interaction between HDR and HVS, where these interactions are complicated and need further understanding. Among the difficulties concerned with these interactions are the complaints of the cinema artists on the software tools they work with, as they consider these tools to be dependant on basic vision science. Therefore, the film-making industry relies on manual alterations done by artists and technicians instead of automated methods, in order to match the appearance of the shot scene to that of the real world. Cyriac *et al.* [74] introduced the Tone Mapping (TM) and Inverse Tone Mapping (ITM) methods for the different processes in film making, including production, exhibition and post-production. Their algorithms are useful for cinema applications, since they are based on the vision models, where the parameters of the methods are fine-tuned by cinema professionals.

To sum up, it is essential for display manufacturers, as well as content creators to apply standardization defining the maximum characteristics achievable by the graded cinema content. If any display is incapable of inhibiting these standard characteristics, adaptive methods are then used to preserve both the required quality and the artistic goals [37].

Although 3D cinemas have gained popularity with more spectators, the usage of 3D gears degrades the overall experience. Accordingly, automultiscopic displays and LFDs are suitable for cinematography due to their ability to display multiple angular perspectives of the scene, resulting in a 3D sense of immersion without the need of additional viewing gears [93, 168]. Since cinematography addresses multiple spectators at once, wide-baseline systems are the ones used. As an attempt for the usage of wide-baseline LFDs in cinematography, the HoloVizio C80 LFD[1] was implemented. Its huge size of 3 m × 1.8 m allows it for such utilization. Figure 5.2 shows the setup for the HoloVizio display composed of multiple projectors, where each of them creates the content visualized from a certain angular perspective.

Applying the concept of light-steering projectors to LFDs shall allow the spectators the experience of visualizing HDR contents in a cinema without the cumbersome nature of 3D glasses. In order to achieve such task, the light-steering concept has to be applied to each of the projectors used for the LFD [122].

[1] https://holografika.com/c80-glasses-free-3d-cinema/

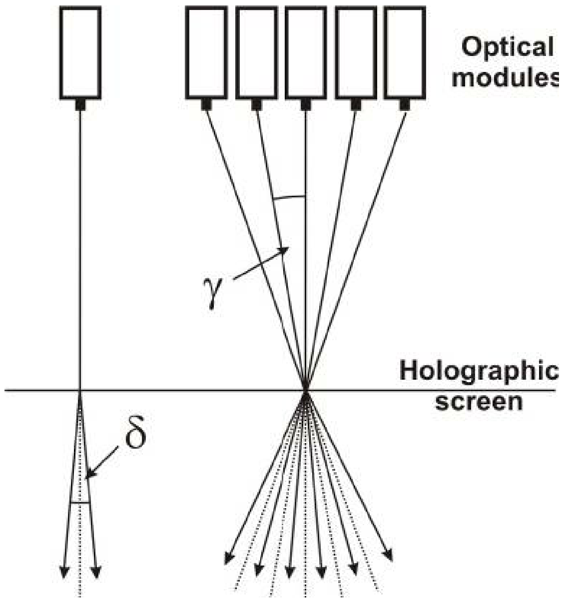


Figure 5.2: The HoloVizio setup [41]

5.3.5 Medical use cases

One of the main concerns in medical imaging is the set of possible limitations regarding the capabilities of the utilized display systems. An appropriate display allows for a reasonable trade-off between the diagnostic accuracy (i.e., the avoidance of false negatives and positives) and productivity (i.e., short interpretation times). Based on the research conducted by Reiner *et al.* [252], the interpretation error rates for radiology include a range of 2% to 15% false positive readings, while false negatives occur more frequent in the range of 20% to 30%.

Since HDR allows more contrast levels and dynamic range compared to LDR, it is thus more efficient for diagnostic procedures, where medical images are supposed to convey information with the highest possible accuracy, in order to facilitate the disease detection task for clinicians. Medical imaging is affected by many factors, among which is image accuracy, bit depth, spatial resolution, dynamic range, viewing angle, arising artefacts (e.g., noise), and perceptual issues (e.g., contrast sensitivity and visual acuity). These factors need to be taken into account when designing displays for medical purposes. Whereas conventional displays have been effective in medical tasks, some medical applications that have fine details require displays with better dynamic range and higher luminance values. According to Ramponi *et al.* [251], in order to display high-quality diagnostic images, at least three requirements must be met: (i) various levels provided by the detector (i.e., bit depth of the obtained datasets), (ii) complex mapping between the source data and the corresponding driving levels and (iii) visualization of source data and the corresponding driving levels as distinct luminance values on the display.

For conventional displays, HDR images undergo dynamic range compression or various techniques of TM. However, doing this for medical imaging is not recommended since photometric distortions can occur to the processed data, rendering inaccurate information. Some attempts were done to display HDR data on conventional displays for medical purposes including the *window-and-level* method. On the other hand, long analysis time and the possibility of details distortion and/or loss occurs in the search phase of this method [328]. Another attempt took advantage of the eye-tracking techniques, where dynamic processing is being carried out on the display, such that the inspected area had its luminance and contrast optimized [61]. Currently, HDR displays have become available in the market among which is the HDR Liquid Crystal Display (LCD)s with 14 bits. These are used by radiologists and physicians, where medical image details are more subtle. While increasing the dynamic and luminance range, some of the image quality parameters are affected. Among which are the increased veiling glare, visual adaptation (done by retina) and optical crosstalk. Hence, more research needs to be considered for the mapping between the obtained datasets and the final visualizations. Due to the nonlinear behavior of HVS, nonlinear mapping needs to be carried out, taking into account the HVS when designing the map while adapting to the display's luminance range [29].

3D imaging has proven its efficiency in many aspects of the medical field, including the diagnosis process, where a better understanding of the complex spatial structures is achieved, as well as better abnormality detection. An example to that is the increased detection rate achieved by stereoscopic devices in breast imaging. Moreover, 3D imaging has proven its importance to the manufacturing of the medical devices and treatments, as well as a better visualization of 3D ultrasound, leading to an increase in the visualization quality of the internal structures. Among the different important use cases of 3D imaging in the medical field is the Minimally Invasive Surgery (MIS), resulting in a decrease in the surgical time while improving the surgical procedure accuracy [302, 214]. Also, such displays have enabled the 3D visualization of the results of Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans, which could further enhance the neurosurgical applications [236]. Displaying HDR medical content on LFDs can be a breakthrough in medical imaging, resulting in an increased accuracy in diagnosis, detection and surgeries, while also resulting in a significant increase in the success rate of the different medical applications [122].

5.3.6 Cultural heritage

Cultural heritage –including archaeological sites– can be better visualized by means of computer graphics using digital surrogates. A digital surrogate allows for better insight and historical understanding by virtually representing real-world elements. Since visual appearance is a crucial factor for creating digital surrogates, light and illumination are key elements for better outputs. Accordingly, the more the light range is covered, the higher the rendering fidelity is, and thus, the more accurate the historical representation is. Unlike LDR, HDR imaging allows for a richer light acquisition in the real world, which can be later used in IBL for studying cultural heritage, thus, increasing the realism of the reproduced outputs. In addition to light acquisition, HDR provides visual documentation with more details when being zoomed in by reducing the clipping of the subsurface information [133, 134].

As a means of visualizing the digital surrogates, LFDs provide an excellent choice, as they allow spectators and researchers to navigate through the 3D scene. In addition to visualizing the contents of cultural heritage, LFDs are best deployed in museums and galleries for viewing exhibitions. Since visitors are highly mobile and can walk freely in museums, using LFDs present such contents better, since they provide the correct angular perspectives from any viewing position within the VVA [159]. Accordingly, visualizing HDR contents on LFDs shall help in studying and researching cultural heritage with easy navigation and manipulation through the 3D scene. Moreover, deploying HDR LFDs in exhibitions shall increase the engagement of visitors due to the combination of vivid colors and general 3D immersion. This is particularly applicable to wide-baseline LFDs [122].

5.3.7 Education

Among the various fields in which HDR can be used as an educational tool is architectural education. Based on the work of Debevec [83], rendering synthetic objects in real-world scenes has become possible, hence, bridging the gap between the physical sites and digital designs. Accordingly, IBL has enabled HDR to become a useful, practical tool for architecture education, where the technical aspects of lights are taught [278]. This allows for a complete design education –according to Watson [319]– where the light is taught in connection to the site and the environment, allowing easier experimentation.

LF technology, with its capability to simulate 3D scenes, serves significant educational and training purposes. This technology enables learners to comprehend the internal structures of complex devices, such as gearboxes and engines, as well as the human body [330]. Consequently, it is particularly well-suited for advanced educational levels and specialized training programs [159]. It is pertinent to highlight, based on our analysis of LF visualization for training and education, that as the level of education increases, the associated KPI requirements become increasingly stringent [121]. Accordingly, visualizing HDR LF content can be highly advantageous in the field of education, as it shall enhance the learning curve due to engaging the students, while creating content with clear details and colors. Both narrow- and wide-baseline LFDs can be used for different educational purposes. Whereas the wide-baseline systems offer visualization to multiple students at once, narrow-baseline systems can be used in single-user scenarios for a more focused and personalized learning experience [122].

5.3.8 Digital signage

With the fast-moving technological developments, the customers of the current era have become sophisticated media users with growing expectations towards visualization quality.

Accordingly, HDR digital signage is becoming a necessity, where high-quality engaging content incorporating a wide range of vivid colors while having a natural effect is expected. Accordingly, LED displays visualizing HDR contents have now been developed and used in the market, where they provide the best viewing experience, especially when compared to LCDs [51, 16].

Among the different attempts for HDR digital signage is the Samsung QMR series[2], which uses Ultra-High Definition (UHD) up-scaling technology for visualizing LDR contents, creating "life-like" images. In addition to the up-scaling, the QMR/QMT series enables good visual experience from all angles by means of non-glare panels, as well as performing noise reduction while using dynamic color crystals for creating a high dynamic range of colors (almost one billion color shades). For easier usage and mount, this series features slim design. Other attempts included Sony with its new BRAVIA 4K HDR professional displays[3] that could be used to present HDR contents in the field of digital signage, and Vestal[4] and LG[5] digital signage with their different series for visualizing HDR content.

Introducing LFDs for digital signage while creating HDR contents shall greatly improve the market due to their ability to grab the users' attention with their vivid colors, as well as the glasses-free 3D visualization. The added 3D effect shall allow spectators to view the advertised contents from multiple perspectives, adding realism and increasing the plausible outcomes from the utilization of digital signage, while engaging the users and consumers. For digital signage, it is preferable to use wide-baseline LFDs to target numerous consumers at once [122].

5.3.9 Telepresence

In recent times and especially since the emergence of the COVID-19 pandemic, online meetings and video conferences have become more of a necessity. This emphasizes the importance of creating high-quality sounds and visuals for calls. As a solution to this issue, 3D telepresence systems are being developed for real-time audiovisual connection. In addition to being used in communication, telepresence can be used to immerse users in remote sites with high degrees of realism, which can be further used in VR [241]. Since outdoor scenes have a high dynamic range of luminance, where the sun is almost 2^{17} times brighter compared to the dark areas in clouds [281], representing those scenes in telepresence by means of LDR imaging results in poor quality outputs. Hence, HDR imaging is deployed [241] by means of exposure bracketing, where multiple LDR images of the scene are captured with different exposures and merged together to create HDR contents [85].

[2] https://displaysolutions.samsung.com/pdf/brochure/5257/Smart_Signage_QMR_QMT_Brochure_210909_WEB.pdf

[3] https://cdn.cnetcontent.com/5f/8a/5f8a2146-0a42-477b-9d4d-dea498205aea.pdf

[4] www.vestelvisualsolutions.com

[5] https://www.lgbusiness.it/wp-content/uploads/2020/07/Catalogo-LG-Signage.pdf

Utilizing LFDs in remote meetings can greatly enhance the visual experience, where LFDs with the same size of the individuals can be used to add realism and create a higher sense of presence, as well as engaging environments [330]. As an attempt for telepresencing by means of LFDs, the HoloVizio 1080T [68] provided by Holografika –having the dimensions of 180 cm × 100 cm and an FOV of 180°– creates a full-size portrait. Moreover, the LightBee [336] implementation addresses the same telepresence problem by displaying only the head of the user. In both telepresence utilization cases, using HDR along with LFDs can create more engaging experiences in the audiovisual calls, as well as immersing the users in remote sites with a higher sense of realism and 3D immersion, close to real-life experiences [122].

5.4 Applying LDR to HDR reconstruction techniques on light field images

Whereas the different HDR reconstruction CNNs were applied for either conventional images or videos, applying the same CNNs for LF images still remains an open question. In this section, we introduce the experimental setup of the tests that addressed some of the implemented HDR image and video reconstruction CNNs on LF images.

The CNNs were evaluated using the *Teddy* dataset [127], captured by Fraunhofer, which comprises high-quality LF images in HDR, acquired through exposure bracketing techniques. This dataset features multiple static LF images with large spatial resolution, making it an ideal benchmark for assessing the performance of HDR reconstruction methods [124].

5.4.1 HDR reconstruction for light field images

From Section 5.2, three CNN architectures are used for experimenting with HDR image generation: *ExpandNet* [220], *HDR-DeepCNN* [94], and *DeepHDRVideo* [57]. We have considered real-world HDR LF dataset *Teddy* [127] that contains geometry and color-calibrated HDR LF images. This dataset consists of 50×50 LF images with horizontal and vertical parallax. From the original 50×50 LF images, we generated 36 non-overlapping subsets of LFs, each containing 8×8 images. For each algorithm, the average performance results are reported over all the 36 LF sets.

For testing the HDR image reconstruction algorithms (*ExpandNet* and *HDR-DeepCNN*), we simulated constant-exposure LDR images from the *Teddy* HDR images and fed each network one image at a time. The performance is measured for one image at a time and then averaged over a given LF subset. For testing the *DeepHDRVideo* method, we considered three alternating exposure versions of this trained algorithm. Precisely, given an LF subset, we extracted three consecutive HDR images at a time and generated three LDR images with varying exposures. The overview of the procedure is provided in Section 5.2, in accordance with the work of Chen *et al.* [57]. Then, these multiple-exposure LF images are fed to this network, three images at a time, for reconstructing

HDR LFs. After generating the corresponding HDR LF subset –similarly to HDR image reconstruction methods– we measure the performance for one image at a time and then average it over a given LF subset [124].

5.4.2 Metrics used for evaluation

In order to test the efficiency of the produced results, quantitative analysis is carried out, where the generated HDR LF images are compared against the ground truth images. In this subsection, we discuss the metrics used in the comparison process including [124]:

- PSNR: Calculated between two gray-scale images f and g, given that their size is $N \times M$ as follows [141]:

$$PSNR(f,g) = 10log_{10}(255^2/MSE(f,g)) \tag{5.5}$$

  where

$$MSE(f,g) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(f_{ij}-g_{ij})^2 \tag{5.6}$$

- SSIM:

$$SSIM(f,g) = l(f,g)c(f,g)s(f,g) \tag{5.7}$$

  where

$$\begin{aligned} l(f,g) &= \frac{2\mu_f\mu_g + C_1}{\mu_f^2+\mu_g^2+C_1}, \\ c(f,g) &= \frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2+\sigma_g^2+C_2}, \\ s(f,g) &= \frac{\sigma_{fg} + C_3}{\sigma_f\sigma_g+C_3} \end{aligned} \tag{5.8}$$

  where l, c and s are the luminance, contrast and structure comparison functions, respectively. The terms μ_f and μ_g are the luminance means for images f and g. The standard deviation is denoted by σ, whereas σ_{fg} is the correlation coefficient between both images (f and g). Finally, C_1, C_2 and C_3 are positive constants used to ensure that the denominator is not null.

- HDR-VDP-2.2.1: This metric is an upgrade for the HDR-VDP metric, which in turn is a modification to the VDP metric. It takes two images as input: the original and the distorted one. Both input images are transformed to their luminance values used in the comparison process. The metric results in a probability map, indicating the differences between the input images [217]. The HDR-VDP-2 [218] enhances the visibility metric, specially for low conditions with regards to luminance.

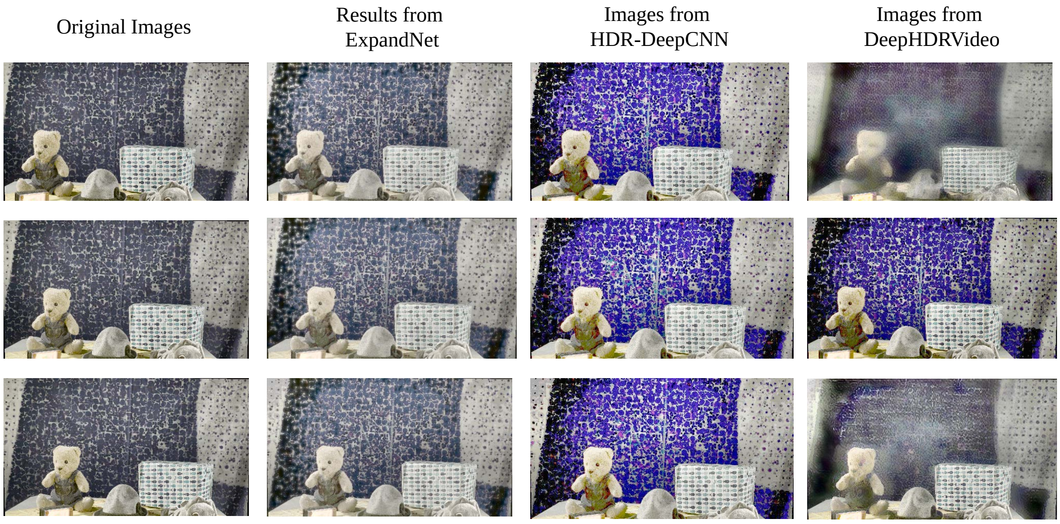


Figure 5.3: Original and predicted tonemapped HDR images from the considered algorithms for experiment [124]

5.4.3 Results

Figure 5.3 shows the results of our experiments on the considered HDR LF dataset *Teddy* [127]. According to Figure 5.3, *ExpandNet* produces plausible results close to the ground truth images. However, visually ghosting artifacts arise in the background on the images. Hence, learning concatenated features from different branches –local, dilation and global– seems to be a good direction for generalizing to other datasets. On the other hand, in the *HDR-DeepCNN*, we see inconsistencies in the colours in the reconstructed textures. These inconsistencies could be the result of using skip connections that include domain transformation from LDR display values to logarithmic HDR. However, the method should be thoroughly investigated further with more LF datasets. Finally, in the *DeepHDRVideo* reconstruction method, inconsistencies are quite visible in the shape and texture of the reconstructed images. The method uses sub network that aligns the input images to a common frame and then does HDR reconstruction using the aligned features. The artifacts seen in the reconstructed images show that there are errors in the optical flow reconstruction due to complex patterns in the scene which are carried forward to the next stage where HDR reconstruction is done. Without a reliable method for calculating accurate optical flow, such a direction for HDR reconstruction does not seem valid for LF images. Furthermore, the *Teddy* images exhibit complex background textures –fine-grained patterns with significant high-frequency components. However, this is not accounted for in the training dataset of the *DeepHDRVideo* CNN, as the dataset primarily contains simple or facade-like backgrounds. Moreover, the tested LF

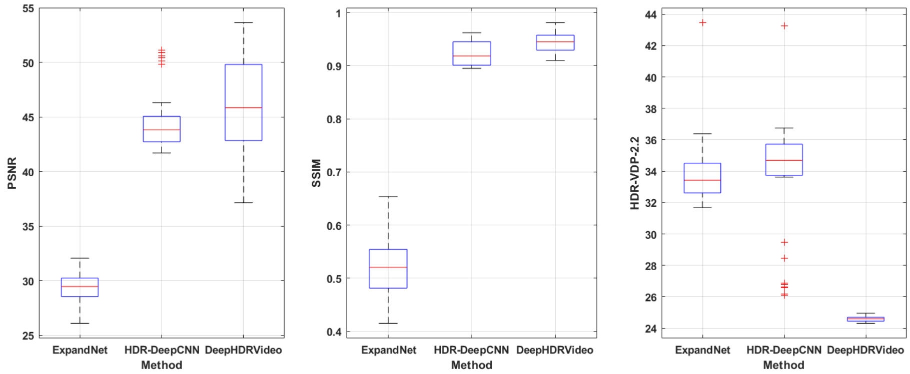


Figure 5.4: Boxplots showing the performance of the considered methods for experiment on the HDR LF dataset *Teddy*. Left: box plot using the PSNR metric. Middle: box plot using the SSIM metric. Right: box plot using the HDR-VDP metric [124].

images exhibit multiple or high disparities, which is not the case in the original training dataset. Consequently, this affects the disparity map estimation, resulting in inaccurate outputs. Additionally, the training dataset encompasses images captured with a static camera, where a person remains mostly stationary with slight movements. In contrast, LF scenes involve motion of the entire scene due to camera movement, rather than just the main character as in the training dataset.

As mentioned earlier, objective performance of the considered methods is measured using three traditional Image Quality Metrics (IQMs): PSNR, SSIM and a perceptually-guided HDR image quality metric HDR-VDP-2.2.1. Results are interpreted by using box plots (see Figure 5.4), which show the min and max scores together with the inter-quartile range (shown in boxes). A larger inter-quartile range translates to more global inconsistencies in the performance.

Our experimental results show that according to metrics PSNR and SSIM, *DeepHDRVideo* method performs better than the other two. It is important to note that the inter-quartile range in obtained PSNR values for the *DeepHDRVideo* method is higher. Also the difference between observed minimum and maximum PSNR values for this method is higher than other two methods showing that there are greater deviations in the performance of this method. In contrast, the SSIM results show no such fluctuations in the performance of this method. HDR-VDP scores show that *HDR-DeepCNN* performs better.

Assuming that HDR-VDP better correlates to the HVS than the other two metrics, our experiments show that the *HDR-DeepCNN* method achieves more consistent results than the other two methods, and also achieves better visual quality of the reconstructed HDR images. Although, HDR video reconstruction methods involve retrieving more scene information than HDR image reconstruction methods, our results indicate that given a properly color-calibrated set of single-exposure LDR LF images of a scene –such as scene *Teddy*– architectures like *HDR-DeepCNN* are capable of reconstructing more globally consistent HDR LFs. This direction is particularly desirable since it only involves capturing single-exposure LDR images, and therefore, also supports faster processing times for HDR reconstruction than methods involving multiple exposures. To further develop such architectures, there is a demand for producing more LF HDR datasets for fine tuning, which is currently lacking in the scientific literature. Furthermore, irregularities in the reported quality scores of various metrics show that there is a great need for novel quality metrics that are more suitable for LFs [124].

5.5 Towards HDR light field datasets: CLASSROOM dataset

According to Metzler *et al.* [228], many techniques can be used to reconstruct HDR images from LDR images, including reverse tone mapping methods, computational photography methods and CNNs. Among those techniques, CNNs have proven to provide the best results with the ability to further improve. Similarly, CNNs can be used for LDR-to-HDR LF image reconstruction. As shown in Section 5.4, different CNNs were tested on HDR LF images to visualize the results in order to proceed with the next steps for better HDR LF image reconstruction. However, one of the main challenges was the lack of HDR LF datasets. Since the outputs of deep learning depend on both the deep complex structures of the networks and the large training datasets [310], acquiring more HDR LF datasets will further improve the research on HDR LF image reconstruction. However, the creation of an HDR LF dataset requires tremendous amounts of storage due to the aforementioned reasons in Sections 2.8 and 5.1.

In our research, we developed a synthetic dataset called “CLASSROOM”, which enables the manipulation of various parameters and adaptation to different conditions, thereby facilitating the creation of additional datasets. Also, increasing the scene complexity is possible by adding more objects or upgrading the scene to have more complex materials. This can be useful in the progressive learning curve of the HDR LF reconstruction field. In addition to the aforementioned reasons, a synthetic dataset is not custom-designed to a certain type of baseline or parallax, as different alterations can be made to render multiple datasets for different baseline and parallax settings.

The reason behind choosing the classroom scene is due to its ability to provide HDR images. The concept of HDR relies on having a big dynamic range of colors in the scene. Considering the classroom scene, this is possible since there are areas where light penetrates the classroom windows, creating over-exposed regions, whereas on the other hand, some regions in the classroom (e.g., cupboards and bookshelves) are under-exposed, hence the high dynamic color range in the produced images [116].

5.5.1 MAYA setup

In order to create the CLASSROOM dataset, we used MAYA (version of 2022). For rendering the modeled classroom, the Arnold renderer was used. This renderer is an advanced Monte Carlo ray tracing renderer, which is both memory-efficient and scalable. Multiple features are integrated in the Arnold renderer, including –but not limited to– subsurface scatter, hair and fur, motion blur, volumes, instances, subdivision and displacement, OSL support, light path expressions, adaptive sampling, toon shader and –most importantly– denoising, which was used as a post-processing step in the dataset to eliminate the noise resulting from the Monte Carlo algorithm. Due to its efficiency and plausible results, the Arnold renderer is integrated in many softwares, such as MAYA, Houdini, Cinema 4D, 3Ds Max and Katana [1]. In addition to the aforementioned capabilities, the Arnold renderer allows the usage of IBL, firstly introduced by Debevec [83] in 2008, which allows synthetic objects to be rendered in real-world scenes. In other words, illumination and lighting can be used from real-world scenes by means of HDR Images (HDRI) to illuminate the modeled synthetic scenes, adding realism to the output content. In order to use IBL for realistically illuminating the classroom scene, an HDRI was imported from the "polyhaven" website[6] (previously named HDRI haven). With the rising importance of HDR imaging, the newer versions of MAYA support HDR formats. In the CLASSROOM dataset, we used OpenEXR 32-bit floating point images [116].

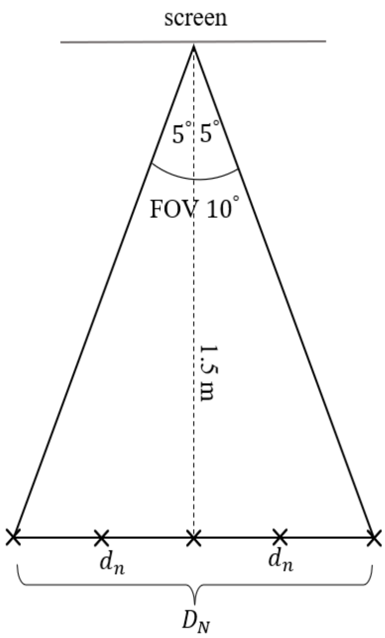
5.5.2 Distance calculation

For the creation of the CLASSROOM dataset, we consider both baselines (i.e., narrow- and wide-baseline systems), the distance between the LFD and the observer line or rectangle and both parallax cases (i.e., HOP and FP). In order to understand the reason behind the chosen distances between the consecutive images in the dataset, we consider the following two cases.

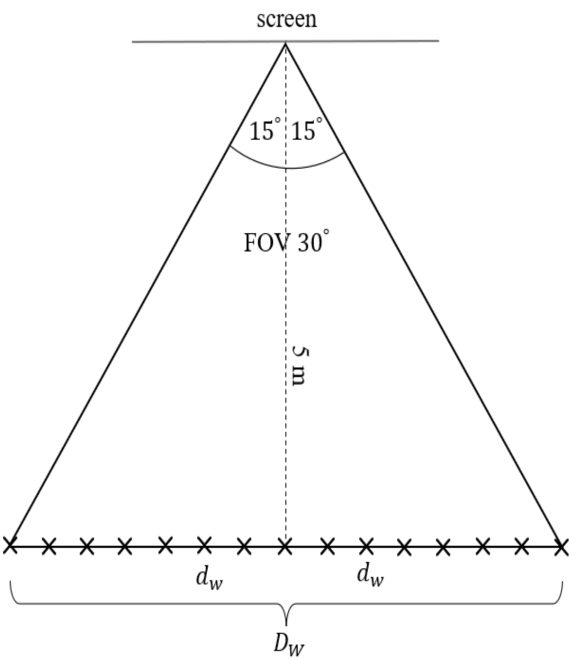
The first case considers a narrow-baseline LF system with an FOV of 10° and a distance of 1.5 m between the observer line or rectangle and the LFD screen. This is illustrated in Figure 5.5a, depicting the top view of the LF system setup. The distance d_n can be calculated as $1.5 * tan(5°) = 0.1312\,m \approx 13\,cm$, with a total distance $D_N = 26\,cm$. In the narrow-baseline FP dataset, we consider 5 images in each direction, hence, the distance between any two consecutive images in the horizontal or vertical directions is $26/4 = 6.5\,cm$.

For the wide-baseline LF system, we consider an FOV of 30°, with a distance of 5 m between the observer line or rectangle and the LFD screen. Accordingly from Figure 5.5b, the distance d_w can be calculated as $5 * tan(15°) = 1.339\,m \approx 133\,cm$, with a total distance $D_W = 266\,cm$. For the wide-baseline HOP dataset, a total of 15 images were rendered, therefore, the distance between each two consecutive images is $266/14 = 19\,cm$ [116].

[6] https://polyhaven.com/



(a) Top view for narrow-baseline LF



(b) Top view for wide-baseline LF

Figure 5.5: Distance calculation in narrow- and wide-baseline LF systems [116]

5.5.3 Rendered results

The CLASSROOM dataset consists of three subsets: (i) narrow-baseline FP, (ii) narrow-baseline HOP and (iii) wide-baseline HOP. The images are rendered using an Intel(R) Core(TM) i7-5820K CPU with 6 cores. For all datasets, we consider an image size of 960×540. The reason for the chosen size is to avoid having small-sized images (i.e., loss of details) and large-sized images (i.e., too much time and complexity when applying HDR LF reconstruction techniques). The creation of the components of the scene (e.g., chairs) followed a public online tutorial on Autodesk Maya[7] [116].

Narrow-baseline FP dataset

Starting off with the narrow-baseline FP dataset, we created 25 images arranged in a 5×5 2D array. The distance between each two consecutive images is 6.5 cm in both the horizontal and vertical directions, covering a total distance of 26 cm spanned in the 10° FOV of the considered narrow-baseline system. The camera used for creating narrow-baseline datasets had a focal length of 35 mm. Figure 5.6 illustrates the distances between the rendered images with respect to the FOV.

[7]Hassaan Owaisi: Classroom interior modeling in maya https://www.youtube.com/watch?v=lRrLqR_5eBM

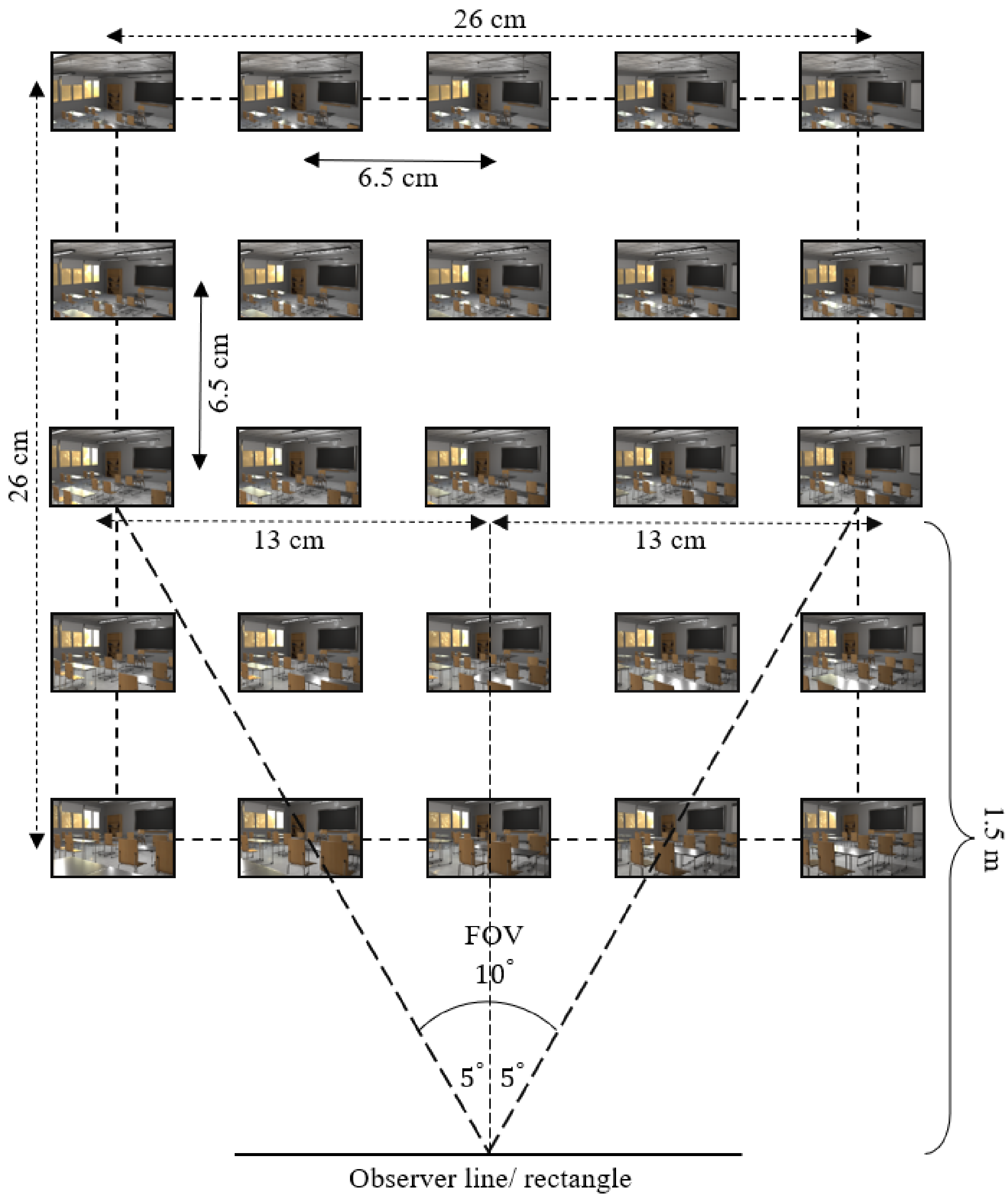


Figure 5.6: Narrow-baseline FP dataset setup [116]

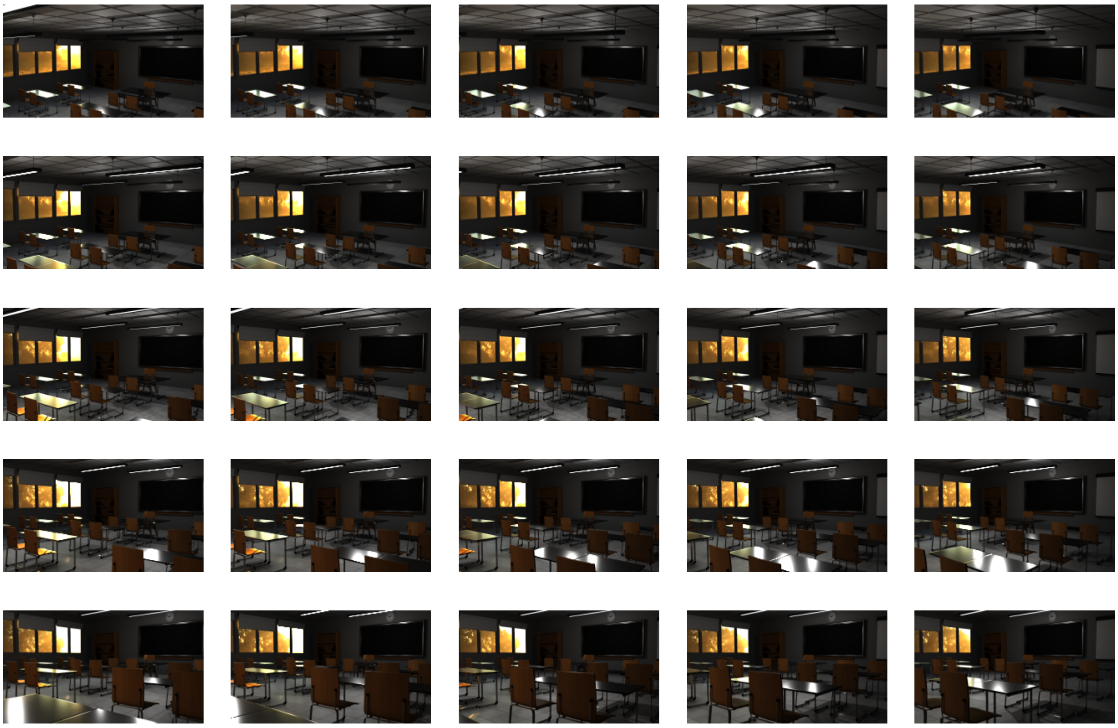


Figure 5.7: Dataset for narrow-baseline FP systems [116]

The final rendered images for the dataset are illustrated in Figure 5.7, with the image EXR file size ranging between 27.1 MB and 30 MB and a total size of 713 MB per dataset. The time taken to render a single image ranged between 6:13 min and 6:56 min with an average of 6:31 min per image [116].

Narrow-baseline HOP dataset

The narrow-baseline HOP dataset is considered to be a subset of the narrow-baseline FP dataset, since the HOP considers horizontal directions only. Accordingly, given the narrow-baseline FP dataset, 5 datasets can be created for the HOP system, as illustrated in Figure 5.8 [116].

Wide-baseline HOP dataset

In this dataset, we consider wide-baseline HOP systems, rendering a total of 15 images for the dataset arranged in a 1D horizontal array. Figure 5.9 depicts the relation between the rendered images and the FOV of the wide-baseline systems, where the distance between any two consecutive images is 19 cm, thus, covering a total distance of 266 cm spanned by the wide-baseline system with an FOV of 30°. For rendering, a camera with a focal length of 20 mm was used to allow for wider motions in the scene.

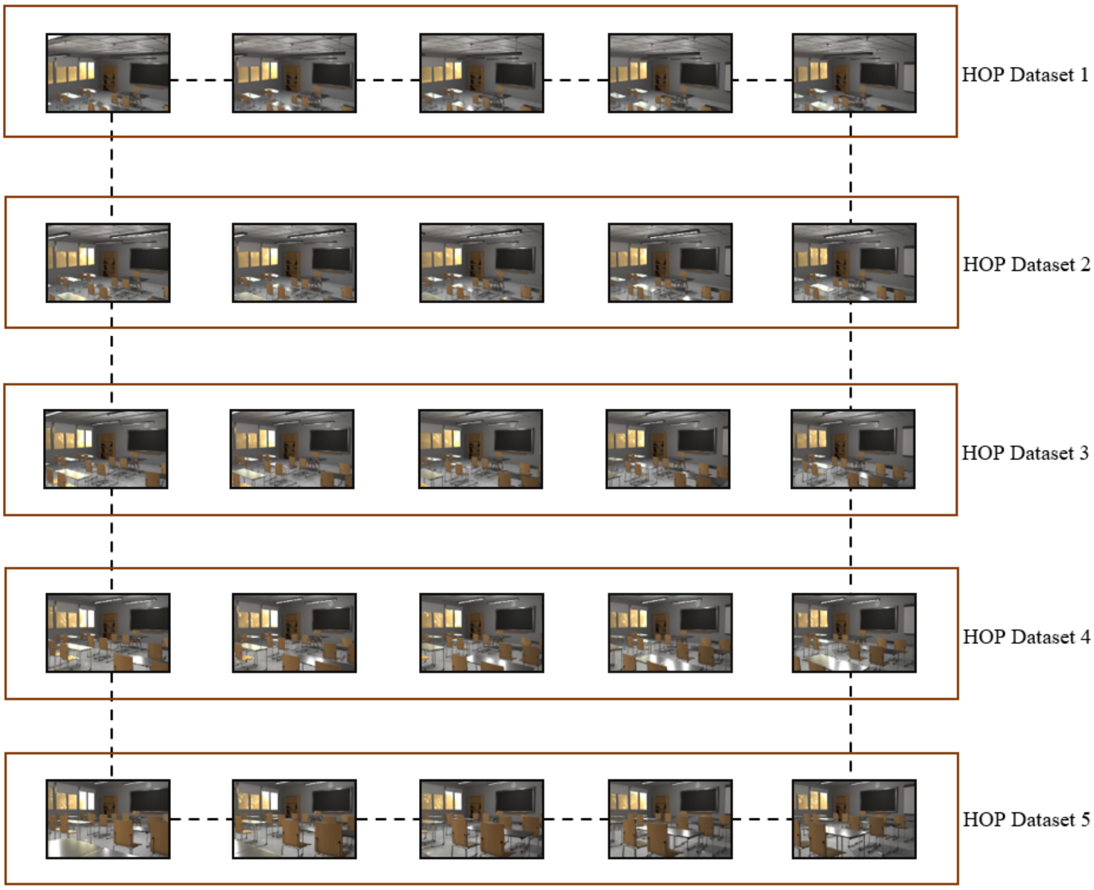


Figure 5.8: Narrow-baseline HOP datasets from the FP dataset [116]

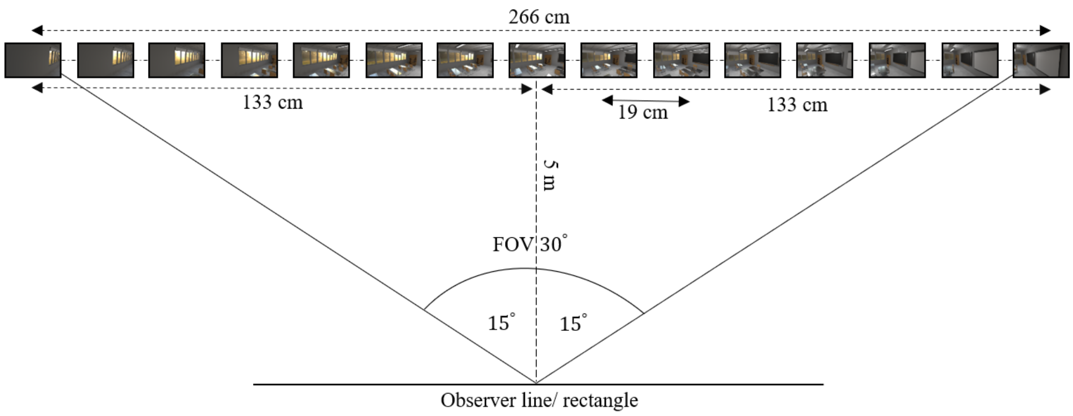


Figure 5.9: Wide baseline dataset setup [116]

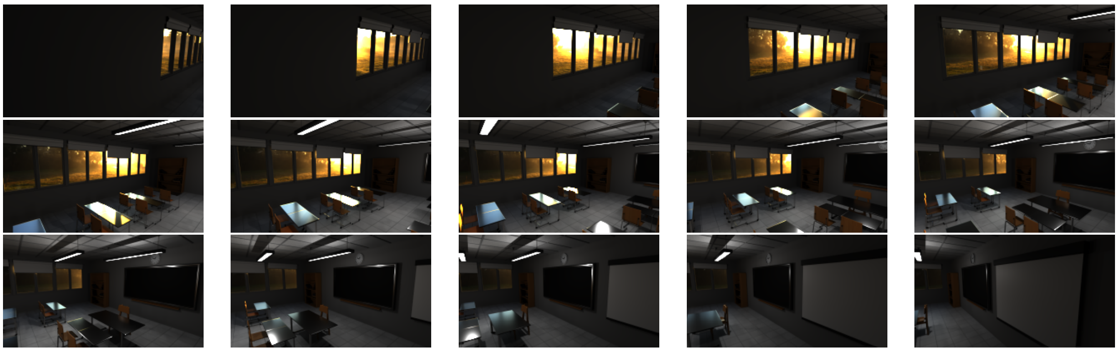


Figure 5.10: Dataset for wide-baseline HOP systems [116]

Figure 5.10 shows the rendered images constituting the wide-baseline HOP dataset, where images are arranged from right to left and top to bottom. The image EXR file size ranges between 19.3 MB and 27.5 MB with a total size of 370 MB for the dataset. The time taken to render an image ranged between 5:18 min and 6:54 min with an average of 6:14 min per image [116].

5.6 Conclusion and future work

In this chapter, we explored the evolution of two impactful imaging technologies –HDR and LF imaging– both of which offer novel visualization capabilities. HDR enhances the color range and detail, providing an experience close to the capabilities of the HVS, while LF imaging allows immersive 3D viewing without additional equipment, increasing user engagement. When combined, HDR LF imaging has significant potential for diverse applications, as discussed throughout this chapter. A critical step in achieving HDR LF imaging from legacy LDR LF is reconstructing the HDR LF images, necessitating varied HDR LF datasets to train and test reconstruction CNNs effectively. To address this, we reviewed various CNN models designed for LDR-to-HDR image and video reconstruction, noting that although considerable progress has been made, applying these models specifically to LF images remains underexplored. Our experiments tested three LDR-to-HDR CNNs –*ExpandNet*, *HDR-DeepCNN*, and *DeepHDRVideo*– on the *Teddy* LF image dataset. Evaluation results, measured against various metrics, indicated that *HDR-DeepCNN* performed best in terms of quality, particularly when assessed with HDR-VDP, though video-based reconstruction methods leveraging temporal coherence were initially anticipated to perform better due to the coherence similarities with LF spatial data. Finally, we presented a newly created dataset for HDR LF applications, designed to support different LF systems with three configurations: narrow-baseline FP, narrow-baseline HOP, and wide-baseline HOP, thereby expanding resources for future HDR LF research and applications [124, 116, 125, 122].

Although we have explored various potential applications of HDR LF imaging, further research is necessary to address each use case individually, as specific requirements and priorities may differ among applications. Additionally, addressing the inherent limitations of both HDR and LF imaging will allow for more tailored and effective design outputs. Since HDR LF applications require high-quality HDR LF images, effective methods for reconstructing HDR LF content from legacy LDR LF images are essential. While single-image-based CNNs provide plausible LDR-to-HDR reconstruction, applying CNNs across multiple LF images yields better results by leveraging spatial coherence and angular information. Testing various existing CNN models on a range of LF images, incorporating key reconstruction principles, and developing CNNs specifically for LF imaging could further improve HDR LF reconstruction, and thus, HDR LF datasets are critical for training and testing CNN models in this area. In our implementation, the classroom scene was intentionally designed with minimal detail, as starting with simpler, lower-complexity scenes allows for more manageable initial testing. As HDR LF reconstruction techniques advance, scene complexity can be progressively enhanced by incorporating additional objects or more sophisticated materials. Additionally, a dataset tailored for arc systems could be developed by rendering images from varied orientations using camera tools within *MAYA*, thereby enriching the dataset's versatility and expanding its application potential. Furthermore, future work includes exploring systems and methods to capture real-world HDR LF content, further supporting the experimental scope and applicability of these datasets [124, 116, 125, 122].

CHAPTER 6

Quality of experience for light field visualization

As research into projection-based LF visualization continues to expand, the investigation of human observer experience presents unique challenges. A primary hurdle in this domain is the absence of standardized testing methodologies, which complicates the design and interpretation of experiments. The introduction of LFDs within research institutions –although limited– has significantly broadened scientific horizons, providing a robust framework for conducting experiments regarding visualization quality and QoE.

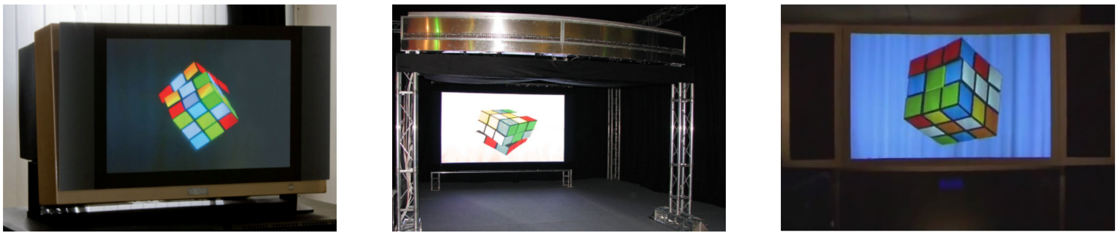
Owing to the 3D perception and immersive experience afforded by LF visualization, numerous potential use cases can be identified, including cinematography, medical applications, 3D digital signage, telepresence, military and industrial applications, among many others. This further emphasizes the importance of guaranteeing that the QoE associated with LF visualization and its applications aligns with or exceeds the expectations of users. This is accomplished through the execution of subjective tests that take into account either a single test variable or multiple variables. Based on the analysis of our work [172], the total number of subjective studies regarding LF visualization published by the first quarter of 2022 is 29. From these works, 20 involved static contents, 4 used LF videos, 1 relied on live video and 2 presented interactive contents [164]. The topics covered in these studies include spatial and angular resolution [161, 72], spatial and angular distortion [292, 287, 288, 25], compression [91], interpolation [67], FOV [165], viewing distance [169], zoom levels [81], ROI [82], viewing conditions [170], LF reconstruction [166, 25], format assessment [69, 72], view synthesis [288], system assessment [336], quality switching [162], content size [183, 182], content characteristics (e.g., complexity [164], alignment [81], and orientation [290]),Human-Computer Interaction (HCI) [19, 20].

To enhance the user experience of LF visualization, we conducted a series of experiments on LFDs to investigate the factors that influence the overall visual experience. The subjective studies conducted in this chapter evaluate various factors, either at a broad level (Sections 6.2 and 6.4) or within the framework of specific use cases (Section 6.3 and 6.5). An alternative method for categorizing the subjective studies presented in this chapter is by considering visual acuity. Experiments involve participants with either normal visual acuity (Sections 6.2 and 6.3) or those with impaired vision (Sections 6.4 and 6.5).

Experimental configurations encompass various elements, among which the viewing distance of test participants holds significance. While conventional 2D displays standardize this parameter extensively based on screen resolution and height (often denoted as H), ongoing research endeavors continually deepen our understanding of how viewing distance influences perceived quality. For example, the study of Amirpour *et al.* [30] evaluates potential bitrate reductions linked to viewing distance, attributing this phenomenon to the more favorable perception of content with lower media encoding quality levels at greater distances. This underscores the need for objective quality metrics that consider viewing distance. In Section 6.2, we delve into our investigation of the perceptually-supported and subjectively-preferred viewing distance in LF visualization.

As previously stated, LFDs can be utilized in various domains, including healthcare, telepresence, educational settings, and numerous other fields. Section 6.3 examines the application of LFDs within industrial contexts. Industrial applications encompass prototype evaluation and diverse modeling applications, wherein the visualized content can exhibit considerable variability in both complexity and dimensions. In the context of LF visualization, complex models are notably influenced by the ray density utilized in visualization. Furthermore, even with high angular resolution, rendering a 3D model at greater depths (i.e., farther from the screen plane) can lead to blurriness. In Section 6.3, we investigate the effects of angular resolution and 3D rendering on the perceived quality of content displayed on LFDs. Our subjective study investigates the industrial application of prototype evaluation through seven static, synthetically rendered industrial objects. The models were designed to exhibit variations in structural complexity and depth, rendered at seven distinct angular resolutions ranging from 0.5 to 2 degrees, yielding a total of 49 test stimuli. Each stimulus was assessed against a reference quality to evaluate degradation levels. The statistical analysis focuses on angular resolutions and source contents, while also tackling methodological challenges related to the study [164].

Furthermore, understanding the effects of LF visualization on users with diverse visual capabilities is crucial. Most subjective studies regarding quality perception in the context of LF visualization feature test participants who have undergone screening for visual acuity using the Snellen chart, in addition to the Ishihara plates for color vision assessment. This measure is undertaken to ensure the reliability of results and to ascertain the accuracy of data that may serve as the basis for ground truth. However, a significant segment of prospective future users wouldn't meet such screening requirements. This is due to a concerning global trend wherein the visual acuity of successive generations is demonstrably deteriorating, accompanied by an increase in the incidence of other ocular impairments, particularly color vision deficiencies. Therefore, it is imperative



(a) HoloVizio 80WLT LFD (b) HoloVizio C80 LFD (c) HoloVizio 640RC LFD

Figure 6.1: HoloVizio LFDs used in the QoE experiments

that long-term innovations in visualization also consider the needs of these users. In this chapter, we assess the LF visualization from the perspective of users with reduced visual capabilities [274], as well as imperfect visual acuity and color blindness [273].

Section 6.4 examines various factors influencing LF visualization, including spatial and angular resolutions, as well as viewing distance. This analysis is conducted twice: first, considering participants with impaired visual acuity, and second, focusing on participants undergoing color-blindness.

Section 6.5 encompasses the experiment involving participants with reduced visual capabilities. In this study, we examine the preferred viewing distances relevant to future applications of LF visualization. Applications in this regard include those involving passive visual engagement, such as cultural heritage exhibition or cinematic viewing experiences.

The structure of this chapter is organized as follows: Section 6.1 presents a comprehensive overview of the experimental framework utilized in the subjective experiments. Sections 6.2, 6.3, 6.4, and 6.5 offer a detailed account of the distinct experiments conducted on the LFDs, along with their methodologies and key findings. Finally, Section 6.6 provides a comprehensive summary of the experimental factors, participant details, and key findings drawn from the experiments on QoE for LF visualization.

6.1 Experimental framework

6.1.1 Light field displays

During the QoE experiments, one or more of three LFDs were utilized. These comprise the HoloVizio 80WLT, the HoloVizio C80 LF cinema and the HoloVizio 640RC (depicted in Figure 6.1). It is important to note that all the LFDs mentioned are of the HOP type.

The HoloVizio 80WLT is a small, television-like back-projection HOP LFD featuring a 30-inch screen. It has an angular resolution of 1 degree and accommodates content spanning a full-horizontal 180-degree FOV. The HoloVizio C80 cinema system is a large-scale HOP front-projection LFD. It operates as a 140-inch front-projection cinema system

Table 6.1: Characteristics of the HoloVizio LFDs

Characteristics	HoloVizio 80WLT	HoloVizio C80	HoloVizio 640RC
Screen size	30 inch	140 inch	72 inch
Angular resolution	1°	0.5°	0.5°
FOV	180°	40°	56°
Projection type	back projection	front projection	back projection

with an angular resolution of 0.5 degrees and an FOV of 40 degrees. The HoloVizio 640RC is a large-scale back-projection HOP LFD featuring a 72-inch screen. It supports an angular resolution up to 0.5 degrees and provides an FOV of 56 degrees. Table 6.1 provides a concise overview of the principal characteristics of each HoloVizio display.

6.1.2 Research environment

All experiments were conducted in a controlled laboratory environment, isolated from audiovisual distractions, with the lighting conditions maintained at 20 lux. The brightness values of the 80WLT, C80, and 640RC were calibrated to 300 cd/m^2, 1500 cd/m^2, and 1000 cd/m^2, respectively.

6.2 The perceptually-supported and the subjectively-preferred viewing distance of projection-based light field displays

As advancements in research and development continue to shape LF visualization technologies, fresh possibilities for new applications emerge. The practical deployment of LFDs across different scenarios crucially hinges upon the observation distance. This section delves into an exploration of the perceptually-supported and subjectively-preferred viewing distances associated with LF visualization. To gain insights into these distances, a comprehensive series of tests were conducted on a variety of projection-based LFDs. These tests engaged both experts and regular test participants in distinct study sessions [163].

6.2.1 Related work

Over the past decade, numerous studies have been conducted to evaluate various KPIs [171], content, and essential attributes in LF visualization through assessments by test participants. Yet, as of 2021, there exists no singular international standard that directly pertains to determining the optimal distance at which LF visualization is best perceived. Due to the lack of standardized methodologies, experimental configurations in LF visualization either choose arbitrary values for the viewing distances without providing reasons for such choices, or do not report the viewing distance in the first place.

In their subjective study, Ahar *et al.* [25] followed the methodology outlined in Rec. ITU-R BT.500-13 guidelines [1], yet they opted for a viewing distance 3.2 times the height of the LFD screen, without providing an explanation for this choice.

The HoloVizio C80 HOP LFD was employed in various studies, including those conducted by Cserkaszky *et al.* [67], Darukumalli *et al.* [82], and Kara *et al.* [162]. In these studies, a viewing distance of 460 cm was established to prevent participants from obstructing the light beams emitted by the display. This precautionary measure was necessary because the C80 utilizes the front-projection technology. Moreover, a study utilizing the C80 LFD established a viewing distance of 6 m without specifying the rationale behind this selection [91].

In their work, Kawakita *et al.* [175] assessed a 200-inch prototype display, whose design calculates the suitable viewing distance according to the following equation:

$$\frac{1}{L} + \frac{1}{D} = \frac{1}{f}, \tag{6.1}$$

where L represents the distance between the projectors and the screen, D denotes the viewing distance, and f signifies the focal length.

The lens maker's law was employed by Lee *et al.* [191] in their system design as follows:

$$\begin{aligned} d_v &= \frac{f d_p}{d_p - f}, \\ p_e &= \frac{p_p d_v}{d_p}, \end{aligned} \tag{6.2}$$

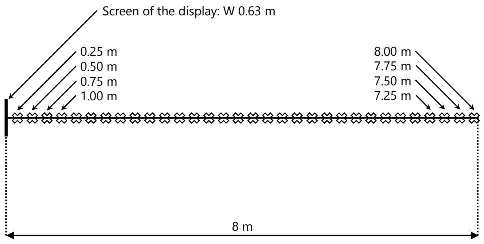
where d_v represents the "ideal" viewing distance, f stands for the focal length, d_p indicates the distance between the projectors and the screen, p_e signifies the interval of viewpoints, and p_p denotes the interval of the adjacent projectors. The system's designated viewing distance was set at 1.2 m, while maintaining a viewpoint interval of 65 mm.

In the study conducted by Kara *et al.* [169] the viewing distance threshold for LFDs (i.e., the maximum distance at which the visualized content is perceived as 3D) is determined through the following calculation:

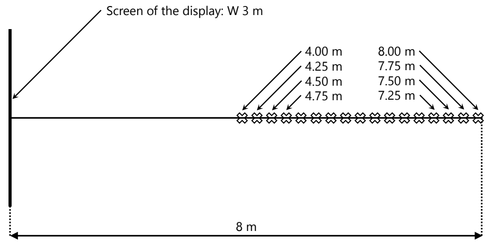
$$D_V = \frac{D_E}{tan(AR)}, \tag{6.3}$$

where D_V, D_E, and AR are the viewing distance threshold, interpupillary distance, and angular resolution, respectively. In a follow-up study [159], the authors discuss the constraints and adaptability of the threshold depending on the specific use case.

[1] https://www.itu.int/dms pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-I!!PDF-E.pdf



(a) Viewing positions for the HoloVizio 80WLT LFD



(b) Viewing positions for the HoloVizio C80 LF cinema.

Figure 6.2: Viewing positions for LFDs [163]

As a means for standardization, the IEEE P3333.1.4[2] standard, focusing on the quality assessment of LF imaging, continues to consider the viewing distance for LF visualization. The work illustrated in this section aims to to provide research data that aids in the standardization endeavors [163].

6.2.2 Experimental setup

For these experiments, the HoloVizio 80WLT and the HoloVizio C80 cinema LFDs were employed. The experiment concerning perceptually-supported viewing distance utilized the 80WLT, while the experiment pertaining to subjectively-preferred viewing distance involved both displays.

Within the research environment, the maximum viewing distance from the screen was determined to be 8 meters. However, it's worth noting that for the perceptually-supported experiment, the viewing distance for the C80 LFD could extend beyond this limit but was excluded as indicated at the end of the preceding subsection. In designing the assessment methodology for the preference tests, this constraint was accounted for, and its impact is reflected in the analysis of the acquired outcomes.

In both experiments, test conductors recorded the preferred viewing distances indicated by the test participants. Visual markers were positioned at intervals of 25 cm from the screen along the central viewing angle line on the laboratory floor for each display. Regarding the 80WLT LFD, the markings began from the screen and continued up to 8 m, totaling 32 marks, as illustrated in Figure 6.2a. Whereas for the C80 LFD, the markings commenced at a distance of 4 m from the screen, extending up to 8 m, totaling 17 marks, as illustrated in Figure 6.2b. The starting point of 4 m was chosen because the C80 utilizes front-projection technology. Consequently, individuals, particularly taller ones, positioned too close to the screen may obstruct the emitted rays from the optical engines, potentially resulting in invalid outcomes.

[2]https://standards.ieee.org/ieee/3333.1.4/10873/

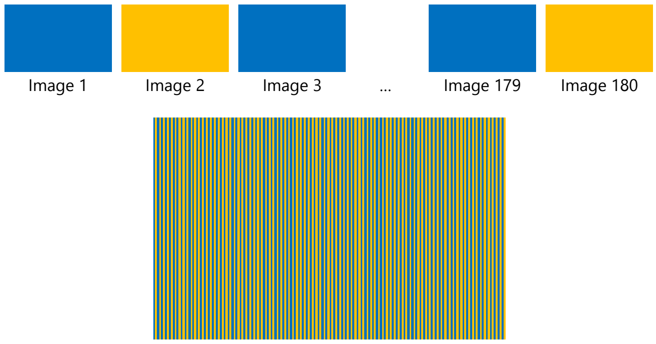


Figure 6.3: Concept of the visual stimulus for the experiment on the perceptually-supported viewing distance [163]

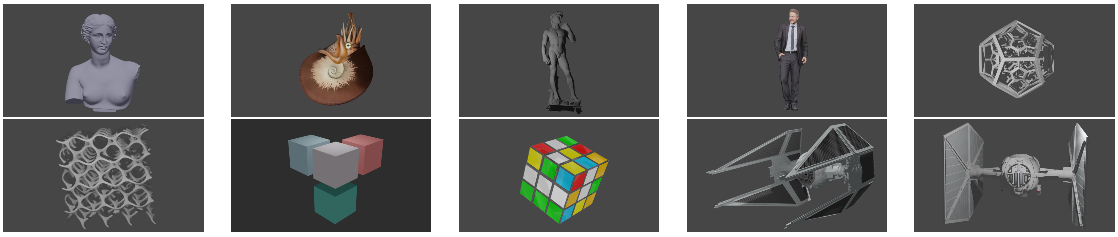


Figure 6.4: The source contents of the experiments [163]

Regarding the perceptually-supported viewing distance experiment, content detail discrimination was used to address the matter by means of generating a visual stimulus consisting of alternating vertical stripes featuring high color contrast. As indicated in Figure 6.3, high contrast was enabled by means of the blue and yellow colors. These were alternately fed to the converter of the 80WLT system creating an image sequence of plain colored images. As stated earlier, the 80WLT has a 1-degree angular resolution and an FOV of 180°. The 1-degree angular resolution practically corresponds to one source image for every degree. Meanwhile, the 180° FOV necessitates 180 alternating images, matching the 2D spatial resolution of the display (1280×768) perceived from any viewpoint.

The perceptually-supported experiment was conducted by instructing test participants to initially stand at a comfortable distance from the display, which could start from as close as 0.25 m. Subsequently, they were asked to gradually increase their viewing distance by moving away from the screen until discriminating content details became unfeasible. The experiment conductors recorded the furthest distance at which visual details were still distinguishable.

In the subjectively-preferred viewing distance experiment, ten source contents were rendered to accommodate the capabilities of both displays (illustrated in Figure 6.4). These contents were rendered separately for each display, with parameters adjusted accordingly each time. All contents shared the same background color, characterized by a medium shade of gray. Content A featured a laser-scanned model of Aphrodite's bust provided by Jotero[3]). Having the same spatial dimensions but textured, content B (Ammonite[4]) was rendered. In order to visualize vertically tall elements, contents C and D were rendered featuring David from The Digital Michelangelo Project[5] and Dennis Posed 004[6], respectively, with the latter being textured. In order to represent complex mathematical bodies, contents E and F (George W. Hart's Rapid Prototyping[7]) were visualized. Contents G and H displayed the Holoxica 3D logo[8] and the animated Rubick's Cube [9], respectively. Finally, the Tie Interceptor[10] and the Tie Fighter[11] models were rendered as contents I and J, respectively, adapted from the Star Wars™ franchise.

During the subjectively-preferred viewing distance experiment, participants were instructed to view the displayed content from various indicated positions (i.e., marked on the floor every 0.25 m) and determine the distance from each display at which they personally preferred visualizing the given content. The necessity of visiting every single position and the sequence in which participants visited the marked positions were inconsequential. Additionally, participants were permitted to revisit positions they had previously visited. Once the test participant indicated the preferred viewing distance, the latter was recorded by the experiment conductors.

Before the experiments commenced, test participants were screened to ensure normal vision using the Snellen chart for visual acuity and the Ishihara plates for color vision. During both experiments, participants were restricted to moving along the middle viewing angle line, with the flexibility for head and body sways. They were instructed to stand still while observing the content.

Finally, in addressing the viewing distance challenge posed by the C80 LFD, participants were prompted to indicate whether their preferred viewing distance extended beyond the 8-meter threshold for the subjectively-preferred experiment. This inquiry carried particular significance for participants generally preferring greater viewing distances.

For both experiments, a total of 22 regular non-expert participants took part. Among them, 12 were male and 10 were female. The age of participants ranged from 21 to 65 years, with an average age of 31. In addition to the regular test participants, both experiments were conducted with individuals possessing expertise in the underlying technology of LF visualization. Their perceptions of the visual quality rendered by the LFDs differ due to their extensive knowledge in this domain [163].

[3] website is no longer available
[4] https://www.turbosquid.com/3d-models/free-ammonite-3d-model/254206
[5] https://graphics.stanford.edu/papers/digmich_falletti/
[6] https://free3d.com/3d-model/dennis-posed-004-812878.html
[7] https://www.georgehart.com/rp/rp.html
[8] https://www.holoxica.com/
[9] https://holografika.com/
[10] https://downloadfree3d.com/3d-models/aircraft/spaceship/tie-interceptor/
[11] https://downloadfree3d.com/3d-models/aircraft/spaceship/tie-fighter-from-star-wars/

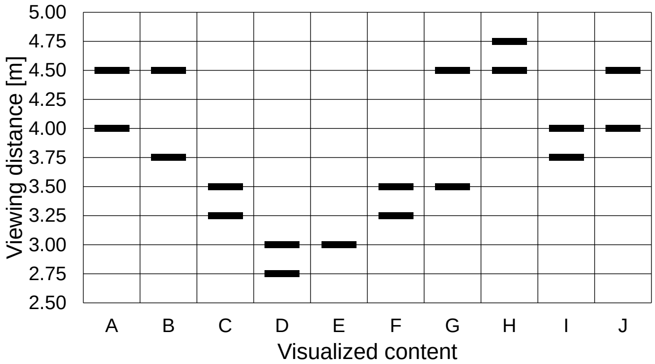


Figure 6.5: Results of the expert analysis on the subjectively-preferred viewing distance via the HoloVizio 80WLT. The markers indicate the intervals used by the experts [163].

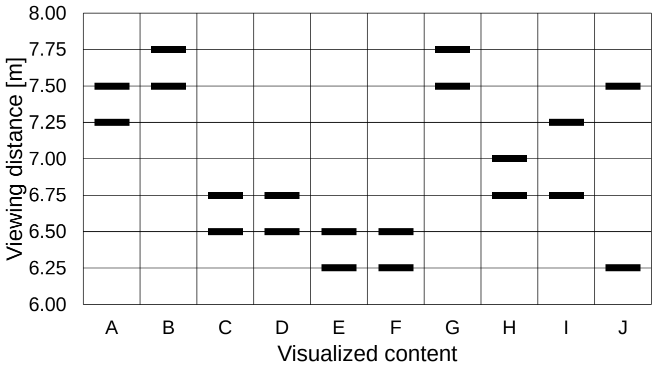


Figure 6.6: Results of the expert analysis on the subjectively-preferred viewing distance via the HoloVizio C80. The markers indicate the intervals used by the experts [163].

6.2.3 Results

Expert analysis

Given the minimal variance among expert assessments, this section presents the ranges within which the results were collected. The results obtained for the perceptually-supported viewing distance fall within the range of 4 m to 5.75 m. According to Equation 6.3, the threshold for the viewing distance for the 80WLT LFD is approximately 3.75 m. Consequently, it appears that the experts' perception extends beyond the point where the perceived content shifts from being predominantly 3D to becoming more 2D. Furthermore, taking into account the height of the display, which measures 390 mm, the distances of 3.75 m, 4 m and 5.75 m correspond to approximately 9.62 H, 10.26 H, and 14.74 H, respectively.

The results concerning the subjectively-preferred viewing distance for both displays (80WLT and C80) are depicted in Figures 6.5 and 6.6. In the case of the HoloVizio 80WLT, for each visualized content, each expert selected a value within a narrow range

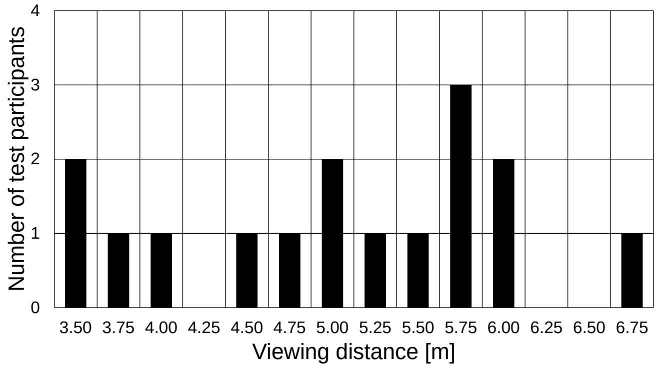


Figure 6.7: Results of the subjective tests on the perceptually-supported viewing distance (excluding outliers) [163]

between the two specified distances, with the exception of content E where unanimous agreement among all experts was reached regarding the optimal distance. Regarding the HoloVizio C80, the unanimity of the results is even more conspicuous, with only two distinct distances being registered for 8 out of 10 contents.

Since experts look for the same visual cues (i.e., particular elements of changes influenced by distance) when assessing perceived quality, their findings tend to be relatively similar. On the other hand, a greater variance in results is anticipated from regular test participants [163].

Perceptually-supported viewing distance

Regarding the perceptually-supported viewing distance, an average of 5.85 m was registered by the 22 regular test participants. However, among them, there were 6 outliers who exhibited greater body sway compared to the others, thereby influencing the results. Furthermore, unlike the remaining 16 participants, all 6 outliers reported perceiving the distinct stripes at the maximum distance of 8 m.

The registered distances for the 16 non-outlier participants are illustrated in Figure 6.7, demonstrating a uniform distribution spanning from 3.5 m to 6.75 m. The mean distance is calculated to be 5.05 m, with the most common distance being 5.75 m, a value comparable to that obtained from experts. As Equation 6.3 is designed for static observation, a distance of 3.75 m was computed. Nevertheless, considering the natural sway, the perceptually-supported viewing distance is observed to be approximately 3.75 m or greater.

Speaking of natural swaying, the experiment conductors observed that the outliers tended to be taller than the other participants. This greater height impacts the results as it leads to a larger horizontal displacement at eye level. Therefore, Equation 6.3 can be

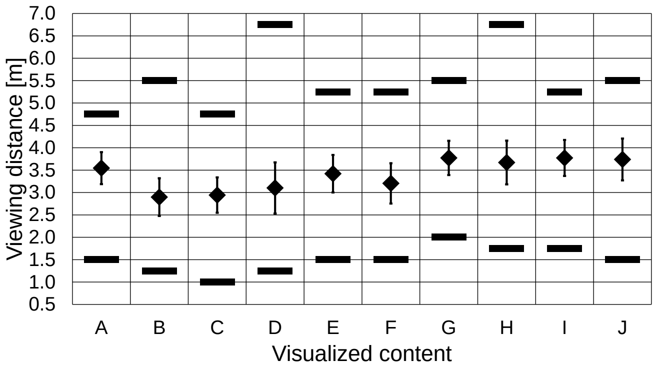


Figure 6.8: Results of the subjective tests on the subjectively-preferred viewing distance via the HoloVizio 80WLT [163]

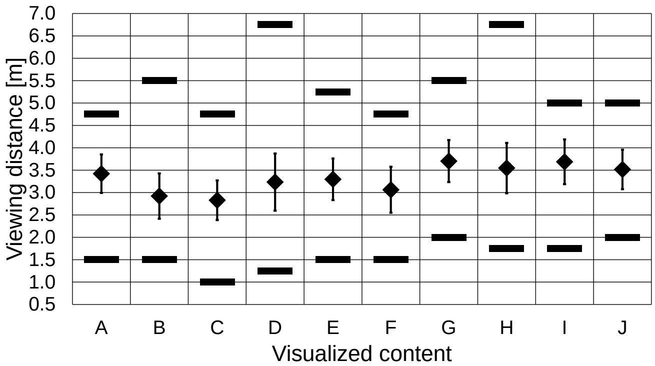


Figure 6.9: Results of the subjective tests on the subjectively-preferred viewing distance via the HoloVizio 80WLT (excluding outliers) [163]

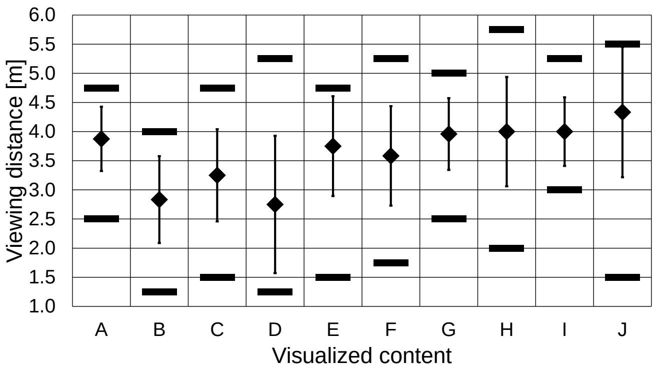


Figure 6.10: Results of the subjective tests on the subjectively-preferred viewing distance via the HoloVizio 80WLT (outliers only) [163]

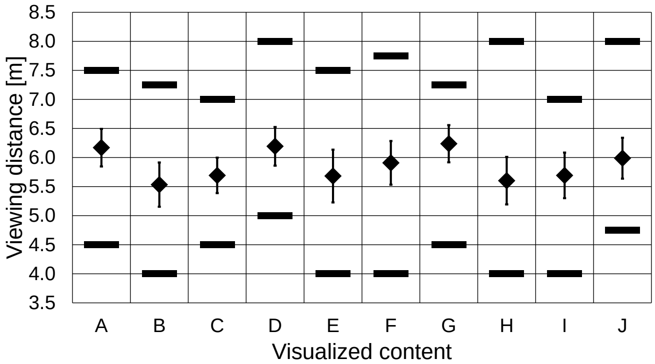


Figure 6.11: Results of the subjective tests on the subjectively-preferred viewing distance via the HoloVizio C80 [163]

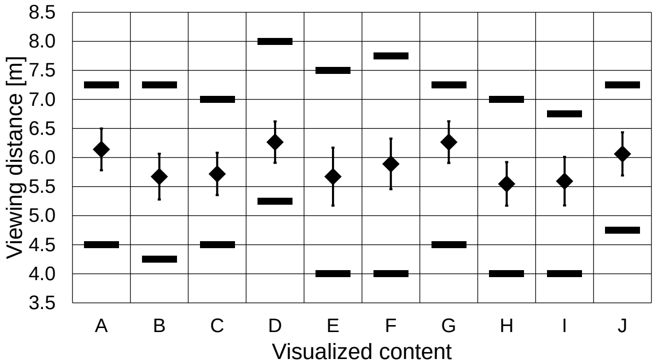


Figure 6.12: Results of the subjective tests on the subjectively-preferred viewing distance via the HoloVizio C80 (excluding outliers) [163].

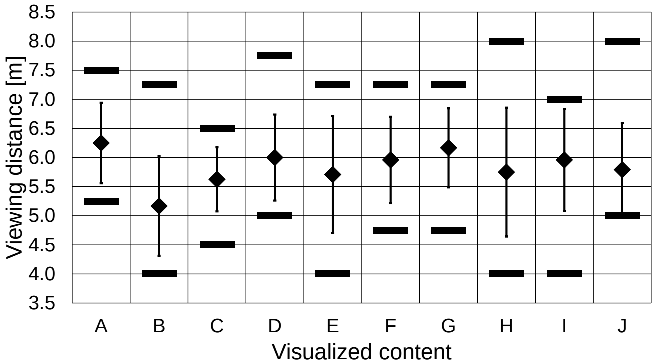


Figure 6.13: Results of the subjective tests on the subjectively-preferred viewing distance via the HoloVizio C80 (outliers only) [163]

adjusted accordingly to:

$$D_V = \frac{D_E + D_S}{tan(AR)}, \tag{6.4}$$

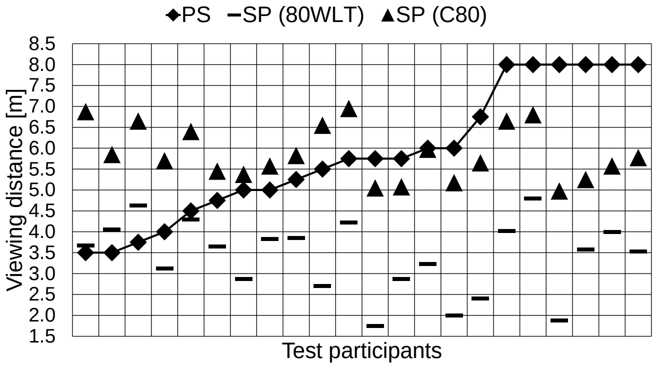


Figure 6.14: Results of the subjective tests on the perceptually-supported (PS) and the subjectively-preferred (SP) viewing distance. Each column represents the mean subjective scores of a test participant, ordered by the results on the perceptually-supported viewing distance [163].

where distance D_S denotes the extent of horizontal displacement caused by swaying. In this scenario, solving the equation for D_S with D_V set to 8 m yields a calculated value of 75 mm. This means that if the sway value reaches or exceeds 75 mm at a distance of 8 m, the perception of a 1-degree angular resolution LFD, which would typically be considered normal for static observation, may be overridden [163].

Subjectively-preferred viewing distance

Drawing from our observations concerning outliers, we proceed to examine the acquired results concerning the subjectively preferred viewing distance both with and without the inclusion of outlier data. Additionally, we conduct a separate analysis focusing specifically on the outlier data. The outcomes for the HoloVizio 80WLT are depicted in Figures 6.8, 6.9, and 6.10. Similarly, the results for the HoloVizio C80 are illustrated in Figures 6.11, 6.12, and 6.13. Each figure displays the average preferred viewing distance, highlighting both the nearest and farthest preferred viewing distances. Additionally, the figures incorporate the 0.95 confidence intervals to enhance clarity and precision.

Examining the results, it is statistically evident that the visual content has a negligible impact on the preferred viewing distance. Whereas some contents are preferably viewed from closer or further positions, their average values –irrespective of test subject classification– tend to cluster within relatively narrow intervals, particularly when taking into account the confidence intervals.

Considering the HoloVizio 80WLT, the mean distances of the visualized contents ranged between 2.9 m and 3.77 m, with a recorded mean viewing distance of 3.4 m. In case of outliers exclusion, the mean viewing distance was 3.32 m. Whereas, the mean viewing distance for the outliers was 3.63 m. As for the HoloVizio C80, the registered values were 5.87 m, 5.53 m, 6.24 m, 5.88 m and 5.83 m, respectively. Regarding the computation of

H values for the 80WLT and C80 LFDs, possessing screen heights of 0.39 m and 1.8 m respectively, the assessments involving all 22 participants yielded values of 8.72 H and 3.26 H, respectively.

As previously stated in Section 6.2.2, considerable focus was directed towards the prospect of the subjectively-preferred viewing distance exceeding 8 m, with respect to the C80. Out of the 220 preferred distances, three instances of an 8 m-viewing distance were observed among three distinct participants, two of whom are outliers. This is indicated in Figures 6.12 and 6.13.

Figure 6.14 presents an overview of the data collected from both experiments. Each test participant's mean subjectively-preferred viewing distance and the corresponding registered perceptually-supported viewing distance are depicted, ordered according to the latter. Consequently, outliers are situated towards the right side of the figure. The comparison reveals significant diversity in preferences, indicating a lack of clear correlation between the experiment results. For the 80WLT, average personal preferences ranged from 1.875 m (4.8 H) to 4.225 m (10.83 H), while for the C80, the range was 4.975 m (2.76 H) to 6.95 m (3.86 H) [163].

6.2.4 Conclusion

Regarding the perceptually-supported experiment conducted on a 30-inch television-like LFD, the values were fairly evenly distributed within the recorded ranges. For experts, the recorded range extended from 4 m to 5.75 m, whereas for regular participants, the interval spanned from 3.5 m to 6.75 m. Despite the presence of outliers in the perceptually-supported experiment, their results concerning the subjectively-preferred viewing distances were not notably distinct from those of the other regular participants. Moreover, no substantial effect was evident in the results concerning the visual stimuli. Overall, the results of the subjectively-preferred experiment for both LFDs exhibited significant variance. Potential areas for future research include addressing the motion of observers across various usage scenarios, as well as examining the extent of head and body sways during static observations [163].

6.3 The effect of angular resolution and 3D rendering on the perceived quality of the industrial use cases of light field visualization

In this study, we focus on the application of LF visualization within industrial contexts. Primarily, LFDs are utilized for prototype visualization, with resource exploration also being a significant application. During prototype reviews, a specific unit or component is often showcased, allowing a group of experts and stakeholders to view it simultaneously. Although the model typically remains static, interactive features –such as rotation and zoom– can significantly improve the effectiveness of these sessions.

The visualized prototypes can exhibit significant variation in their spatial characteristics, including differences in content dimensions and structural complexities. Hence, the main technological challenge associated with the actual shape of the industrial model lies in its potential for high-frequency variations. As the interval of change decreases, a greater angular resolution becomes necessary. Consequently, insufficient angular resolution may lead to the occurrence of crosstalk effect between adjacent distinct sections of the visualized model. Another challenge arises from the depth of rendering. Regarding LFDs, the content is rendered with the highest sharpness on the screen plane, while its quality deteriorates (i.e., becomes blurry) as it extends away from the screen, either towards the viewer or in the opposite direction.

Although angular resolution and 3D rendering of the model may appear to be independent, their effects are, in fact, closely interconnected. For instance, content rendered away from the screen plane is not only blurry, but also more susceptible to the effects of insufficient angular density. In other words, large depth values can cause visual issues, even when the angular resolution is sufficient. While one possible solution to address the aforementioned issue is to render the content on the screen plane, this undermines the purpose of LFDs, as the content would appear more two-dimensional rather than truly three-dimensional.

This study explores the impact of angular resolution and 3D rendering on the perceived quality of LF visualization in industrial scenarios via subjective assessment methodology. The source contents, exhibiting a wide range of structural variations, are rendered at multiple depth levels, with the test stimuli generated across seven distinct angular resolutions. The results obtained, not only illustrate the collective influence of the examined factors, but also draw attention to significant methodological aspects [164].

6.3.1 Related work

There is a substantial amount of scientific literature covering several aspects of LF visualization. In the scope of subjective studies and QoE, investigated topics included spatial and angular resolutions, zoom levels, FOV, content orientation, ROI, among many others.

Regarding angular resolution in the scientific literature on QoE for LF visualization, the highest value is typically 0.5 degrees, with values often reduced to at least 1 degree. Some studies investigate lower angular resolutions, such as 2 degrees, but these often result in significantly diminished perceived quality. Research generally shows that an angular resolution of 0.5 degrees provides excellent visualization quality, while 1 degree is considered sufficient –or acceptable– for most use case scenarios [164].

In this subjective study, we adhered to the best practices established by the scientific community. The details of the experimental setup are presented in the following section.

6.3.2 Experimental setup

For this study, the HoloVizio HV640RC was utilized due to its capability to support an angular resolution up to 0.5 degrees. The display features an FOV of 56 degrees, which means that achieving an angular resolution of 0.5 degrees necessitates the use of 112 source views for rendering. All source views were rendered at a spatial resolution of 1024×768 corresponding to that of the display.

The study was conducted in a conventional laboratory environment with only two individuals present during the test: the experiment conductor and the test participant actively performing the experiment.

Angular resolution was the sole test variable in the study, with the highest and lowest values set to 0.5 and 2 degrees, respectively. These correspond to 112 rendered source views for the 0.5-degree resolution and 28 rendered source views for the 2-degree resolution. A difference of 14 views was incorporated between adjacent values, resulting in 7 test conditions rendered from 28, 42, 56, 70, 84, 98, and 112 source views.

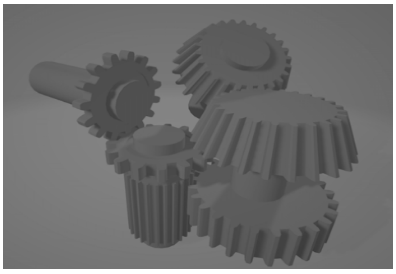
For this experiment, we created 7 industrial models, each featuring distinct spatial characteristics. The source contents are illustrated in Figure 6.15. Contents A and E featured a set of gears and a ratchet, respectively. Both exhibited the most pronounced depth variation in the study. In other words, a significant portion of the content extended beyond the plane of the screen. Contents B and C, featuring heatsink models, had less depth. However, the fin design was intentionally crafted to evaluate the angular density of the setup. Content D was a lathe, featuring smooth surfaces and consistent depth. However, the sharp yet continuous edge of the spindle made it prone to degradation. Lastly, contents F and G, which included models of a suspension and a turbine blade, respectively, exhibited the least depth variation.

Each source content was rendered 7 times, corresponding to each of the 7 test conditions, resulting in a total of 49 visual stimuli. The models were rendered following the established protocol for the 972-face polyhedron as described in previous studies [67, 161, 160, 289, 72, 273]. This approach involved using a single color for all models while placing them against a background of a different color. The color schemes applied were based on those used for the polyhedron rendering.

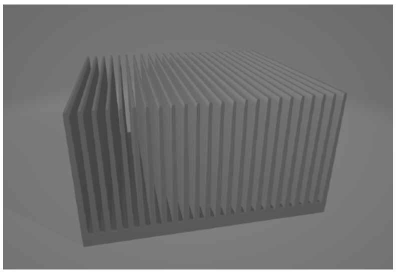
The task for participants in this study was to evaluate the visual degradation associated with reduced angular resolution. The technical cause of the degradation was not revealed to the participants. The evaluation utilized a 5-point Degradation Category Rating (DCR) scale [12] and implemented the hidden reference method. Within this framework, the reference visual stimulus, generated from 112 views, was also evaluated by the test participants.

For assessment, the DCR scale was employed due to its dual functionality, utilizing numerical values ranging from 1 to 5. The scale values from 1 to 5 corresponded,

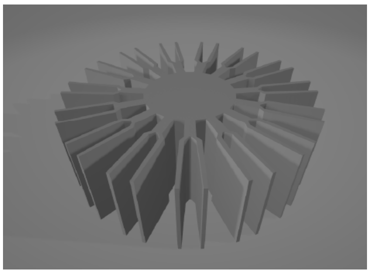
[12] ITU-T Rec. P.910: Subjective video quality assessment methods for multimedia applications



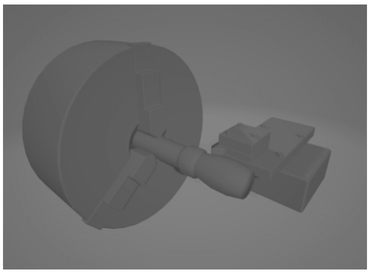
(a) Content A



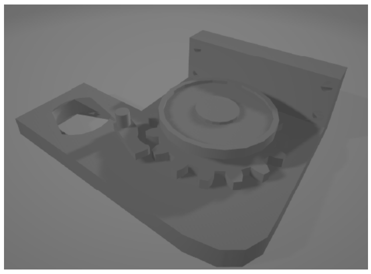
(b) Content B



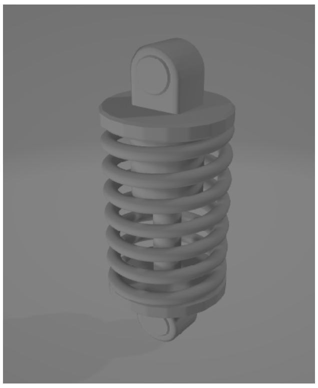
(c) Content C



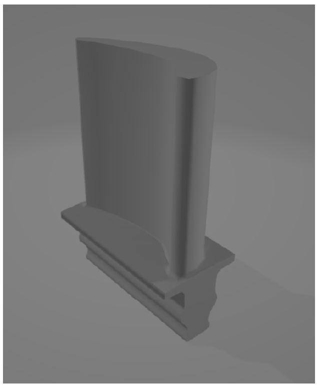
(d) Content D



(e) Content E



(f) Content F



(g) Content G

Figure 6.15: Source contents of the subjective tests [164]

respectively, to “Very annoying”, “Annoying”, “Slightly annoying”, “Perceptible but not annoying”, and “Imperceptible”. The DCR scale serves a dual function: it first determines whether degradation is perceptible compared to the reference, with a rating of “Imperceptible” indicating no noticeable difference. If the degradation is distinguishable from the reference, the scale further measures the degree of annoyance experienced.

During the recruitment phase, test participants underwent vision screening using the Snellen chart and the Ishihara plates to ensure normal vision. Participants were extensively trained for the assessment task to avoid significant rating issues. Inadequate training, particularly regarding the effects on parallax smoothness, could lead to substantial inaccuracies in their evaluations [167].

A viewing distance of 1.86 meters was selected, based on the following equation [169, 159]:

$$Viewing\ distance \leq \frac{Interpupillary\ distance}{tan(Angular\ resolution)}, \tag{6.5}$$

where, according to the scientific community, the interpupillary distance is approximately 6.5 cm. Substituting the interpupillary distance and the minimum angular resolution of 2 degrees results in a maximum viewing distance of 1.86 m. Greater distances were feasible; however, visual stimuli with lower angular resolution would be perceived as 2D rather than 3D. Consequently, to eliminate the influence of this factor, the distance derived from the equation was adopted. Participants observed the various models on LFDs from a fixed central position, maintaining a distance of 1.86 m from the screen. Additionally, participants were permitted to lean sideways to improve their perception of the parallax effect, provided they did not move sideways or step in any lateral direction.

The subjective study involved 43 participants, consisting of 25 males and 18 females, with ages ranging from 20 to 66 and an average age of 25. Participants verbally rated the quality of each visual stimulus, with these evaluations being documented by the test conductor [164].

6.3.3 Results

Given the 49 visual stimuli, each participant provided 49 separate assessments, resulting in a total of 2107 ratings across all participants.

Figures 6.16 and 6.17 illustrate the mean DCR scores and the distribution of ratings for each test condition, respectively. A total of 301 ratings (43 participants $\times$ 7 visual stimuli) represent each test condition.

The consistency among average scores is evident, and significant differences between certain pairs of adjacent test conditions are observed, as indicated by the non-overlapping 0.95 confidence intervals. This can be seen with the 98-source view test condition, which received notably higher ratings than the 84-source view test condition. Likewise, the 42-source view condition outperformed the 28-source view condition.

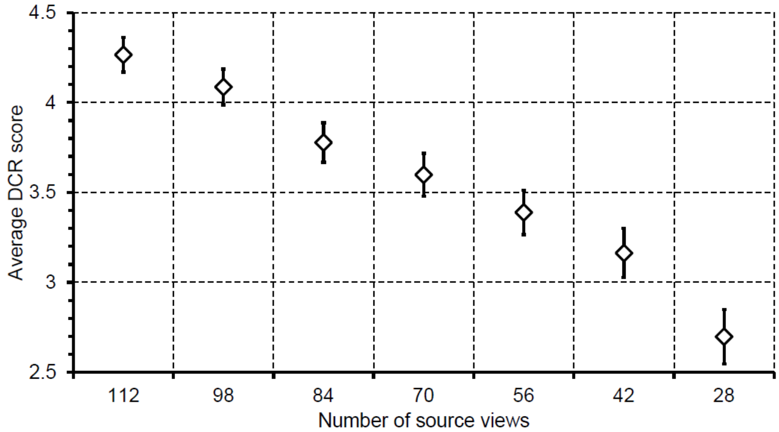


Figure 6.16: Average DCR scores of the test conditions [164]

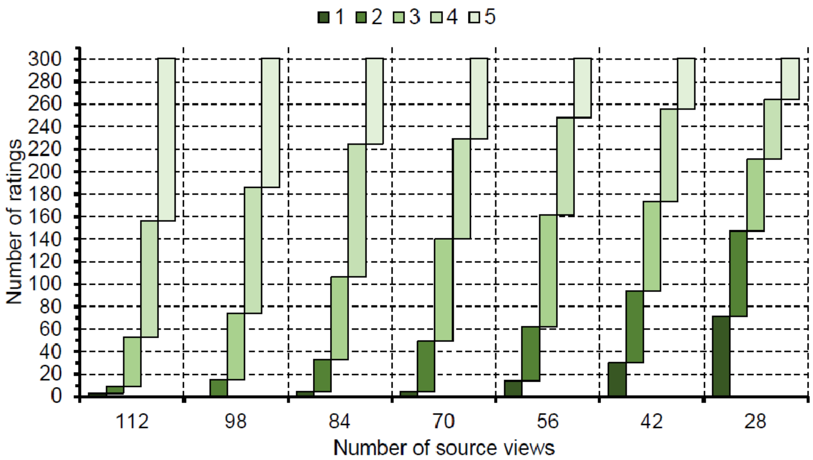


Figure 6.17: Rating distribution of the test conditions [164]

Based on the actual mean values, test conditions with 112 and 98 source views are rated between "imperceptible" and "perceptible but not annoying". Test conditions with 84, 70, 56, and 42 source views fall within the range of "perceptible but not annoying" to "slightly annoying". Lastly, the test condition with 28 source views is situated between "slightly annoying" and "annoying". A notable issue is that the 112-source view condition acted as the hidden reference, thus it was essentially evaluated against itself. Despite this, an analysis of the rating distribution shows that just 145 ratings, or 48.17%, were deemed "imperceptible". This suggests that more than half of the ratings detected a noticeable difference between the reference and the hidden reference. This observation can be attributed to the manner in which the source contents were rendered. Given that these stimuli exhibited significant depth variations compared to typical studies on general visualization quality, even with 112 source views, both the nearest and farthest parts of

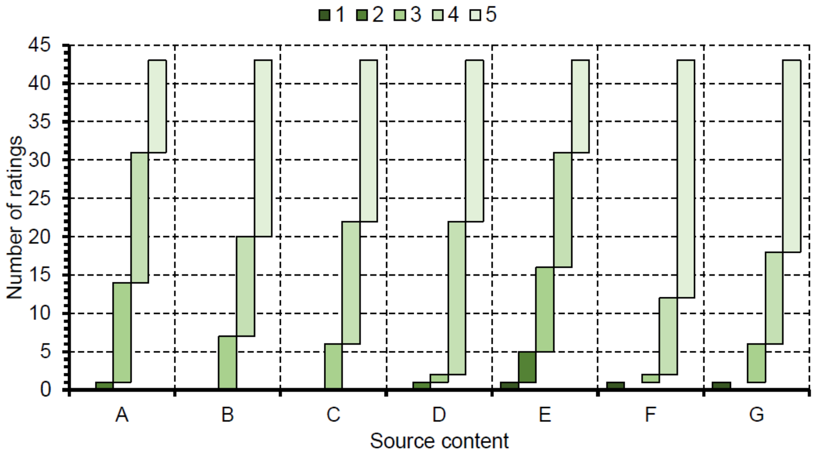


Figure 6.18: Rating distribution of the source contents at 112 source views [164]

the models experienced some degree of degradation. This effect was less noticeable for models F and G, but more pronounced for models A and E. Since the reference stimulus itself was not flawless, it introduced a degree of cognitive bias, leading to distortion in the collected data.

A prominent type of cognitive bias is the misinformation effect [212]. While the reference stimulus is intended to provide an ideal representation of the source content, theoretically exhibiting perfect quality, the hidden reference, despite being intended for close scrutiny and evaluation, does not possess the same level of perfection. Consequently, although no discernible difference should theoretically exist between the reference and the hidden reference, the actual assessment reveals imperfections in the hidden reference. Even if there are recollections of the reference stimulus being imperfect, these memories can be influenced or modified by integrating its conceptual understanding with the new visual information received.

The alteration in ratings can be attributed to the misinformation effect, but only insofar as new information modifies the old one. In this context, the distortion is compounded by the expectation that the reference stimulus is supposed to be flawless. Even if the reference is degraded, it is assumed that this degradation is minimal and should be less noticeable than that of the test stimuli.

The distortion in the ratings varied across different source contents. The distribution of ratings for the source contents rendered at 112 source views, illustrated in Figure 6.18, reflects this. As anticipated, given the depth values during rendering, contents A and E experienced the most significant impact, whereas contents F and G experienced minimal effects. Among the 43 ratings submitted by participants, only 12 (27.9%) considered the degradation in contents A and E at reference quality as "imperceptible" when compared to their respective reference stimuli. Conversely, for contents F and G, 31 (72.09%) and 25 (58.14%) participants, respectively, found the degradation to be "imperceptible". It

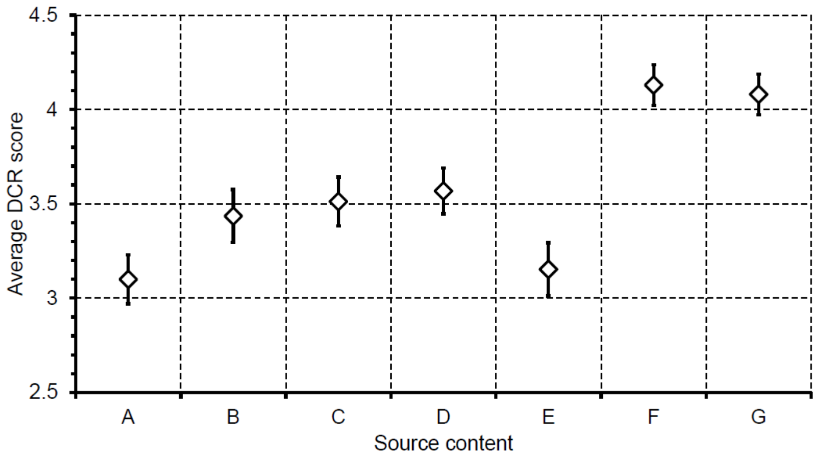


Figure 6.19: Average DCR scores of the source contents [164]

is noteworthy that 14 participants (32.56%) found content A to be annoying, while 16 participants (37.21%) expressed the same sentiment for content E. This indicates that more participants experienced annoyance with these two contents compared to those who were unable to differentiate between the reference and the hidden reference. It's important to mention that contents F and G each received a single "very annoying" rating, both from the same participant.

If any objective quality metric was used to assess the test stimuli, the hidden reference would receive "imperceptible" ratings due to the absence of discernible differences. In QoE research, the goal is to close the gap between subjective and objective (i.e., predicted subjective) ratings by refining objective measures. However, in this specific situation, it would be more appropriate for the subjective ratings to more closely resemble the objective ones.

This particular bias arises because stimuli are presented in sequence, leading to comparisons between the perception of one stimulus and the memory of another. A simple solution to avoid this issue is to present stimuli simultaneously, such as through a side-by-side comparison. While this approach works well for 2D displays, it cannot be applied to LFDs, as the stimuli would be viewed from different angles.

Figure 6.19 presents the average DCR scores for the source content, with each score based on 301 ratings. Similar patterns to those in Figure 6.18 emerge, where contents A and E received the lowest scores in all test conditions, contents B, C, and D performed moderately better, and contents F and G were rated the highest by the participants.

These results are statistically significant, as the 0.95 confidence intervals for the three groups (A and E; B, C, and D; F and G) do not overlap. The comprehensive statistical analysis of the ratings for both the test conditions and source contents is detailed in Tables 6.2 and 6.3, respectively.

Table 6.2: Statistical analysis of the test conditions[164]

Test conditions	***p* value**
112 / 98	0.01
112 / 84	< 0.01
112 / 70	< 0.01
112 / 56	< 0.01
112 / 42	< 0.01
112 / 28	< 0.01
98 / 84	< 0.01
98 / 70	< 0.01
98 / 56	< 0.01
98 / 42	< 0.01
98 / 28	< 0.01
84 / 70	0.03
84 / 56	< 0.01
84 / 42	< 0.01
84 / 28	< 0.01
70 / 56	0.02
70 / 42	< 0.01
70 / 28	< 0.01
56 / 42	0.02
56 / 28	< 0.01
42 / 28	< 0.01

We conducted a Student's t-test for each combination of a test condition and a source content, resulting in 21 pairwise comparisons for each analysis, given the 7 test conditions and 7 source contents. To control the family-wise error rate and reduce Type I errors (i.e., false positives), the Bonferroni correction adjusts the α level for the p value from 0.05 to 0.00238. In the results mentioned, whenever p is less than 0.01, it also falls below 0.00238. As a result, the findings related to the source contents remain valid even with the Bonferroni-corrected α level, as the statistical significance is still upheld. However, this adjustment impacts the significance of four comparisons related to the test conditions.

Figure 6.20 illustrates the mean DCR scores for the source contents across the various numbers of source views. The results demonstrate the response of each source content to decreases in angular resolution, specifically highlighting the effect of this degradation on each individual content. The content with the smallest difference in average DCR

Table 6.3: Statistical analysis of the source contents [164]

Source contents	p value
A / B	< 0.01
A / C	< 0.01
A / D	< 0.01
A / E	0.59
A / F	< 0.01
A / G	< 0.01
B / C	0.44
B / D	0.16
B / E	< 0.01
B / F	< 0.01
B / G	< 0.01
C / D	0.53
C / E	< 0.01
C / F	< 0.01
C / G	< 0.01
D / E	< 0.01
D / F	< 0.01
D / G	< 0.01
E / F	< 0.01
E / G	< 0.01
F / G	0.52

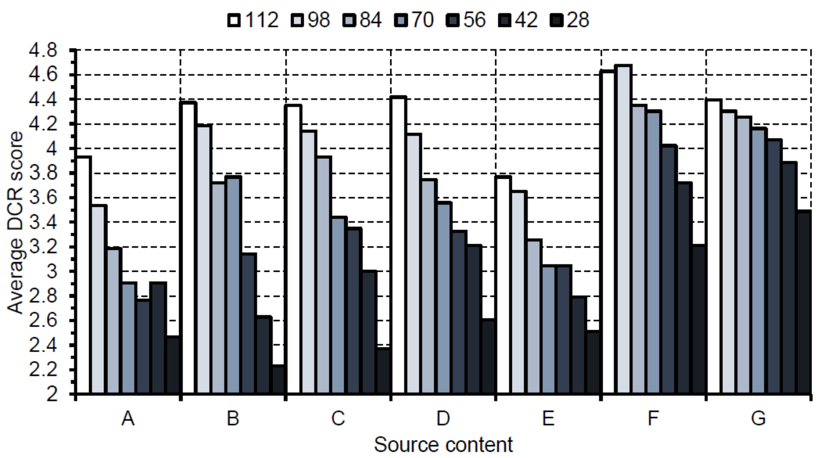


Figure 6.20: Average DCR scores of the source contents for the different numbers of source views [164]

score was content G, showing a variation of 0.9 between the lowest and highest angular resolutions. In contrast, content B exhibited the largest difference, with a variation of 2.14. Contents C and D also demonstrated substantial differences in average scores, with variations of 1.98 and 1.81, respectively. In comparison, contents A, E, and F had smaller differences, recording variations of 1.47, 1.26, and 1.42, respectively. The significant difference between the extremes of angular resolution for contents B, C, and D can be attributed to two factors. First, these contents displayed more pronounced depth variations than contents F and G, making them more susceptible to the effects of lower angular resolutions. Second, the ratings for the hidden reference were higher for these contents compared to A and E, which prevented the remaining rating tasks from being compressed into a narrower scale. Consequently, the depth of the content was sufficient to negatively impact visual quality when angular resolution was reduced, while still being low enough to allow for a broader range in the quality assessment scale. Moreover, slight inconsistencies were observed in similar test conditions, though they were rare across the results [164].

6.3.4 Conclusion

In this study, we investigated the impact of angular resolution and 3D rendering on the perceived LF visualization quality within the framework of industrial applications. The findings suggest that the angular resolution and 3D rendering (i.e., depth of source content) have a statistically significant effect on QoE. Significant differences were identified within the examined angular resolution range of 0.5 to 2 degrees, even when there was a 14-source view difference across a 56-degree FOV between adjacent test scenarios. Regarding 3D rendering, the quality ratings directly corresponded to the classification based on source content depth, and the statistical significance of the findings remained unchanged despite applying a Bonferroni correction to the α level. In future research, we aspire to investigate various use case contexts while considering their unique characteristics and attributes [164].

6.4 The perceived quality of light field visualization assessed by test participants with imperfect visual acuity and color blindness

Regarding LF QoE, understanding how individuals without visual impairments perceive objective metrics is essential. However, with a growing number of younger individuals experiencing sight-related issues, there is an urgent need to understand how they interpret LF visualization. Despite this, the lack of subjective tests involving participants with impaired vision hinders our comprehension in this area.

In this section, we introduce our preliminary investigation into the quality of LF visualization as assessed by participants with impaired visual capabilities. As this study marks the initial phase of an extensive research series, we commence by examining the

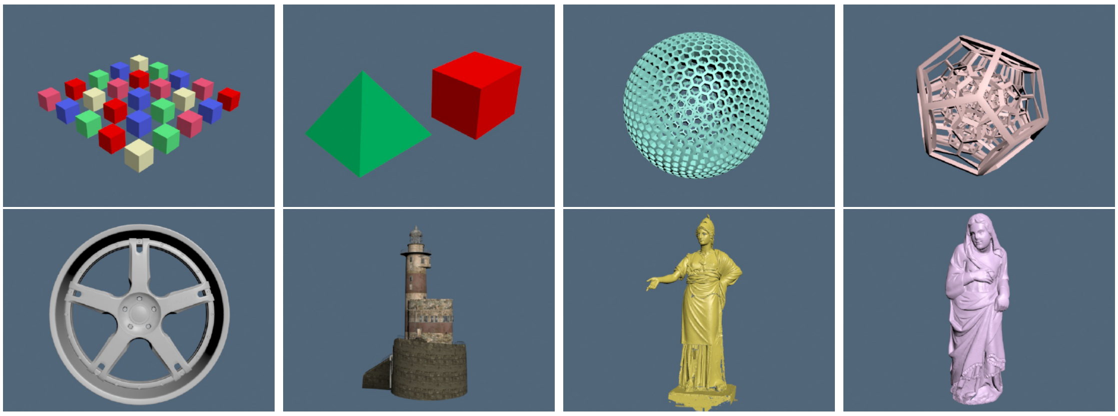


Figure 6.21: The source contents of the subjective study [273]

spatial and angular resolutions, recognized as critical KPIs in LF visualization [171]. Acknowledging the potential implications of varying viewing distances for participants with visual impairments, we conduct an investigation across multiple viewing distances. In addition, observer motion is integrated into the experimental setup. However, its thorough analysis is reserved for future investigations [273].

6.4.1 Experimental setup

The experiment utilized the HoloVizio HV640RC LFD. Within this study, three factors influencing LF visualization were explored: spatial resolution, angular resolution, and viewing distance, each adjusted to 2, 3, and 2 values, respectively. The spatial resolution was configured to either 640×480 or 1024×768. For the angular resolution, LFs were generated using 56, 84, and 112 source views. Given the display's 56-degree FOV, these values translate to angular resolutions of 1 degree, 0.66 degrees, and 0.5 degrees, respectively. The two viewing distances used in the study (1.86 m and 3.72 m) were calculated using Equation 6.5 [169, 159], with an average interpupillary distance of about 6.5 cm. Consequently, the maximum viewing distance (denoted as D_V) is directly determined by the angular resolution of the system and the visualized content. In this context, the lowest angular resolution of 1 degree corresponds to a D_V of 3.72 m. Another distance selected for the tests was established at half of the recommended maximum viewing distance.

The visual stimuli for the test comprised eight static LF scenes, selected for their performance in previous research. As illustrated in Figure 6.21, the 3D models encompassed a spectrum of shapes, ranging from simple to complex, with a variety of colors, structures, depths, and textures. In our analysis, we label the source contents as A, B, C, D, E, F, G, and H, arranged from left to right and top to bottom.

The combination of all test variables produced 12 distinct test conditions (2 spatial resolutions $\times$ 3 angular resolutions $\times$ 2 viewing distances). Two of the twelve test conditions were selected as reference points, featuring the highest spatial and angular resolutions at the two distinct viewing distances. The remaining ten test stimuli were assessed by the participants. In total, 48 visual stimuli were generated from the 8 source contents for every possible resolution combination.

A 10-point ACR scale was employed to give more options to offer greater differentiation among the 48 visual stimuli, with 10 denoting the reference quality and 1 representing the lowest quality.

The test was administered individually to each participant. In terms of the procedure, the test stimuli were presented in a random order. Initially, the reference visual stimulus was shown to the participant without being subject to evaluation. Subsequently, six test stimuli, all of the same model and necessitating evaluation, were displayed, each separated by a plain screen. These six stimuli included the reference stimulus, in accordance with the hidden reference methodology, whereby the reference stimulus was assessed without the participant's awareness.

The assessments were conducted at the designated viewing distances (previously calculated) marked on the laboratory floor. The primary viewing position was centrally located, allowing for slight lateral movement limited to a single step to either side from the center.

The study included 15 participants, whose ages ranged from 26 to 72 years, with an average age of 40. The group comprised 10 males and 5 females. Each participant used corrective eyewear, either glasses or contact lenses, with diopter values between -6 and +2. Eight participants in the study were color-blind, predominantly exhibiting difficulties in distinguishing red and green hues. The remaining 7 participants displayed significantly higher diopters. These participants are referred to as Group 1 and Group 2, respectively. Prior to the test procedure, all participants underwent a training phase to acquaint themselves with the assessment task and the parallax effect that impacts the visualization experience, which, in case of degradation, can lead to the occurrence of crosstalk effect [273].

6.4.2 Results

Each participant provided a total of 96 ratings, resulting in 1440 subjective scores across all 15 participants. The Mean Opinion Scores (MOS) of all participants is depicted in Figure 6.22, highlighting the significance of the hidden reference methodology. Although a score of 10 denotes the reference quality, the MOS assessed for the reference at distances of 1.86 m and 3.72 m were 8.016 and 8.316, respectively. Among the 240 rated evaluations for the reference stimuli at both viewing distances, only 59 were rated as 10, which accounts for less than 25%. Without incorporating the hidden reference, one might anticipate a generally lower perceived quality level.

Results indicate that for the stimulus rendered from 84 views, participants were unable to distinguish it between the reference stimulus and the one rendered at the higher spatial resolution. Indeed, the degraded stimulus achieved a higher MOS of 8.025 compared to the reference stimulus at the closer viewing distance.

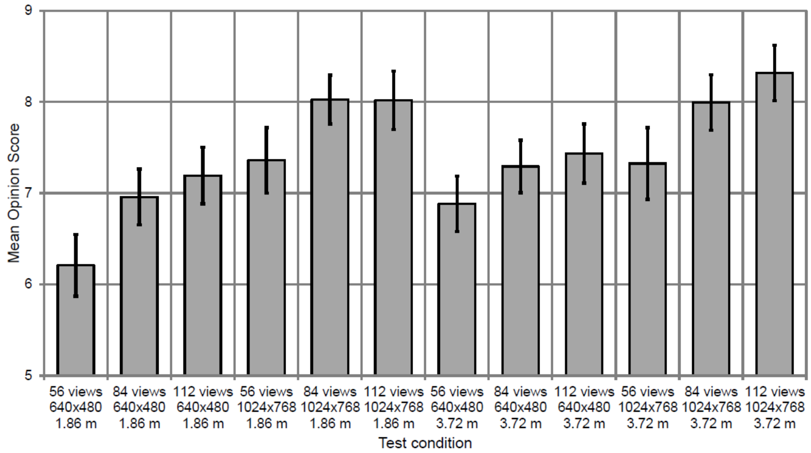


Figure 6.22: MOS of the test conditions for all test participants [273]

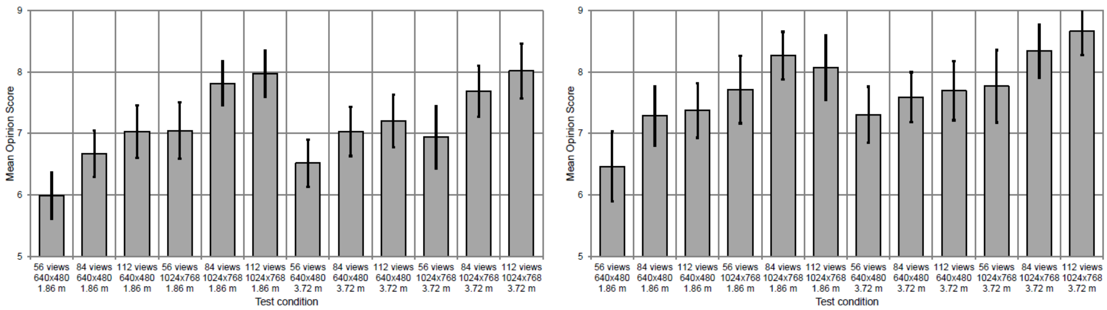


Figure 6.23: MOS of the test conditions for Group 1 (left) and Group 2 (right) [273]

Hence, the impact of viewing distance on the obtained results is evident. Participants perceived more impairment indicators in the degraded quality stimuli when observed from the closer viewing distance. This is especially true for the blurring effect caused by low spatial resolution. Analyzing the data obtained at the closer viewing distance shows a statistically significant difference when comparing every spatial resolution across all three angular resolutions.

As shown in Figure 6.23, similarities in the rating tendencies are evident for both groups. However, overall, Group 1 (color-blind) ratings exhibited lower values. It is noteworthy that for Group 2, at the closer viewing distance, there is an 0.2 MOS difference between the stimuli produced from 84 views at high spatial resolution and the reference stimulus.

Figure 6.24 illustrates the impact of each test variable. The rating tendencies for the lowest and highest angular resolutions are similar to those observed for the two spatial resolutions, with a statistically significant difference being evident. Between the two viewing distances, there is a visible difference of approximately 0.25 MOS. As previously noted, this difference can largely be attributed to the scoring associated with the variation

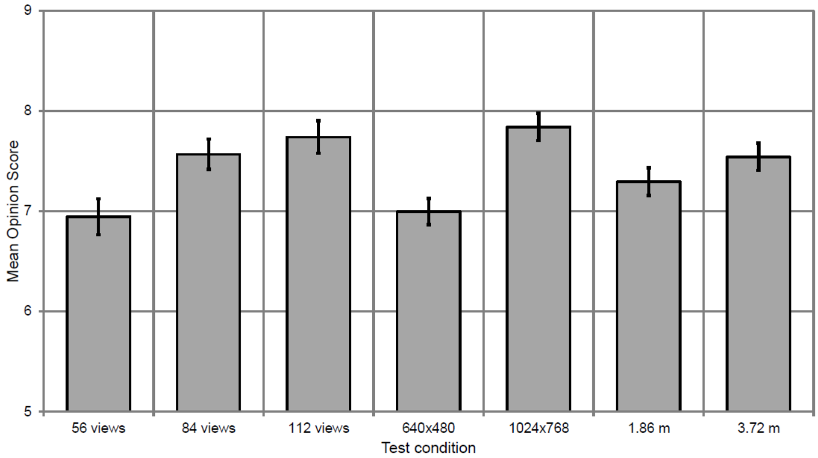


Figure 6.24: MOS of the test conditions for each test variable for all participants [273]

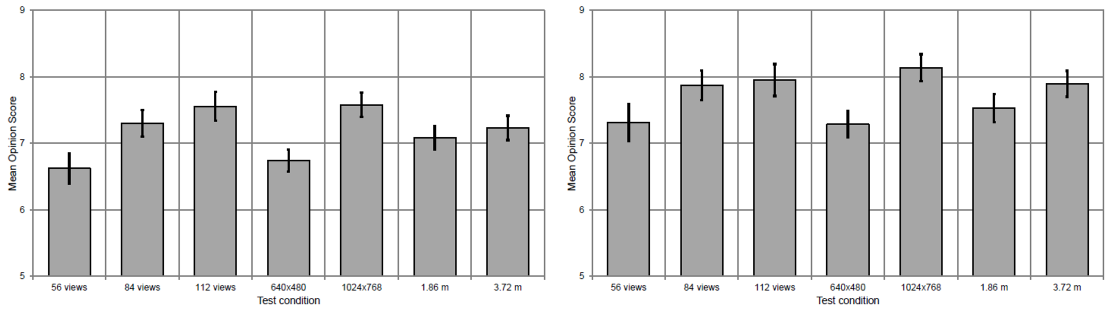


Figure 6.25: MOS of the test conditions for each test variable for Group 1 (left) and Group2 (right) [273]

in spatial resolution.

The influence of each test variable on each group is depicted in Figure 6.25, highlighting the notable differences in their respective MOS values. The differences between the respective three angular resolutions, two spatial resolutions, and two viewing distances are 0.69, 0.57, 0.4, 0.55, 0.56, 0.44, and 0.66, respectively. In the first group, the influence of viewing distances outweighed that of the top two angular resolutions. Conversely, in the second group, the opposite was true.

A significant limitation of MOS analysis is its failure to report rating inconsistencies. Rating inconsistencies refer to instances where a representation of lower quality receives a higher or equivalent rating. The issue of rating inconsistency regarding experimental validity has been examined in the scientific literature [167]. However, this was attributed to inadequate training procedures, unlike the current study, where participants underwent a prior training phase.

In the context of the current experiment, the angular resolution exhibited the highest

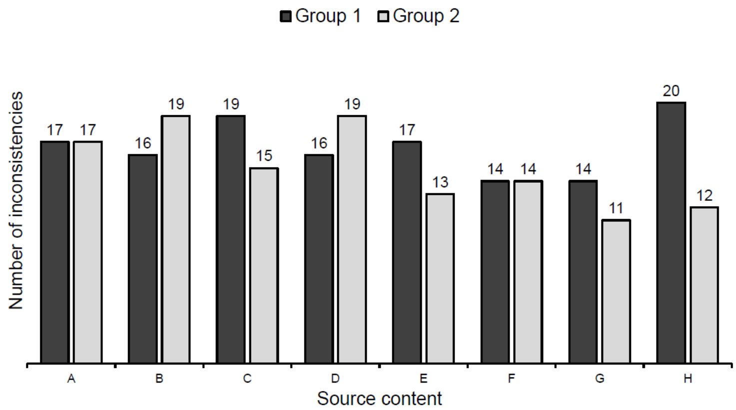


Figure 6.26: Number of rating inconsistencies related to angular resolution per source content [273]

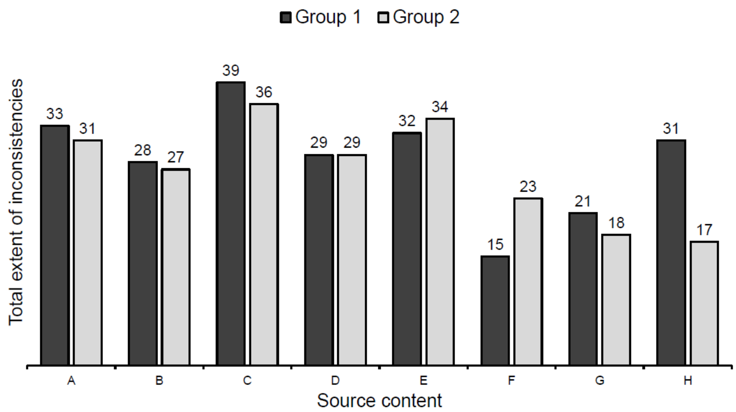


Figure 6.27: Total extent of rating inconsistencies related to angular resolution per source content [273]

rating inconsistencies. Twelve quality ratings were provided for each source content. Accordingly, each test participant assessed 32 stimulus triplets of angular resolution (2 spatial resolutions $\times$ 2 distances $\times$ 8 source contents), where each source content is characterized by 60 triplets (4×15). For instance, a triplet might consist of 56 views/1024 $\times$ 768/3.72 m; 84 views/1024 $\times$ 768/3.72 m; and 112 views/1024 $\times$ 768/3.72 m. A triplet is deemed inconsistent if a higher quality rating is assigned to a stimulus with a lower angular resolution. The degree of inconsistency is measured by the largest discrepancy in ratings. In the context of the preceding triplet example, if the scores assessed are 8, 7 and 8, respectively, then the degree of inconsistency is 1. If the scores are 8, 7 and 6, respectively, then the degree of inconsistency is 2 and not 3, since only the greatest difference is taken into account. In that case, the difference between the first and third ratings is considered.

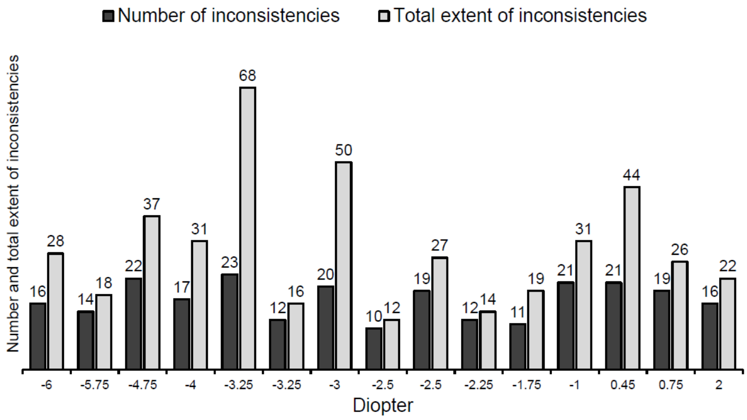


Figure 6.28: Number and total extent of rating inconsistencies related to angular resolution per diopter [273]

Figures 6.26 and 6.27 illustrate the quantity and overall extent of inconsistencies with regards to the angular resolution for each source content, respectively. Regarding most visual stimuli, the rating inconsistencies show no significant variation between the two groups, except for the laser-scanned content H. This could be attributed to the minimal variations in depth. The other laser-scanned statue (content G) features an arm extending outward, which could provide a distinct visual cue for detecting variations in the smoothness of the parallax effect.

It is noteworthy that there were between 15 and 20 inconsistencies observed across the majority of contents. In other words, for many source contents, almost third of the assessments of the stimulus triplets were inconsistent regarding the angular resolution, since each source content is characterized by 60 triplets.

Another perspective on rating inconsistencies can be examined through the diopter values of participants. Figure 6.28 demonstrates the negligible to nonexistent effect of diopter values on the extent and number of rating inconsistencies. For each test participant, the number of inconsistencies varied between 10 and 23, with an average of around 17. The majority of participants exhibited inconsistency extents around 1 and 2, with an average value of approximately 1.75. Additionally, there were outliers among the participants, displaying an average inconsistency extent of 2.96 [273].

6.4.3 Conclusion and future work

This study initiated our research into the QoE of LF visualization as experienced by individuals with color deficiencies and reduced visual acuity. The study results indicate that participants with color blindness assigned lower scores to the visual stimuli. Additionally, the importance of viewing distance and spatial resolution is emphasized, alongside the number and extent of rating inconsistencies related to angular resolution.

For future work, it is recommended that the KPIs analyzed in this study be examined more thoroughly and individually to address their interdependencies. In such cases, unlike this study, where a single variable was limited to a maximum of three values to avoid prolonging the test duration, more values can be assigned to a single variable when examined individually. Moreover, the exploration of further KPIs is recommended. Given the use of LFDs, it is important to investigate their long-term utilization, particularly concerning the potential for perceptual fatigue. In the case of passive and active use cases, movement of observers and task performance should be investigated, respectively [273].

6.5 Analysis of the suitable viewing distance ranges of light field visualization usage contexts for observers with reduced visual capabilities

This study elaborates on our research regarding the preferred viewing distance for potential applications of LF visualization, as perceived by users with reduced visual capabilities. The study extends the previous research outlined in Section 6.4, which, to the authors' understanding, was the first attempt to involve such users in assessing LF QoE. In contrast to the prior study that examined multiple factors including angular and spatial resolutions along with viewing distance, this study concentrates solely on determining the preferred viewing distance. Hence, this study utilized six viewing distances instead of two, assessing the quality of two combinations of resolution values. Furthermore, participants in this experiment experienced issues related to visual acuity rather than deficiencies related to color vision. Among the test participants, one experienced a vision loss exceeding 90%, the findings of which are presented separately [274].

6.5.1 Related work

Numerous factors, including HCI, viewing distance, content type, display parameters, among others, can be utilized to determine the potential use cases of LF visualization. The viewing distance, specifically, has prompted a multitude of research inquiries, primarily due to its direct association with technical parameters. Technically, as the viewing distance increases, the perceived angular resolution required for a 3D experience diminishes. It is apparent that diverse use cases exhibit varying requirements concerning viewing distance. The common practice calculates the recommended maximum viewing distance according to the angular resolution of visualization [159], based on Equation 6.5.

Unlike the majority of studies on subjective LF visualization that include participants pre-screened for visual acuity and color vision, our prior research encompassed a series of subjective tests conducted with participants exhibiting imperfect visual acuity and color vision deficiencies [273]. In that study, various factors were involved; therefore, the variables were restricted to minimize the overall test duration. To ensure time efficiency, only two viewing distances were considered in this study: D_V and $0.5 \times D_V$, with D_V

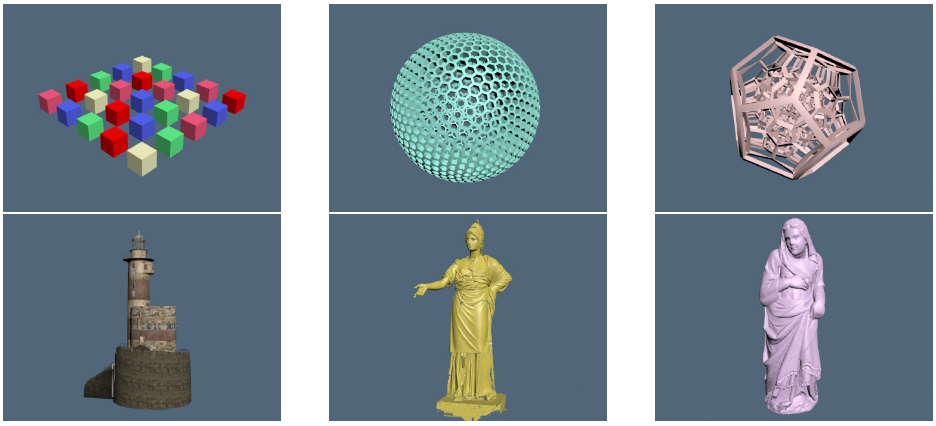


Figure 6.29: The source contents of the subjective study [274]

calculated to be 3.72 m based on a 1-degree angular resolution. Regarding the designated viewing distances, the overall ratings for both of them did not exhibit a statistically significant difference [274].

6.5.2 Experimental setup

The experiment was conducted using the 640RC HoloVizio LFD, with six marked viewing distances on the floor of the laboratory: 1.39 m, 1.86 m, 2.32 m, 2.79 m, 3.25 m, and 3.72 m. Solving Equation 6.5 by substituting the lowest angular resolution used in the experiment, which is 1 degree, yields a maximum viewing distance (D_V) of 372 cm. Based on our earlier findings [273], the viewing distances primarily consisted of two. These included the distances measured as $0.5 \times D_V$ (1.86 m) and D_V (3.72 m). Then, three more distances were evenly distributed within the two aforementioned distances with a span of 0.47 m between any consecutive pair. Finally, an additional distance was marked in front of $0.5 \times D_V$.

Regarding the resolutions implemented in this study, two quality representations were utilized. One featured low resolutions in both spatial and angular aspects, with a spatial resolution of 640×480 and an angular resolution of 1 degree. The other featured high quality in both resolutions, with a spatial resolution of 1024×768 and an angular resolution of 0.5 degrees. It is important to highlight that the initial determination of D_V was derived from the minimum angular resolution utilized in the study, which amounted to 1 degree.

In a manner akin to our prior study [273] (discussed in Section 6.4), the same source materials were utilized, with the omission of two, to accommodate the augmented number of viewing distances. The source contents depicted in Figure 6.29 portray static 3D objects against a plain background. These contents encompass a simple model rendered as a set of cubes, complex mathematical objects, a textured model of a lighthouse, and laser-scanned statues.

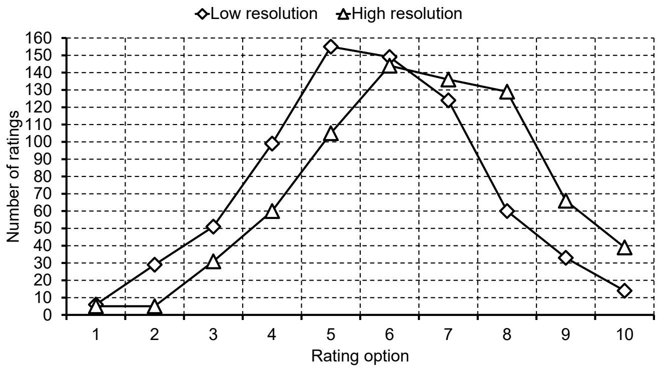


Figure 6.30: Rating distribution for low and high resolutions [274]

For this study, participants were instructed to evaluate the quality of visualized content using the ACR scale, which encompassed 10 rating options spanning from 1 (indicating the lowest quality) to 10 (indicating the highest quality). Assessments were conducted for both resolutions across each source content, at every designated viewing distance. Consequently, each participant contributed 72 assessment values, resulting from the combination of two resolutions, six viewing distances, and six source contents.

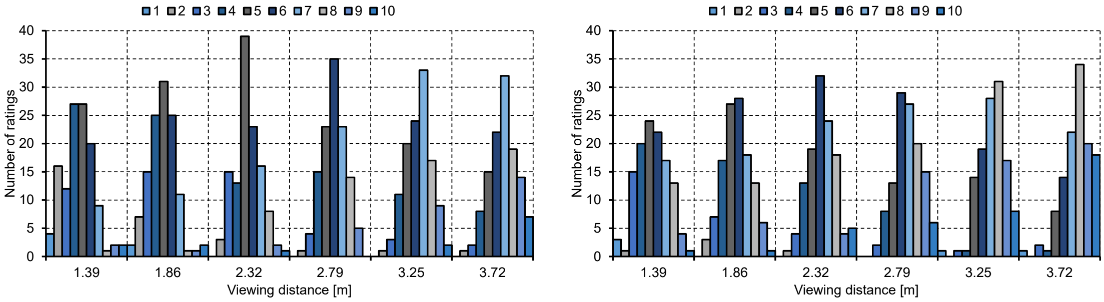
Concerning the test procedure, only one individual was permitted to participate at a time. The visual stimuli were presented to the participant in a randomized order. For each stimulus, the test participant was required to provide a rating at each of the six designated viewing distances.

The study targeted use cases involving static observation (e.g., cinematography). Therefore, participants were instructed to rate the visualized content from a central viewpoint, as the default viewing angle, without moving. However, minor head and body movements were permitted. This raised the question of having seated participants in the study to resemble the case of cinematography; however, this was discarded for numerous reasons. These reasons included the inability to assign a chair for each marked distance due to the small viewing distance intervals, and the additional requirement of a specific seating posture to avoid leaning backward or forward, which could affect the results.

The study enlisted 20 test participants, all of whom wore glasses with high diopter values. These participants comprised young adults, with an average age of 23, among whom 13 were male and 7 female. Additionally, the study incorporated one participant with a vision impairment exceeding 90%, who did not utilize corrective eyewear [274].

6.5.3 Results

A total of 1512 subjective ratings were registered by the 21 (20+1) participants. We commence our analysis by concentrating on the 20 test participants, followed by the participant with significant vision impairment.



(a) Low resolution

(b) High resolution

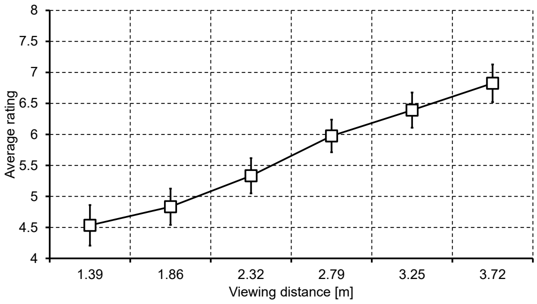
Figure 6.31: Rating distribution at the different viewing distances [274]

Figure 6.30 depicts a total of 1440 ratings recorded by the 20 test participants for both resolutions. As illustrated in the figure, both resolutions exhibit a similar distribution, with a discernible shift in ratings. The average ratings for the low and high resolution stimuli are 5.65 and 6.53, respectively.

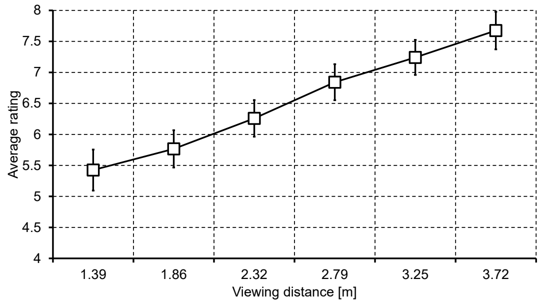
The depicted distribution suggests an effectively utilized rating scale by participants. As previously indicated, a 10-point ACR scale was selected over the 5-point scale defined by the ITU-T Rec. P.910 [13]. The rationale for this choice was the enhanced capacity to discern smaller perceptual differences. However, it's worth noting that both extremes of the scale might be underutilized. The psychological rationale for this conduct stems from the inclination to retain options within the ACR scale, allowing for the expression that particular test stimuli are either better or worse than those previously evaluated. This concept could apply to a subjective study where test stimuli are degraded to varying degrees and presented to participants in a randomized sequence. In this scenario, if a participant rates one stimulus as a 1, they may find it challenging to express if a subsequent stimulus is even worse than that. In this research endeavor, the evaluation task prioritized the consideration of viewing distance over quality degradation. As a result, participants were able to more effectively employ the scale, given the absence of necessity for reserving rating options.

A further notable observation is the overall consistency in the ratings. In our earlier research [273], we utilized six distinct combinations of resolution values, yielding subtle perceptible differences. Conversely, the present study exclusively employed the lowest and highest resolution combinations from the previous work, resulting in more pronounced perceptible differences within the test stimuli. This is clearly depicted in Figure 6.30, demonstrating a clear and sufficient separation between the two distributions. Consequently, rating consistency is not further examined in this study. Moreover, this observation is evident from both the rating distributions and the average ratings across the different viewing distances.

[13] ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications



(a) Low resolution



(b) High resolution

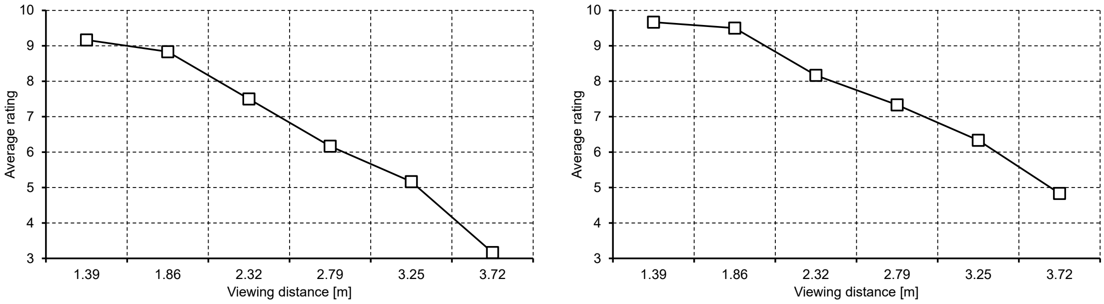
Figure 6.32: Average ratings at the different viewing distances [274]

Figures 6.31a and 6.31b depict the rating distribution recorded at the various viewing distances for the low and high resolutions, respectively. Each resolution received 120 ratings for every viewing distance. A conspicuous trend towards greater viewing distances was observed for both resolutions. Consistently, the closest distance received the lowest ratings, particularly evident with low resolution, which notably received the highest numbers of ratings 1 and 2 (4 and 16, respectively). Conversely, the farthest viewing distance attained the highest number of ratings 9 and 10 (20 and 18, respectively), presenting a notable contrast to the ratings assigned to the closest distance. While the majority of test participants exhibited a preference for greater distances, there were disparities in individual assessments: certain participants rated the closest viewing distance highly, while others assigned low ratings to the farthest viewing distance

Figures 6.32a and 6.32b illustrate the average ratings for low and high resolutions across the various viewing distances, respectively. The lack of overlap between the 0.95 confidence intervals signifies substantial distinctions among the examined viewing distances. Additionally, notable statistical differences are present between the two resolutions at each viewing distance. While these results bear resemblance to our prior research concerning distance-related preferences [273], statistical significance was attained in this study. This is largely attributed to the increased focus on the investigated subject within the experimental framework, coupled with the augmentation of the number of test participants. Indeed, as outlined in ITU-T Rec. BT.500 [14], a minimum of 15 test participants are required for conducting a QoE study of this nature. Although our prior study [273] did involve 15 test participants, they were subdivided into groups comprising 7 and 8 individuals. This division resulted in insufficient sample sizes within each group to achieve statistical significance. Furthermore, the subgroup of color-blind participants in that investigation did not exhibit a pronounced preference for greater distances.

Out of the 20 test participants, only two displayed interest in closer distances. Unlike the remaining 18 participants, whose average rating differences sometimes exceeded 5,

[14]ITU-T Recommendation BT.500: Methodologies for the subjective assessment of the quality of television images.



(a) Low resolution

(b) High resolution

Figure 6.33: Average ratings of the test participant with high vision loss at the different viewing distances [274]

theirs were within 1. Moreover, out of these two participants, only one distinctly favored closer distances, while the other evaluated 2.79 m with the highest scores, albeit with even smaller average rating differences.

As mentioned earlier, apart from the 20 test participants, an additional participant with severe vision loss (exceeding 90%) also completed the study. Figures 6.33a and 6.33b illustrate the average ratings for this particular test participant across the various viewing distances for the low and high resolutions, respectively. Unlike the previous results, this participant exhibited a distinct preference for closer viewing distances, likely due to vision impairment. The disparities in ratings extend to levels that equal or exceed those observed among the 18 other test participants (such as 8 for one of the laser-scanned statues). Additionally, the findings consistently indicate uniformity in terms of quality [274].

6.5.4 Conclusion and future work

This study aimed to investigate the preferred viewing distance for LF visualization among individuals with impaired visual acuity, in the case of static observation. The results suggest a preference for greater viewing distances. The acquired data aligns with previous findings [273] and exhibits less diversity in rating patterns compared to research on subjectively-preferred viewing distances [163], where participants underwent screening for normal vision using the Snellen chart for visual acuity and Ishihara plates for color vision. In this study, a maximum distance of 3.72 m was employed, determined by the 1-degree angular resolution of visualization. The results for the low and high resolutions exhibit statistically pronounced differences between the 1.86 m and 3.72 m viewing distances. For both resolutions, the ratings for these distances were significantly distinct from those at 2.79 m, further emphasizing the consistency in preference. Moreover, the study incorporated a participant with vision loss exceeding 90%. The findings for this individual diverged from those of the other participants, indicating a preference for closer viewing distances.

Table 6.4: Overview of the experimental factors, participant details, and key results for the QoE experiments on LF visualization

Experiment	Visualization factors				Participants		Results
	Viewing distance	Angular resolution	3D rendering	Spatial resolution	Impaired visual acuity	Color blind	
Experiment 1	✓	—	—	—	—	—	Section 6.2.3
Experiment 2	—	✓	✓	—	—	—	Section 6.3.3
Experiment 3	✓	✓	—	✓	✓	✓	Section 6.4.2
Experiment 4	✓	—	—	✓	✓	—	Section 6.5.3

In terms of future research, several key questions need to be explored. These include investigating the potential to surpass the theoretical limit of viewing distance and examining whether greater distances are preferred for visualization scenarios where objects appear more flat 2D than 3D. Additionally, the scope of the study should be expanded to include various observer motion models, such as sideways movement, as well as dynamic adjustments in viewing distance. The influence of unfavorable lighting conditions, encompassing distracting external light sources, requires examination. This is particularly important as individuals with diminished visual capabilities may exhibit distinct reactions in such situations. Finally, it is essential to investigate the correlation between the preferred viewing distance and the efficiency of interaction with LF systems, both in a general context and concerning individuals with reduced visual capabilities [274].

6.6 Conclusion

To wrap up this chapter, the experiments conducted to address the QoE for LF visualization have provided valuable insights into the factors affecting the perceived quality and user experience. These findings, gathered through subjective assessments across various contexts and visual acuity conditions, contribute to the ongoing efforts to enhance LFDs for both general and specialized applications. A summary of the experimental factors, participant details, and key results is provided in Table 6.4.

CHAPTER 7

Conclusion and future work

Among new capture and visualization technologies, LF has seen substantial progress, bringing it closer to practical, everyday usage. This technology has emerged as a method to represent the 3D world, acting as a window by filling the 3D space with light rays [42]. To view these captured LFs, LFDs were developed. Unlike many conventional 3D display systems, LFDs offer a full 3D experience without requiring viewers to wear any special devices. This advantage allows multiple people to view the display at once, making it possible for an almost unlimited number of users to interact with the content, depending only on the VVA of the display [117, 120]. Despite its potential, LF technology faces several challenges, which this study aims to address.

One key challenge is the difficulty of applying conventional cinematography techniques to LF cameras, particularly in wide-baseline systems. To address this, we developed a framework for evaluating camera motions on LFDs using virtual cameras. The study incorporated realistic physical motion formats, which were evaluated on a real LFD using multiple metrics. Results indicated that certain motion types were unsuitable for LFDs due to optical limitations, underscoring the importance of considering these constraints when designing camera movements for LFDs. Additionally, empirical studies on user preferences for different physical camera motions revealed that excessive oscillations or collisions led to a loss of focus on the LFD, with varied opinions on the effectiveness of these movements [119, 117, 123, 120].

Camera motion design for LFDs remains an unresolved research challenge, particularly in selecting optimal motions and addressing controversial movements that minimize visual issues. Further investigation is required, especially concerning the use of first-person camera views on LFDs. Several approaches can be explored to replicate first-person camera perspectives without introducing artifacts caused by the inherent limitations of LFDs [119, 117, 123, 120].

Building on the exploration of camera motion for LFDs, this research also examined interaction techniques for wide-baseline LFDs, a growing and promising field. After evaluating possible presentation models, the theater model was chosen for testing. User feedback indicated that LFDs generally offer a stronger sense of 3D immersion compared to traditional 2D displays, especially for specific content types. Participants also expressed that additional interactive features could further enhance this 3D perception. However, a key finding was an inverse relationship between interaction complexity and user mobility on LFDs: the less interactions and movements on the LFD, the better for users to walk around, and vice versa [118, 123].

As a continuation of this work, exploring additional presentation models on LFDs is recommended to further validate the feasibility of 3D interactions. The challenge of managing 3D interactions on LFDs, particularly for the application/system control task, remains an open research question. Future work could investigate new approaches for providing application/system control feedback on LFDs, aiming to display the GUI and offer feedback without disrupting the 3D immersion [118, 123].

We then explored the combination of HDR and LF technologies, highlighting their significant potential across various applications. To achieve HDR LF imaging from legacy LDR LF images, we reviewed existing CNN models for LDR-to-HDR reconstruction and tested three models –*ExpandNet*, *HDR-DeepCNN*, and *DeepHDRVideo*– using the *Teddy* LF dataset. The results revealed that *HDR-DeepCNN* outperformed the other models in terms of image quality, particularly when evaluated with HDR-VDP. In contrast, video-based methods, which were expected to benefit from temporal coherence, did not perform as well as anticipated. Finally, we introduced a new dataset specifically designed for HDR LF applications, supporting various LF systems with three configurations: narrow-baseline FP, narrow-baseline HOP, and wide-baseline HOP. This dataset extends the available resources for future HDR LF research and applications [124, 116, 125, 122].

Although we have explored various applications of HDR LF imaging, further research is needed to address the unique requirements of each use case. Regarding LDR-to-HDR LF reconstruction, applying CNNs across multiple LF images, rather than single-image approaches, yields better results by leveraging spatial coherence and angular information. To improve HDR LF reconstruction, future research should test various CNN models, develop CNNs specific to LF imaging, and create HDR LF datasets for training. In our work, we began with a simple dataset –simulating a classroom scene– to facilitate CNN training and testing, with plans to gradually increase scene complexity as reconstruction techniques improve. Additionally, future efforts could include developing a dataset for arc systems by rendering images from multiple orientations in *MAYA*, and exploring methods to capture real-world HDR LF content to enhance the practical applicability of these datasets [124, 116, 125, 122].

Lastly, to enhance the user experience of LF visualization, we conducted a series of experiments across multiple LFDs to identify the factors that shape the overall visual experience. These experiments not only explored general aspects of LF visualization but also delved into specific use cases, involving participants with both normal and reduced visual capabilities. One of the key unresolved questions in LF research is the optimal viewing distance, which prompted our first experiment to assess both perceptually-supported and subjectively-preferred viewing distances. As our investigation progressed, we discovered that factors such as angular resolution and 3D rendering significantly influence the quality of visualized models, particularly complex ones. This finding led to our second experiment, which focused on examining their impact in industrial contexts. These two factors are intricately connected and, therefore, require joint consideration. With the growing trend of vision deterioration among younger generations, the final two experiments addressed the factors affecting LF visualization from the perspective of individuals with impaired visual acuity, including those with color blindness and a participant with over 90% vision impairment. Through these experiments, we gained valuable insights into how to optimize LF visualization for diverse users and contexts [163, 164, 273, 274].

Building on the findings of the experiments, several future research directions emerge. For the first experiment, which focused on viewing distances, future work should examine the motion of observers across various usage scenarios, as well as examining the extent of head and body sways during static observations [163]. Regarding the second experiment, which investigated angular resolution and 3D rendering in industrial contexts, future studies should explore various use case contexts, considering their unique characteristics and attributes [164]. For the final two experiments, which involved participants with reduced visual acuity, future research should address the long-term utilization of LFDs, the potential for perceptual fatigue, and how individuals with visual impairments respond to lighting conditions. This latter issue is particularly important, as individuals with reduced visual capabilities may react differently in such situations. Furthermore, several key questions remain to be explored, including the potential to surpass the theoretical limit of viewing distance and whether greater distances are preferred in scenarios where objects appear more flat 2D than 3D. Additionally, the scope of the study should be expanded to include various observer motion models, such as sideways movement, and dynamic adjustments in viewing distance [273, 274].

In conclusion, this thesis provides a comprehensive exploration of LF technology, from its foundational principles to its most advanced applications. Through a series of novel contributions, the research examines critical aspects of LF visualization, including camera animation, interactive UIs, and HDR LF imaging, while also addressing the subjective QoE across different LFDs. By integrating both theoretical and practical elements, the work advances the understanding of how LF technology can be optimized for various use cases, with particular attention to enhancing accessibility for users with diverse visual capabilities. The findings not only offer valuable insights into the current state of LF technology but also pave the way for future developments, guiding the field toward more effective, high-quality, and user-centered applications.

CHAPTER 8

New scientific contributions

The primary scientific contributions of this dissertation are articulated in the following key theses:

Thesis 1 *Light field camera animation* **(Chapter 3)**

Related publications: [119, 123, 117, 120]

To advance the study of camera animation in LF visualization, I designed and developed a novel simulation framework that uniquely incorporates the properties of LF cameras, rigorously testing it on a real LFD. Through this framework, I established a foundation for LF camera animations –an under-explored area– by developing and testing various virtual camera animations on a HOP LFD, namely the HoloVizio C80 cinema system, thereby paving the way for future research.

Main features of this work are as follows:

- This new and original framework, built using Holografika's clustered rendering modules, is the first to support both lenticular and projection-based displays while utilizing a GPU cluster for real-time, multi-view rendering optimized for HoloVizio LFDs.
- This simulation framework enables real-time rendering of diverse scenarios, simulating physical environments and common camera movements used in film production.
- I integrated path planning for wide-baseline LF cameras and physical camera simulation, allowing users to set key parameters like speed, mass, and acceleration for camera movements.

- Optical and sensor properties are automatically aligned with the LFD to ensure seamless compatibility between LF cameras and displays without the need for additional conversions.
- Through the framework, I addressed the challenge of matching captured LFs with those of the LFDs by means of virtual LF cameras, with findings applicable to physical LF cameras with comparable baselines. I managed the camera movement by means of ROI matrix, where display rays were evaluated and transformed into world space coordinates, making it easier to render objects and lights within the same system.
- I defined the capture surface of the LF camera by determining sensor positions per pixel and tessellating a flat surface among neighboring points. I devised error metrics to evaluate system performance on LFDs by using a 4×4 affine transformation (ROI) to align observer and capture planes.
- I created realistic simulation environments using the Bullet Physics Library [65] to model physical scenes with basic shapes. The framework tested various camera animation scenarios by adjusting parameters like weight, size, and motion for both cameras and scene objects. This helped generate diverse scenarios to assess the effectiveness of different camera movements for LF camera simulations.
- I implemented different camera animations to establish the foundation of LF camera animations for LF visualization. Camera animations included cinematography camera animations including pan, tilt, zoom, dolly, truck, and pedestal. Additionally, I created simulation camera animations for both first- and third-person perspectives. I also developed three physical scenarios to simulate collision, falling, and suspension cameras.
- I developed a set of criteria to assess different aspects of camera animation including general visibility of the scene along the observer's line during animations, the frequency of immersion-breaking occluders, collision occurrences, depth-related artifacts, and changes in the depth of field. Based on the results of the expert assessments, I identified which LF camera animations are suitable for LFDs and which require further investigation.
- Results of perceptual assessments indicated that pan, tilt, truck, and pedestal camera movements produced clear outputs, while dolly and zoom movements caused blurriness. First-person camera simulations also showed artifacts, while third-person camera animations were more reliable. These findings pave the way for future LF camera animations, highlighting effective camera movements and areas for refinement to enhance visual quality and user experience.
- To evaluate the plausibility of the generated physical simulations (i.e. collision, falling, and suspension cameras), I devised several objective metrics to be measured, designed for HOP LFDs:
  - Camera collision metric: counts the number of intersections between the AABBs of the objects in the scene and the AABB of the camera.

  - Blurriness metric: measures the number of blurry objects in the scene by counting the intersections between the objects' AABBs and the frustum defining the blurry region of the LFD.
  - Occlusion metric: used in case of third-person cameras.

- I conducted subjective tests to further evaluate the plausibility of the realistic physical simulations. The results showed that 76.2% of participants preferred third-person cameras on LFDs due to the blurriness and discomfort caused by first-person cameras, which also led to dizziness and focus loss, indicating the need for further research.
- A key finding in the subjective assessment was the inverse relationship between participant movement and camera motion. Evaluations revealed that increased camera motion resulted in more occlusions, blurriness, and collisions, which reduced visual quality. Based on these findings, slight camera movements are recommended for LFDs.
- Beyond the implemented framework, I theoretically explored the development and assessment of LF camera animation techniques, analyzing their implications, limitations, potential applications, and directions for future research from the perspectives of use cases, visual content, quality assessment, and capture and display hardware.

Thesis 2 *Interaction techniques for light field displays* **(Chapter 4)**

Related publications: [118, 123]

In order to test different interaction methods on LFDs, which have thus far only seen the development of basic UIs, I first analyzed the challenges imposed by LFDs for each of the 3D interaction tasks (i.e., navigation, selection and manipulation, and application/system control). Then, I proposed several presentation models for LFDs, including line-up, carousel, 3D sphere, CAD/CAM, medical, and theater model, where the latter was chosen.

I implemented a theater model using *MAYA*[1], and visualized it on the HoloVizio C80 LFD. The theater model was selected because it parallels LFDs, allowing multiple viewers to observe content simultaneously in an angularly-dependent manner. Considering the capabilities and limitations of LFDs, I analyzed and modified the three interaction tasks involved in 3D environments as follows:

- **Navigation** in LFDs, due to their multi-camera setups, presents unique challenges that require modifications to the observer line/rectangle, for precise adjustments. To address this, I implemented a static camera configuration designed to meet these requirements within the theater model.
- In the theater model, I implemented several selection and manipulation techniques, including a rotating stage positioned in the sharp region of the LFD to prevent blurriness during movement. I also animated objects along designated paths and used curtains to hide/reveal elements. To avoid transitioning into blurry regions, presentation elements were positioned on a plane parallel to the screen (e.g., animating curtains and flying systems). Additionally, I employed rotating stages with one half in the sharp region and animated spotlights within a limited range to minimize LFD issues.
- **Application/system control** on LFDs is challenging, as overlay rendering relies on image space, which disrupts the 3D depth perception essential to LFDs. I proposed several possible solutions including rendering the UI into 2D areas, akin to selection methods, or spatially separating 3D controls from the main scene to provide scene feedback on the control geometry. In my work, I implemented a monitor room to provide high-quality visual feedback. View switching is triggered by pressing buttons, which activate corresponding animations and lighting in the theater model. The monitor room displays the current view, and after activation, navigation resumes through a static camera within the theater model.

I conducted subjective assessments of the three implemented theater scenarios to gather feedback data, which is crucial for the long-term development of such applications. The following summarizes the novel findings for each interaction task:

[1] https://www.autodesk.com/products/maya

- I evaluated user preferences for the **navigation** task, finding that the majority of participants favored a static camera. This preference appeared to enhance the 3D effect of the LFDs, with further improvement achieved by allowing users to move around the screen. My findings suggest that static cameras are effective for navigation tasks, as they reduce discomfort while preserving immersion.
- I assessed interaction models for **selection and manipulation** on the LFD, finding a strong preference for the “multiple carousels” model, along with positive responses to “curtain” and “flying system” motions and backstage theater scenes. These findings indicate a clear preference among participants for highly interactive methods on LFDs, with increased interest in moving around the display for better immersion. Overall, participants favored interaction techniques on LFDs over traditional 2D displays and expressed a desire for more advanced interactive features.
- I assessed user preferences for the **application/system control task** and found that most participants preferred buttons within the main scene, although this could disrupt 3D immersion. This highlights an ongoing challenge in providing effective feedback for 3D scenes, offering insights for future immersive system design.

Finally, subjective evaluation revealed an inverse relationship between the level of interaction on LFDs and participant mobility.

Thesis 3 *Towards HDR light field imaging* **(Chapter 5)**

Related publications: [**124**, **125**, **116**, **122**]

In this thesis, I integrate both HDR technology and multi-autostereoscopic systems, such as LFDs, to achieve powerful and impactful results, while also examining the potential challenges. HDR technology enhances the realism of visual content, while multi-autostereoscopic systems deliver immersive 3D experiences without the need for specialized viewing equipment.

To achieve HDR LF imaging, the following steps were undertaken:

- I carried out a comprehensive analysis of HDR LF imaging applications and explored future use cases with substantial practical potential. Key applications examined include physically-based rendering, digital photography, image editing, cinematography, various medical use cases, cultural heritage, education, digital signage, and telepresence.
- Reconstructing HDR LF content from LDR LF images poses challenges but can yield higher-quality outputs, as scene information is encoded across multiple images. In my work, I investigated the theoretical possibilities of combining CNN architectures utilized for HDR images and videos, in order to enhance the outputs of HDR LF image reconstruction.
- As a starting point for LDR-to-HDR LF reconstruction research, I tested several HDR reconstruction CNNs on the *Teddy* LF image dataset [127]. The insights gained from the output images have provided valuable guidelines for developing CNNs for HDR LF image reconstruction.
  - I found that *ExpandNet* [220] produced visually plausible images, though it introduced ghosting artifacts in the background. This suggests that integrating concatenated feature branches could improve the model's adaptability to various datasets.
  - I discovered that *HDR-DeepCNN* [94] exhibited color inconsistencies, likely due to skip connections involving domain transformations from LDR display values to logarithmic HDR.
  - I observed that *DeepHDRVideo* [57] exhibited visible artifacts in shape and texture, which can be attributed to alignment errors in optical flow.
- I evaluated the performance of the CNNs using three objective metrics: (i) PSNR, (ii) SSIM, and (iii) HDR-VDP. The following findings were observed:
  - Results showed that *DeepHDRVideo* achieved the highest PSNR and SSIM scores, while *HDR-DeepCNN* excelled in HDR-VDP scores, better aligning with the HVS. This was reflected in the reconstructed HDR images, which exhibited superior consistency and visual quality.

  - Although video reconstruction techniques were expected to perform well by leveraging temporal coherence –analogous to spatial coherence in LF images, results show that *HDR-DeepCNN* ultimately delivered more convincing quality results.
  - These findings highlight the need for developing more HDR LF datasets and creating quality metrics tailored to evaluate the unique characteristics of LF imaging.

- I developed a synthetic HDR LF dataset called "CLASSROOM" to address the limited availability of such datasets for CNN training and testing. This dataset allows manipulation of various parameters and scene complexity, supporting the creation of additional datasets. It is not limited to a specific baseline or parallax, enabling the generation of datasets with varying configurations, thus advancing the field of HDR LF reconstruction. I created the "CLASSROOM" dataset using *MAYA* 2022 and rendered it with the Arnold renderer, considering both narrow- and wide-baseline systems. I created the following datasets:
  - A narrow-baseline FP dataset with 5×5 images.
  - A narrow-baseline HOP dataset, a subset of the first with selectable rows.
  - A wide-baseline HOP dataset with 15 images.

  I calculated the inter-image distance based on the FOV of the LFD, the number of images, and the distance between the display and observer's line/rectangle. To create the narrow- and wide-baseline datasets, I adjusted the camera's focal length to 35 mm and 20 mm, respectively.

Thesis 4 *Quality of experience for light field visualization* **(Chapter 6)**

Related publications: [**163**, **159**, **274**, **164**, **273**]

This thesis incorporates subjective studies that evaluate a range of factors impacting the visual experience on LFDs, both broadly and within specific use cases, involving participants with both normal and reduced visual capabilities. In these experiments, I rendered the content on the LFDs and conducted the experiments.

Experiment 1: Regarding LFDs, the optimal viewing distance remains an open research question. Building on the findings by Kara *et al.* [169], the study investigates both **perceptually-supported and subjectively-preferred viewing distances for LF visualization**, conducted on the HoloVizio 80WLT LFD and HoloVizio C80 cinema system.

- I used the Holo Qt Converter to render content for the perceptually-supported viewing distance experiment and Holografika's clustered renderer for the subjectively-preferred experiment, generating ten source contents. I conducted each experiment twice, once with experts and once with 22 regular participants.
- The perceptually-supported viewing distance experiment showed that experts preferred distances between 4 m and 5.75 m, while non-experts favored 3.5 m to 6.75 m. Although some outliers existed, their subjectively-preferred viewing distances aligned with other participants.
- Outliers were observed to be taller than other participants, which impacted the results due to the larger horizontal displacement at their eye level. To account for this, the maximum viewing distance threshold for LFDs is recalculated as $D_V = \frac{D_E + D_S}{tan(AR)}$, where D_S accounts for the horizontal displacement from participant swaying.

Experiment 2: For complex models, angular resolution plays a critical role, as insufficient resolution can result in crosstalk, while higher resolution may improve detail. On the other hand, deeper 3D rendering can still lead to blurriness. The interconnection between these factors highlights the need for careful optimization to achieve the best visualization quality. Therefore, this experiment investigates the **effect of angular resolution and 3D rendering on the perceived quality of content in LF visualization for industrial contexts**, particularly for prototype evaluation, given the complexity of industrial models.

- I conducted the experiment on the HoloVizio HV640RC LFD. I rendered 7 different static industrial objects at 7 angular resolutions (ranging from 0.5 to 2 degrees), with a fixed spatial resolution of 1024×768. The experiment used the hidden reference method and involved 43 participants.

- The results showed that source contents with greater depth variations were more affected by reduced angular resolution. Minor inconsistencies in similar test conditions were noted but had little impact.
- Overall, both angular resolution and 3D rendering significantly influenced the QoE, with quality ratings being directly linked to the classification based on the depth of the source content.

Experiment 3: As visual impairments become more prevalent among younger individuals, understanding how both unimpaired and impaired individuals perceive LF visualization quality is crucial. This study presents our preliminary investigation into **LF visualization as evaluated by participants with imperfect visual acuity and color blindness**.

- I conducted the experiment on the HoloVizio HV640RC LFD to examine various factors influencing LF visualization, with two participant groups: Group 1 consisting of 8 participants with impaired visual acuity and Group 2 consisting of 7 participants with color blindness.
- I rendered 8 static scenes with varying complexity, depth, textures, and structures, across 12 test conditions defined by 2 spatial resolutions (640×480 and 1024×768), 3 angular resolutions (1°, 0.66°, and 0.5°), and 2 viewing distances (1.86 m and 3.72 m).
- Results showed that viewing distance significantly impact perceived quality, with closer distances highlighting impairments, particularly blurriness from low spatial resolution. Statistically significant differences in spatial resolution were observed across all angular resolutions at the closer distance.
- Group 1 and Group 2 showed similar rating tendencies, but color-blind participants generally gave lower scores, especially at closer distances. In Group 1, angular resolution had a greater impact, while for Group 2, viewing distance was more influential.
- Rating inconsistencies, mostly related to angular resolution and content with minimal depth variation, were more frequent among color-blind participants. Diopter values did not significantly affect the rating inconsistencies.

Experiment 4: This study investigated the **preferred viewing distance for LF visualization among individuals with impaired visual acuity during static observation.**

- I used the same source material as in the previous experiment, excluding two scenes, to account for the large number of viewing distances. A total of 21 participants took part: 20 with high diopter glasses and one with more than 90% vision impairment. Six viewing distances were marked: 1.39 m, 1.86 m, 2.32 m, 2.79 m, 3.25 m, and 3.72 m. The study employed two quality settings: low resolution (640×480 spatial, 1° angular) and high resolution (1024×768 spatial, 0.5° angular).

- Results indicate a strong preference for greater viewing distances across both resolutions, with closer distances receiving lower ratings, especially at low resolution.
- Notably, the participant with over 90% vision impairment preferred closer distances, likely due to his/her impaired vision, which contrasted with the general trend observed in other participants.

Overview of the subjective studies: The experiments on QoE for LF visualization provided valuable insights into factors affecting perceived quality and user experience, contributing to improvements for general and specialized applications. Table 8.1 provides an overview of the factors and participants involved in each experiment.

Table 8.1: Overview of the experiments addressing the QoE for LF visualization

Experiment	**Visualization factors**				**Participants**	
	Viewing distance	**Angular resolution**	**3D rendering**	**Spatial resolution**	**Impaired visual acuity**	**Color blind**
Experiment 1	✓	—	—	—	—	—
Experiment 2	—	✓	✓	—	—	—
Experiment 3	✓	✓	—	✓	✓	✓
Experiment 4	✓	—	—	✓	✓	—

List of Figures

List of Tables

Acronyms

AABB Axis-Aligned Bounding Box. 44–46, 58, 60, 136, 137

ACR Absolute Category Rating. 48, 64, 120, 127, 128

AFR Adaptive Feature Remixing. 17

AR Augmented Reality. 2, 19, 20, 25, 36, 53

CAVE Cave Automatic Virtual Environment. 19

CNN Convolutional Neural Network. 2, 16, 17, 69–74, 76, 83, 85, 87, 93, 94, 133, 140, 141

CRF Camera Response Function. 72

CT Computed Tomography. 80

CTU Coding Tree Unit. 28

DCR Degradation Category Rating. 110, 112, 113, 115–117, 147

DOF Degrees Of Freedom. 19, 55

EPI Epipolar Plane Image. 17

EXR EXtended Range format. 71, 91, 93

FALL Frame Average Light Level. 77, 78

FOV Field Of View. 7–9, 11, 19, 22–25, 34, 37, 54, 56, 60, 83, 88, 89, 91, 95, 97, 98, 101, 109, 110, 118, 119, 141

FP Full Parallax. viii, 2, 8, 11, 12, 15, 23, 55, 56, 70, 88–93, 133, 141, 146

FSW Full-Screen White. 77

GAN Generative Adversarial Network. 28

GUI Graphical User Interface. 64, 66, 67, 133

HCI Human-Computer Interaction. 95, 125

HDR High Dynamic Range. 2, 3, 68–89, 93, 94, 133, 134, 140, 141, 146

HDRI HDR Images. 88

HLRA Homography-based Low-Rank Approximation. 27

HMD Head-Mounted Display. 19, 20, 24, 53

HOP Horizontal-Only-Parallax. viii, 2, 8, 9, 11, 12, 15, 22, 23, 25, 26, 37, 44, 45, 47, 55, 56, 64, 70, 88, 89, 91–93, 97–99, 133, 135, 136, 141, 146

HVS Human Visual System. 2, 4, 6, 28, 69, 72, 75, 78, 80, 86, 93, 140

IBL Image-Based Lighting. 75, 81, 88

ILM Industrial Light and Magics. 71

IQMs Image Quality Metrics. 86

ITM Inverse Tone Mapping. 78

KPI Key Performance Indicator. 22, 81, 98, 119, 125

LCD Liquid Crystal Display. 80, 82

LDR Low Dynamic Range. 2, 68–73, 75, 76, 79, 80, 82, 83, 85–87, 93, 94, 133, 140, 146

LED Light-Emitting Diode. 22, 30, 82

LF Light Field. 1–9, 11–17, 21–31, 36–41, 44, 47, 52, 54, 56, 58, 60, 64, 69, 70, 75, 76, 81, 83–89, 93–100, 102, 108, 109, 118, 119, 124, 125, 130–137, 140–146, 148

LFD Light Field Display. 1–3, 5, 9–11, 20–27, 30, 31, 36–67, 74, 76, 78, 80–83, 88, 95–100, 102, 103, 107–109, 112, 115, 119, 125, 126, 131–143, 145, 146, 148

MIS Minimally Invasive Surgery. 80

MLA MicroLens Array. 13–15, 145

MOS Mean Opinion Scores. 120–122, 147

MPI MultiPlane Image. 28

MR Mixed Reality. 2, 53

MRI Magnetic Resonance Imaging. 80

POV Point-Of-View. 33, 34

PSNR Peak-Signal-to-Noise-Ratio. 70, 84, 86, 140, 146

QoE Quality of Experience. 3, 22, 23, 26, 27, 95, 97, 109, 115, 118, 124, 125, 129, 131, 134, 143, 144, 146, 148

RGB Red Green Blue. 2, 71, 72

ROI Region Of Interest. 11, 21, 37–40, 44, 49, 56, 95, 109, 136

S3D Stereoscopic 3D. 19, 20, 36

SADN Spatial-Angular-Decorrelated Network. 28

SAI Sub-Aperture Image. 28

SDR Standard Dynamic Range. 68

SLM Spatial Light Modulator. 77

SOP State Of Polarization. 19

SSIM Structural Similarity Index Measure. 70, 84, 86, 140, 146

TIFF Tagged Image File Format. 71

TM Tone Mapping. 78, 80

UHD Ultra-High Definition. 82

UI User Interface. 2, 3, 53–55, 59, 134, 138

VDP Visible Difference Predictor. 70, 84, 86, 93, 133, 140, 146

VOP Vertical-Only-Parallax. 11, 23

VR Virtual Reality. 2, 19, 20, 22, 36, 53, 82

VVA Valid Viewing Area. 1, 8, 22, 23, 25, 47, 48, 64, 81, 132

Bibliography

[1] Autodesk Arnold, tech. rep., https://damassets.autodesk.net/content/dam/autodesk/www/pdfs/arnold-features.pdf.

[2] ELF-SR1 Spatial Reality Display - Sony Pro. https://pro.sony/ue_US/products/spatial-reality-displays/elf-sr1#TEME502131AllYouNeedIsYourEyes-elf-sr1. [Online; accessed 26-07-2024].

[3] ELF-SR2 Spatial Reality Display - Sony Pro. https://pro.sony/ue_US/products/spatial-reality-displays/elf-sr2. [Online; accessed 26-07-2024].

[4] HoloVizio 722RC Large-scale 3D displaying. https://holografika.com/722rc/. [Online; accessed 26-07-2024].

[5] HoloVizio 80WLT Full-angle 3D displaying. https://holografika.com/80wlt/. [Online; accessed 26-07-2024].

[6] HoloVizio C80 glasses-free 3D cinema system. https://holografika.com/c80-glasses-free-3d-cinema/. [Online; accessed 26-07-2024].

[7] Looking Glass 16" Spatial Display. https://lookingglassfactory.com/16-spatial-oled. [Online; accessed 26-07-2024].

[8] Looking Glass 32" Spatial Display. https://lookingglassfactory.com/looking-glass-32. [Online; accessed 26-07-2024].

[9] Looking Glass 65". https://lookingglassfactory.com/looking-glass-65. [Online; accessed 26-07-2024].

[10] Looking Glass Go. https://lookingglassfactory.com/looking-glass-go. [Online; accessed 26-07-2024].

[11] Looking Glass Portrait. https://lookingglassfactory.com/looking-glass-portrait. [Online; accessed 26-07-2024].

[12] Lume Pad 2. https://www.leiainc.com/lume-pad-2. [Online; accessed 26-07-2024].

[13] Raytrix: 3D Light Field Vision. https://raytrix.de/. [Online; accessed 26-07-2024].

[14] Static Planar Light-Field Test Dataset. https://www.iis.fraunhofer.de/en/ff/amm/dl/lightfielddataset.html. [Online; accessed 26-07-2024].

[15] The Plenoptic 2.0 Toolbox: Benchmarking of Depth Estimation Methods for MLA-Based Focused Plenoptic Cameras. https://zenodo.org/records/3558284#.YeXpMHrP2Hs. [Online; accessed 26-07-2024].

[16] Why samsung LED displays are a great fit for the retail market. https://displaysolutions.samsung.com/pdf/leaflet/3348/Samsung_CM1711_LED_Eprint.pdf. [Online; accessed 26-07-2023].

[17] Edward H. Adelson and James R. Bergen. *The plenoptic function and the elements of early vision*, volume 2. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

[18] Edward H. Adelson and John Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (2):99–106, 1992.

[19] Vamsi K. Adhikarla, Grega Jakus, and Jaka Sodnik. Design and evaluation of freehand gesture interaction for light field display. In *Human-Computer Interaction: Interaction Technologies: 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part II 17*, pages 54–65. Springer, 2015.

[20] Vamsi K. Adhikarla, Jaka Sodnik, Peter Szolgay, and Grega Jakus. Exploring direct 3D interaction for full horizontal parallax light field displays using leap motion controller. *Sensors*, 15(4):8642–8663, 2015.

[21] Vamsi K. Adhikarla, Marek Vinkler, Denis Sumin, Rafal K. Mantiuk, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. Towards a quality metric for dense light fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 58–67, 2017.

[22] Amar Aggoun. A 3D DCT compression algorithm for omnidirectional integral images. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 2. IEEE, 2006.

[23] Amar Aggoun. Compression of 3D integral images using 3D wavelet transform. *Journal of Display Technology*, 7(11):586–592, 2011.

[24] Marco Agus, Enrico Gobbetti, José Iglesias Guitian, Fabio Marton, and Giovanni Pintore. GPU Accelerated Direct Volume Rendering on an Interactive Light Field Display. *Computer Graphics Forum*, 27(2):231–240, 2008.

[25] Ayyoub Ahar, Maksymilian Chlipala, Tobias Birnbaum, Weronika Zaperty, Athanasia Symeonidou, Tomasz Kozacki, Malgorzata Kujawinska, and Peter Schelkens. Suitability analysis of holographic vs light field and 2D displays for subjective quality assessment of Fourier holograms. *Optics Express*, 28(24):37069–37091, 2020.

[26] Waqas Ahmad, Luca Palmieri, Reinhard Koch, and Märten Sjöström. Matching light field datasets from plenoptic cameras 1.0 and 2.0. In *2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2018.

[27] Martin Alain and Aljosa Smolic. Light field super-resolution via LFBM5D sparse coding. In *2018 25th IEEE international conference on image processing (ICIP)*, pages 2501–2505. IEEE, 2018.

[28] M. Zeshan Alam and Bahadir K. Gunturk. Hybrid light field imaging for improved spatial resolution and depth range. *Machine Vision and Applications*, 29(1):11–22, 2018.

[29] Luigi Albani, Lode De Paepe, Silvio Bonfiglio, Gabriele Guarnieri, and Giovanni Ramponi. HDR medical display based on dual layer LCD. In *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 511–515. IEEE, 2013.

[30] Hadi Amirpour, Raimund Schatz, Christian Timmerer, and Mohammad Ghanbari. On the impact of viewing distance on perceived video quality. In *2021 International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–5. IEEE, 2021.

[31] Hirofumi Aoki, Charles M Oman, Daniel A Buckland, and Alan Natapoff. Desktop-VR system for preflight 3D navigation training. *Acta astronautica*, 63(7):841–847, 2008.

[32] Darmont Arnaud. *High Dynamic Range Imaging: Sensors and Architectures.* SPIE Digital Library. SPIE Press, 2012.

[33] David A. Atchison, Emma L. Markwell, Sanjeev Kasthurirangan, James M. Pope, George Smith, and Peter G. Swann. Age-related changes in optical and biometric characteristics of emmetropic eyes. *Journal of vision*, 8(4):29–29, 2008.

[34] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001.

[35] Ronald T. Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997.

[36] A. Balázs, A. Barsi, P. T. Kovács, and T. Balogh. Towards mixed reality applications on light-field displays. In *3DTV Conference*, pages 1–4. IEEE, 2014.

[37] Anders Ballestad, Ronan Boitard, Gerwin Damberg, and Goran Stojmenovik. Advances in HDR Display Technology for Cinema Applications, Including Light-Steering Projection. *Information Display*, 35(3):16–19, 2019.

[38] Tibor Balogh. 3D Light Field: Displaying, Formats, End-to-end Systems, Jun 2019.

[39] Tibor Balogh, Attila Barsi, Peter A. Kara, Mary Guindy, Aniko Simon, and Zsolt Nagy. 3D light field LED wall. In *Digital Optical Technologies 2021*, volume 11788, pages 1–11. SPIE, 2021.

[40] Tibor Balogh and Péter Tamás Kovács. Real-time 3D light field transmission. In *Real-Time Image and Video Processing 2010*, volume 7724, pages 53–59. SPIE, 2010.

[41] Tibor Balogh, Péter Tamás Kovács, and Attila Barsi. Holovizio 3D display system. In *3DTV Conference*, pages 1–4. IEEE, 2007.

[42] Nikhil Balram and Ivana Tošić. Light-field imaging and display systems. *Information Display*, 32(4):6–13, 2016.

[43] Martin S. Banks, David M. Hoffman, Joohwan Kim, and Gordon Wetzstein. 3D Displays. *Annual review of vision science*, 2:397–435, 2016.

[44] Michael Barz, Sebastian Kapp, Jochen Kuhn, and Daniel Sonntag. Automatic recognition and augmentation of attended objects in real-time using eye tracking and a head-mounted display. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–4. Association for Computing Machinery, 2021.

[45] John Bercovitz. Image-side perspective and stereoscopy. In *Stereoscopic Displays and Virtual Reality Systems V*, volume 3295, pages 288–298. SPIE, 1998.

[46] Tom E. Bishop and Paolo Favaro. The light field camera: Extended depth of field, aliasing, and superresolution. *IEEE transactions on pattern analysis and machine intelligence*, 34(5):972–986, 2011.

[47] Christoffer Björkskog, Giulio Jacucci, Bruno Lorentin, and Luciano Gamberini. Mobile Implementation of a Web 3D Carousel with Touch Input. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Association for Computing Machinery, 2009.

[48] Pierre-Alexandre Blanche. Holography, and the future of 3D display. *Light: Advanced Manufacturing*, 2(4):446–459, 2021.

[49] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. An introduction to 3-D user interface design. *Presence*, 10(1):96–108, 2001.

[50] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[51] Jorge Perez Bravo. 25-3: HDR Solution; Dynamic Drive on LED Video Screens. In *SID Symposium Digest of Technical Papers*, volume 49, pages 322–323. Wiley Online Library, 2018.

[52] Philip Brey. Virtual reality and computer simulation. *The handbook of information and computer ethics*, pages 361–384, 2008.

[53] Blain Brown. *Cinematography: theory and practice: image making for cinematographers and directors.* Focal Press, 2016.

[54] Andreas Buehler. Third Person Camera View in Games - a record of the most common problems in modern games, solutions taken from new and retro games. https://www.gamedeveloper.com/design/third-person-camera-view-in-games---a-record-of-the-most-common-problems-in-modern-games-solutions-taken-from-new-and-retro-games, 2019. [Online; accessed 04-03-2021].

[55] Ernest Callenbach. The Five C's of Cinematography: Motion Picture Filming Techniques Simplified by Joseph V. Mascelli, 1966.

[56] Chuo-Ling Chang, Xiaoqing Zhu, Prashant Ramanathan, and Bernd Girod. Light field compression using disparity-compensated lifting and shape adaptation. *IEEE transactions on image processing*, 15(4):793–806, 2006.

[57] Guanying Chen, Chaofeng Chen, Shi Guo, Zhetong Liang, Kwan-Yee Kenneth Wong, and Lei Zhang. HDR Video Reconstruction: A Coarse-to-fine Network and A Real-world Benchmark Dataset. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2482–2491, 2021.

[58] Jie Chen, Junhui Hou, and Lap-Pui Chau. Light field compression with disparity-guided sparse coding based on structural key views. *IEEE Transactions on Image Processing*, 27(1):314–324, 2017.

[59] Wei-Chao Chen, Jean-Yves Bouguet, Michael H Chu, and Radek Grzeszczuk. Light field mapping: Efficient representation and hardware rendering of surface light fields. *ACM Transactions on Graphics (TOG)*, 21(3):447–456, 2002.

[60] Yilei Chen, Ping An, Xinpeng Huang, Chao Yang, Deyang Liu, and Qiang Wu. Light field compression using global multiplane representation and two-step prediction. *IEEE Signal Processing Letters*, 27:1135–1139, 2020.

[61] Wei-Chung Cheng and Aldo Badano. A gaze-contingent high-dynamic range display for medical imaging applications. In *Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment*, volume 7627. International Society for Optics and Photonics, 2010.

[62] Caroline Conti, Péter Tamás Kovács, Tibor Balogh, Paulo Nunes, and Luís Ducla Soares. Light-field video coding using geometry-based disparity compensation. In *2014 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2014.

[63] Caroline Conti, Paulo Nunes, and Luís Ducla Soares. HEVC-based light field image coding with bi-predicted self-similarity compensation. In *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–4. IEEE, 2016.

[64] Tom Coughlin. 4K, HDR, HFR: Calculating the Storage Impact in Media Workflows. Technical report, 2016.

[65] Erwin Coumans and Yunfei Bai. PyBullet, a Python module for physics simulation for games, robotics and machine learning. `http://pybullet.org`, 2016–2021.

[66] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, pages 135—-142. Association for Computing Machinery, 1993.

[67] Aron Cserkaszky, Attila Barsi, Peter A. Kara, and Maria G. Martini. To interpolate or not to interpolate: Subjective assessment of interpolation performance on a light field display. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 55–60. IEEE, 2017.

[68] Aron Cserkaszky, Attila Barsi, Zsolt Nagy, Gabor Puhr, Tibor Balogh, and Peter A. Kara. Real-time light-field 3D telepresence. In *2018 7th European Workshop on Visual Information Processing (EUVIP)*, pages 1–5. IEEE, 2018.

[69] Aron Cserkaszky, Peter A. Kara, Attila Barsi, and Maria G. Martini. Expert evaluation of a novel light-field visualization format. In *2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2018.

[70] Aron Cserkaszky, Peter A. Kara, Attila Barsi, and Maria G. Martini. The potential synergies of visual scene reconstruction and medical image reconstruction. In *Novel Optical Systems Design and Optimization XXI*, volume 10746, pages 1–7. SPIE, 2018.

[71] Aron Cserkaszky, Peter A. Kara, Roopak R. Tamboli, Attila Barsi, Maria G. Martini, and Tibor Balogh. Light-field capture and display systems: limitations, challenges, and potentials. In *Novel Optical Systems Design and Optimization XXI*. International Society for Optics and Photonics, 2018.

[72] Aron Cserkaszky, Peter A. Kara, Roopak R. Tamboli, Attila Barsi, Maria G. Martini, Laszlo Bokor, and Tibor Balogh. Angularly continuous light-field format: concept, implementation, and evaluation. *Journal of the Society for Information Display*, 27(7):442–461, 2019.

[73] David G. Curry, Gary L. Martinsen, and Darrel G. Hopper. Capability of the human visual system. *Cockpit Displays X*, 5080:58–69, 2003.

[74] Praveen Cyriac, Trevor Canham, David Kane, and Marcelo Bertalmío. Vision models fine-tuned by cinema professionals for High Dynamic Range imaging in movies. *Multimedia Tools and Applications*, 80(2):2537–2563, 2021.

[75] Leonardo Da Vinci. *The notebooks of Leonardo da Vinci*, volume 1. Courier Corporation, 2012.

[76] Feng Dai, Xianyu Chen, Yike Ma, Guoqing Jin, and Qiang Zhao. Wide Range Depth Estimation from Binocular Light Field Camera. In *BMVC*, pages 1–11, 2018.

[77] Feng Dai, Jun Zhang, Yike Ma, and Yongdong Zhang. Lenselet image compression scheme based on subaperture images streaming. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4733–4737. IEEE, 2015.

[78] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[79] Oscar Danielsson, Magnus Holm, and Anna Syberfeldt. Augmented reality smart glasses for operators in production: Survey of relevant categories for supporting operators. *Procedia CIRP*, 93:1298–1303, 2020.

[80] Donald G. Dansereau, Bernd Girod, and Gordon Wetzstein. LiFF: Light field features in scale and depth. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8042–8051, 2019.

[81] Subbareddy Darukumalli, Peter A. Kara, Attila Barsi, Maria G. Martini, and Tibor Balogh. Subjective quality assessment of zooming levels and image reconstructions based on region of interest for light field displays. In *International Conference on 3D Imaging (IC3D)*. IEEE, 2016.

[82] Subbareddy Darukumalli, Peter A. Kara, Attila Barsi, Maria G. Martini, Tibor Balogh, and Ahmed Chehaibi. Performance comparison of subjective assessment

methodologies for light field displays. In *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 28–33. IEEE, 2016.

[83] Paul Debevec. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, pages 189—198, New York, NY, USA, 1998. Association for Computing Machinery.

[84] Paul Debevec. Image-based lighting. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06. Association for Computing Machinery, 2006.

[85] Paul E. Debevec and Jitendra Malik. *Recovering High Dynamic Range Radiance Maps from Photographs*. Association for Computing Machinery, New York, NY, USA, 1 edition, 2023.

[86] Yuri N Denisyuk. On the reflection of optical properties of an object in a wave field of light scattered by it. *Doklady Akademii Nauk SSSR*, 144(6):1275–1278, 1962.

[87] Elian Dib, Mikaël Le Pendu, and Christine Guillemot. Light field compression using Fourier disparity layers. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3751–3755. IEEE, 2019.

[88] Jeffrey M. DiCarlo and Brian A. Wandell. Rendering high dynamic range images. In *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications*, volume 3965, pages 392–401. SPIE, 2000.

[89] Xu Dong, Dai Qionghan, and Xu Wenli. Data compression of light field using wavelet packet. In *2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No. 04TH8763)*, volume 2, pages 1071–1074. IEEE, 2004.

[90] Oleksii Doronin, Attila Barsi, Peter A. Kara, and Maria G. Martini. Ray tracing for HoloVizio light field displays. In *International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2017.

[91] Antoine Dricot, Joel Jung, Marco Cagnazzo, Béatrice Pesquet, Frédéric Dufaux, Péter Tamás Kovács, and Vamsi K. Adhikarla. Subjective evaluation of Super Multi-View compressed contents on high-end light-field 3D displays. *Signal Processing: Image Communication*, 39:369–385, 2015.

[92] Charles R. Dyer. High Dynamic Range Digital Photography. Technical report, RPS journal, 2006.

[93] Netalee Efrat, Piotr Didyk, Mike Foshey, Wojciech Matusik, and Anat Levin. Cinema 3D: large scale automultiscopic display. *ACM Transactions on Graphics (TOG)*, 35(4):1–12, 2016.

[94] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K. Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM transactions on graphics (TOG)*, 36(6):1–15, 2017.

[95] Waqas Ellahi, Toinon Vigier, and Patrick Le Callet. Analysis of public light field datasets for visual quality assessment and new challenges. In *European Light Field Imaging Workshop*, 2019.

[96] CJ. Ellis. The pupillary light reflex in normal subjects. *British Journal of Ophthalmology*, 65(11):754–759, 1981.

[97] Gabriel Evans, Jack Miller, Mariangely Iglesias Pena, Anastacia MacAllister, and Eliot Winer. Evaluating the Microsoft HoloLens through an augmented reality assembly application. In *Degraded environments: sensing, processing, and display 2017*, volume 10197, pages 282–297. International Society for Optics and Photonics, SPIE, 2017.

[98] Hanzhi Fan, Dong Liu, Zhiwei Xiong, and Feng Wu. Two-stage convolutional neural network for light field super-resolution. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1167–1171. IEEE, 2017.

[99] Michael Faraday. LIV. Thoughts on ray-vibrations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 28(188):345–350, 1846.

[100] Saber Farag and Vladan Velisavljevic. A novel disparity-assisted block matching-based approach for super-resolution of light field images. In *2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2018.

[101] Reuben A. Farrugia and Christine Guillemot. Light field super-resolution using a low-rank prior and deep convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(5):1162–1175, 2019.

[102] Gregg E. Favalora. Volumetric 3D displays and application infrastructure. *Computer*, 38(8):37–44, 2005.

[103] Anton Fedosov, Evangelos Niforatos, Ivan Elhart, Teseo Schneider, Dmitry Anisimov, and Marc Langheinrich. Design and evaluation of a wearable AR system for sharing personalized content on ski resort maps. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, pages 141–152. Association for Computing Machinery, 2016.

[104] Joseph L Gabbard, Gregory M Fitch, and Hyungil Kim. Behind the glass: Driver challenges and opportunities for AR automotive applications. *Proceedings of the IEEE*, 102(2):124–136, 2014.

[105] D. Gabor. Microscopy by Reconstructed Wave-Fronts. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 197(1051):454–487, 1949.

[106] Dennis Gabor. A new microscopic principle. *Nature*, 161(4098):777–778, 1948.

[107] Todor Georgeiv, Ke Colin Zheng, Brian Curless, David Salesin, Shree Nayar, and Chintan Intwala. Spatio-Angular Resolution Tradeoffs in Integral Photography. In Tomas Akenine-Moeller and Wolfgang Heidrich, editors, *Symposium on Rendering*. The Eurographics Association, 2006.

[108] Todor Georgiev, Georgi Chunev, and Andrew Lumsdaine. Superresolution with the focused plenoptic camera. In *Computational Imaging IX*, volume 7873, pages 232–244. SPIE, 2011.

[109] Todor Georgiev, Andrew Lumsdaine, and Georgi Chunev. Using focused plenoptic cameras for rich image capture. *IEEE Computer Graphics and Applications*, 31(1):62–73, 2010.

[110] Andreĭ Gershun. The light field. *Journal of Mathematics and Physics*, 18(1-4):51–151, 1939.

[111] B. Girod, Chuo-Ling Chang, P. Ramanathan, and Xiaoqing Zhu. Light field compression using disparity-compensated lifting. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, volume 4, pages IV–760, 2003.

[112] Bastian Goldlücke, Oliver Klehm, Sven Wanner, and Elmar Eisemann. Plenoptic Cameras., 2015.

[113] J. W. Goodman. Digital Image Formation From Electronically Detected Holograms. In *Computerized Imaging Techniques*, volume 0010, pages 176–181. International Society for Optics and Photonics, SPIE, 1967.

[114] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 43—-54, New York, NY, USA, 1996. Association for Computing Machinery.

[115] Laurent Guillo, Xiaoran Jiang, Gauthier Lafruit, and Christine Guillemot. *Light field video dataset captured by a R8 Raytrix camera (with disparity maps)*. PhD thesis, Interational Organisation for Standardisation ISO/IEC JTC1/SC29/WG1 & WG11, 2018.

[116] Mary Guindy, Vamsi K. Adhikarla, Peter A. Kara, Tibor Balogh, and Aniko Simon. CLASSROOM: synthetic high dynamic range light field dataset. In *Applications of Digital Image Processing XLV*, volume 12226, pages 153–162. International Society for Optics and Photonics, SPIE, 2022.

[117] Mary Guindy, Attila Barsi, Peter A. Kara, Vamsi K. Adhikarla, Tibor Balogh, and Aniko Simon. Camera animation for immersive light field imaging. *Electronics*, 11(17), 2022.

[118] Mary Guindy, Attila Barsi, Peter A. Kara, Tibor Balogh, and Aniko Simon. Interaction methods for light field displays by means of a theater model environment. In *Holography: Advances and Modern Trends VII*, volume 11774, pages 109–118. SPIE, 2021.

[119] Mary Guindy, Attila Barsi, Peter A. Kara, Tibor Balogh, and Aniko Simon. Realistic physical camera motion for light field visualization. In *Holography: Advances and Modern Trends VII*, volume 11774, pages 70–77. SPIE, 2021.

[120] Mary Guindy and Peter A. Kara. Lessons Learned from Implementing Light Field Camera Animation: Implications, Limitations, Potentials, and Future Research Efforts. *Multimodal Technologies and Interaction*, 8(8), 2024.

[121] Mary Guindy and Peter A. Kara. Light Field Visualization for Training and Education: A Review. *Electronics*, 13(5), 2024.

[122] Mary Guindy, Peter A. Kara, Tibor Balogh, and Aniko Simon. Analysis of high dynamic range light field images in practical utilization contexts. In *Novel Optical Systems, Methods, and Applications XXV*, volume 12216, pages 144–152. SPIE, 2022.

[123] Mary Guindy, Peter A. Kara, Tibor Balogh, and Aniko Simon. Perceptual preference for 3D interactions and realistic physical camera motions on light field displays. In *Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations III*, volume 12125, pages 156–164. SPIE, 2022.

[124] Mary Guindy, Adhikarla K. Vamsi, Peter A. Kara, Tibor Balogh, and Aniko Simon. Performance evaluation of HDR image reconstruction techniques on light field images. In *2021 International Conference on 3D Immersion (IC3D)*, pages 1–7. IEEE, 2021.

[125] Mary Guindy, Adhikarla K. Vamsi, Peter A. Kara, Tibor Balogh, and Aniko Simon. Towards reconstructing HDR light fields by combining 2D and 3D CNN architectures. In *Big Data IV: Learning, Analytics, and Applications*, volume 12097, pages 192–197. SPIE, 2022.

[126] Karen Mc Guinness. Understanding 16 types of camera shots and angles (with GIFs!). https://boords.com/blog/16-types-of-camera-shots-and-angles-with-gifs. [Online; accessed 05-06-2021].

[127] M. Shahzeb Khan Gul, Thorsten Wolf, Michel Bätz, Matthias Ziegler, and Joachim Keinert. A high-resolution high dynamic range light-field dataset with an application to view synthesis and tone-mapping. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2020.

[128] Mark Guttenbrunner, Christoph Becker, and Andreas Rauber. Keeping the game alive: Evaluating strategies for the preservation of console video games. *The International Journal of Digital Curation*, 2010.

[129] Daniel A. Guttentag. Virtual reality: Applications and implications for tourism. *Tourism management*, 31(5):637–651, 2010.

[130] Mark Haigh-Hutchinson. *Real time cameras: A guide for game designers and developers*. CRC Press, 2009.

[131] Matthew Hamilton, Nicholas Wells, and Amilcar Soares. On Requirements for Field of Light Displays to Pass the Visual Turing Test. In *2022 IEEE International Symposium on Multimedia (ISM)*, pages 86–87. IEEE, 2022.

[132] Chris Hand. A survey of 3D interaction techniques. In *Computer graphics forum*, volume 16, pages 269–281. Wiley Online Library, 1997.

[133] Jassim Happa and Alessandro Artusi. Studying Illumination and Cultural Heritage. In *Visual Computing for Cultural Heritage*, pages 23–42. Springer, 2020.

[134] Jassim Happa, Alessandro Artusi, Piotr Dubla, Tom Bashford-Rogers, Kurt Debattista, Vedad Hulusic, and Alan Chalmers. The Virtual Reconstruction and Daylight Illumination of the Panagia Angeloktisti. In *VAST: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*. The Eurographics Association, 2009.

[135] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[136] Eugene Hecht. *Optics*. Pearson Education India, 2012.

[137] Shinsaku Hiura, Ankit Mohan, and Ramesh Raskar. Krill-eye: Superposition compound eye for wide-angle imaging via grin lenses. *IPSJ Transactions on Computer Vision and Applications*, 2:186–199, 2010.

[138] Bernhard Holzer. High dynamic range image formats. *Institute for Computer Graphics and Algorithms, TU Wien*, 2006.

[139] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4D light fields. In *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13*, pages 19–34. Springer, 2017.

[140] Darrel G. Hopper. 1000 X difference between current displays and capability of human visual system: payoff potential for affordable defense systems. In *Cockpit Displays VII: Displays for Defense Applications*, volume 4022, pages 378–389. SPIE, 2000.

[141] Alain Hore and Djemel Ziou. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.

[142] Xinjue Hu, Chenchen Wang, Yuxuan Pan, Yunming Liu, Yumei Wang, Yu Liu, Lin Zhang, and Shervin Shirmohammadi. 4DLFVD: A 4D light field video dataset. In *Proceedings of the 12th ACM Multimedia Systems Conference*, pages 287–292, 2021.

[143] Xinpeng Huang, Ping An, Yilei Chen, Deyang Liu, and Liquan Shen. Low bitrate light field compression with geometry and content consistency. *IEEE Transactions on Multimedia*, 24:152–165, 2020.

[144] Olivier Hugues, Philippe Fuchs, and Olivier Nannipieri. New augmented reality taxonomy: Technologies and features of augmented environment. *Handbook of augmented reality*, pages 47–63, 2011.

[145] Ianir Ideses and Leonid Yaroslavsky. New methods to produce high quality color anaglyphs for 3-D visualization. In *Image Analysis and Recognition: International Conference, ICIAR 2004, Porto, Portugal, September 29-October 1, 2004, Proceedings, Part II 1*, pages 273–280. Springer, 2004.

[146] Ianir Ideses and Leonid Yaroslavsky. Three methods that improve the visual quality of colour anaglyphs. *Journal of Optics A: Pure and Applied Optics*, 7(12), 2005.

[147] IEEE Standards Association. IEEE P3333.1.4-2022: Recommended Practice for the Quality Assessment of Light Field Imaging. `https://standards.ieee.org/ieee/3333.1.4/10873/`, 2023. [Online; accessed 14-08-2024].

[148] Ivo Ihrke, Timo Stich, Heiko Gottschlich, Marcus Magnor, and Hans-Peter Seidel. Fast incident light field acquisition and rendering. 2008.

[149] Frederic E. Ives. Parallax stereogram and process of making same., 1903. US725567A.

[150] Ashish Jagmohan, A. Sehgal, and Narendra Ahuja. Compression of lightfield rendered images using coset codes. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 1, pages 830–834. IEEE, 2003.

[151] Jacek Jankowski and Martin Hachet. A survey of interaction techniques for interactive 3D environments. In *Eurographics 2013-STAR*, 2013.

[152] Xiaoran Jiang, Mikaël Le Pendu, Reuben A Farrugia, and Christine Guillemot. Light field compression with homography-based low-rank approximation. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1132–1145, 2017.

[153] Panqi Jin, Gangyi Jiang, Yeyao Chen, Zhidi Jiang, and Mei Yu. Perceptual Light Field Image Coding with CTU Level Bit Allocation. In *International Conference on Computer Analysis of Images and Patterns*, pages 255–264. Springer, 2023.

[154] Xin Jin, Haixu Han, and Qionghai Dai. Image reshaping for efficient compression of plenoptic content. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1173–1186, 2017.

[155] Florian Kainz, Rod Bogart, and Drew Hess. The OpenEXR image file format. *ACM SIGGRAPH Technical Sketches*, 2003.

[156] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep High Dynamic Range Imaging of Dynamic Scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*, 36(4), 2017.

[157] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep HDR video from sequences with alternating exposures. In *Computer Graphics Forum*, volume 38, pages 193–205. Wiley Online Library, 2019.

[158] Kenji Kansaku, Naoki Hata, and Kouji Takano. My thoughts through a robot's eyes: An augmented reality-brain–machine interface. *Neuroscience research*, 66(2):219–222, 2010.

[159] Peter A. Kara, Attila Barsi, Roopak R. Tamboli, Mary Guindy, Maria G. Martini, Tibor Balogh, and Aniko Simon. Recommendations on the viewing distance of light field displays. In *Digital Optical Technologies 2021*, volume 11788, pages 166–179. SPIE, 2021.

[160] Peter A. Kara, Aron Cserkaszky, Attila Barsi, Tamas Papp, Maria G. Martini, and László Bokor. The interdependence of spatial and angular resolution in the quality of experience of light field visualization. In *2017 International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2017.

[161] Peter A. Kara, Aron Cserkaszky, Subbareddy Darukumalli, Attila Barsi, and Maria G. Martini. On the edge of the seat: Reduced angular resolution of a light field cinema with fixed observer positions. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2017.

[162] Peter A. Kara, Aron Cserkaszky, Maria G. Martini, Attila Barsi, Laszlo Bokor, and Tibor Balogh. Evaluation of the concept of dynamic adaptive streaming of light field video. *IEEE Transactions on Broadcasting*, 64(2):407–421, 2018.

[163] Peter A. Kara, Mary Guindy, Tibor Balogh, and Aniko Simon. The perceptually-supported and the subjectively-preferred viewing distance of projection-based light field displays. In *2021 International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2021.

[164] Peter A. Kara, Mary Guindy, Qiu Xinyu, Vince A. Szakal, Tibor Balogh, and Aniko Simon. The effect of angular resolution and 3D rendering on the perceived quality of the industrial use cases of light field visualization. In *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 600–607. IEEE, 2022.

[165] Peter A. Kara, Péter T. Kovács, Maria G. Martini, Attila Barsi, Kristóf Lackner, and Tibor Balogh. From a different point of view: How the field of view of light field displays affects the willingness to pay and to use. In *8th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2016.

[166] Peter A. Kara, Peter T. Kovacs, Suren Vagharshakyan, Maria G. Martini, Attila Barsi, Tibor Balogh, Aleksandra Chuchvara, and Ahmed Chehaibi. The effect of light field reconstruction and angular resolution reduction on the quality of experience. In *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 781–786. IEEE, 2016.

[167] Peter A. Kara, Maria G. Martini, Peter T. Kovács, Sandor Imre, Attila Barsi, Kristof Lackner, and Tibor Balogh. Perceived quality of angular resolution for light field displays and the validy of subjective assessment. In *2016 International Conference on 3D Imaging (IC3D)*, pages 1–7. IEEE, 2016.

[168] Peter A. Kara, Maria G. Martini, Zsolt Nagy, and Attila Barsi. Cinema as large as life: Large-scale light field cinema system. In *2017 International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2017.

[169] Peter A. Kara, Roopak R. Tamboli, Aron Cserkaszky, Attila Barsi, Aniko Simon, Agnes Kusz, Laszlo Bokor, and Maria G. Martini. Objective and subjective assessment of binocular disparity for projection-based light field displays. In *International Conference on 3D Immersion (IC3D)*. IEEE, 2019.

[170] Peter A. Kara, Roopak R. Tamboli, Aron Cserkaszky, Maria G. Martini, Attila Barsi, and Laszlo Bokor. The viewing conditions of light-field video for subjective quality assessment. In *2018 International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2018.

[171] Peter A. Kara, Roopak R. Tamboli, Oleksii Doronin, Aron Cserkaszky, Attila Barsi, Zsolt Nagy, Maria G. Martini, and Aniko Simon. The key performance indicators of projection-based light field visualization. *Journal of Information Display*, 2019.

[172] Peter A. Kara, Roopak R. Tamboli, Edris Shafiee, Maria G. Martini, Aniko Simon, and Mary Guindy. Beyond perceptual thresholds and personal preference: Towards novel research questions and methodologies of quality of experience studies on light field visualization. *Electronics*, 11(6), 2022.

[173] Hirokazu Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 85–94. IEEE, 1999.

[174] Masaki Kawakami, Chihiro Tsutake, Keita Takahashi, and Toshiaki Fujii. Compressing Light Field as Multiplane Image. *ITE Transactions on Media Technology and Applications*, 11(2):27–33, 2023.

[175] Masahiro Kawakita, Shoichiro Iwasawa, Mikio Sakai, Yasuyuki Haino, Masahito Sato, and Naomi Inoue. 3D image quality of 200-inch glasses-free 3D display system. In *Stereoscopic Displays and Applications XXIII*, volume 8288, pages 63–70. SPIE, 2012.

[176] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus H. Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, 32(4), 2013.

[177] Georg Klein and Tom Drummond. Sensor fusion and occlusion refinement for tablet-based AR. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 38–47. IEEE, 2004.

[178] Georg Klein and David W Murray. Simulating low-cost cameras for augmented reality compositing. *IEEE transactions on visualization and computer graphics*, 16(3):369–380, 2010.

[179] Keunsoo Ko, Yeong Jun Koh, Soonkeun Chang, and Chang-Su Kim. Light field super-resolution via adaptive feature remixing. *IEEE Transactions on Image Processing*, 30:4114–4128, 2021.

[180] George Alex Koulieris, Kaan Akşit, Michael Stengel, Rafał K Mantiuk, Katerina Mania, and Christian Richardt. Near-eye display and tracking technologies for virtual and augmented reality. In *Computer Graphics Forum*, volume 38, pages 493–519. Wiley Online Library, 2019.

[181] Péter Tamás Kovács, Atanas Boev, Robert Bregovic, and Atanas Gotchev. Quality measurement of 3d light-field displays, 2014.

[182] Péter Tamás Kovács, Robert Bregović, Atanas Boev, Attila Barsi, and Atanas Gotchev. Quantifying spatial and angular resolution of light-field 3-D displays. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1213–1222, 2017.

[183] Péter Tamás Kovács, Kristóf Lackner, Attila Barsi, Ákos Balázs, Atanas Boev, Robert Bregović, and Atanas Gotchev. Measurement of perceived spatial resolution in 3D light-field displays. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 768–772. IEEE, 2014.

[184] Shinjini Kundu. Light field compression using homography and 2D warping. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1349–1352. IEEE, 2012.

[185] Douglas Lanman, Daniel Crispell, Megan Wachs, and Gabriel Taubin. Spherical catadioptric arrays: Construction, multi-view geometry, and calibration. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pages 81–88. IEEE, 2006.

[186] Douglas Lanman, Ramesh Raskar, Amit Agrawal, and Gabriel Taubin. Shield fields: modeling and capturing 3d occluders. *ACM Transactions on Graphics (TOG)*, 27(5):1–10, 2008.

[187] Gregory Ward Larson. Overcoming gamut and dynamic range limitations in digital images. In *Color and imaging conference*, volume 6, pages 214–219. Society of Imaging Science and Technology, 1998.

[188] Gregory Ward Larson, Holly Rushmeier, and Christine Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, 1997.

[189] Steven M LaValle, Anna Yershova, Max Katsev, and Michael Antonov. Head tracking for the Oculus Rift. In *2014 IEEE international conference on robotics and automation (ICRA)*, pages 187–194. IEEE, 2014.

[190] Mikaël Le Pendu, Christine Guillemot, and Aljosa Smolic. High dynamic range light fields via weighted low rank approximation. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1728–1732. IEEE, 2018.

[191] Chang-Kun Lee, Soon-gi Park, Seokil Moon, Jong-Young Hong, and Byoungho Lee. Compact multi-projection 3D display system with light-guide projection. *Optics express*, 23(22):28945–28959, 2015.

[192] Titus Leistner, Hendrik Schilling, Radek Mackowiak, Stefan Gumhold, and Carsten Rother. Learning to Think Outside the Box: Wide-Baseline Light Field Depth Estimation with EPI-Shift. In *International Conference on 3D Vision (3DV)*, pages 249–257. IEEE, 2019.

[193] Emmett N. Leith and Juris Upatnieks. New techniques in wavefront reconstruction. *J. Opt. Soc. Am*, 51(11):1469–1473, 1961.

[194] Cédric Lejeune. Encoding and displaying HDR content for cinema. Technical report, 2017.

[195] Marc Levoy, Billy Chen, Vaibhav Vaish, Mark Horowitz, Ian McDowall, and Mark Bolas. Synthetic aperture confocal imaging. *ACM Transactions on Graphics (ToG)*, 23(3):825–834, 2004.

[196] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996.

[197] Chen Li and Xu Zhang. High dynamic range and all-focus image from light field. In *2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, pages 7–12. IEEE, 2015.

[198] Li Li, Zhu Li, Bin Li, Dong Liu, and Houqiang Li. Pseudo-sequence-based 2-D hierarchical coding structure for light-field image compression. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1107–1119, 2017.

[199] Yun Li, Roger Olsson, and Mårten Sjöström. Compression of unfocused plenoptic images using a displacement intra prediction. In *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–4. IEEE, 2016.

[200] Yun Li, Märten Sjöström, Roger Olsson, and Ulf Jennehag. Efficient intra prediction scheme for light field image compression. In *2014 IEEE International conference on acoustics, speech and signal processing (ICASSP)*, pages 539–543. IEEE, 2014.

[201] Yun Li, Mårten Sjöström, Roger Olsson, and Ulf Jennehag. Scalable coding of plenoptic images by using a sparse set and disparities. *IEEE Transactions on Image Processing*, 25(1):80–91, 2015.

[202] Chia-Kai Liang, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu, and Homer H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *ACM siggraph 2008 papers*, pages 1–10. 2008.

[203] Chia-Kai Liang and Ravi Ramamoorthi. A light transport framework for lenslet light field cameras. *ACM Transactions on Graphics (TOG)*, 34(2):1–19, 2015.

[204] JaeGuyn Lim, HyunWook Ok, ByungKwan Park, JooYoung Kang, and SeongDeok Lee. Improving the spatail resolution based on 4D light field data. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1173–1176. IEEE, 2009.

[205] Jan Lindström, Markus Hulthén, Michael Sandborg, and Åsa Carlsson-Tedgren. Development and assessment of a quality assurance device for radiation field–light field congruence testing in diagnostic radiology. *SPIE Journal of Medical Imaging*, 7(6), 2020.

[206] Gabriel Lippmann. Epreuves reversibles donnant la sensation du relief. *J. Phys. Theor. Appl.*, 7(1):821–825, 1908.

[207] Ce Liu. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009.

[208] Deyang Liu, Xinpeng Huang, Wenfa Zhan, Liefu Ai, Xin Zheng, and Shulin Cheng. View synthesis-based light field image compression using a generative adversarial network. *Information Sciences*, 545:118–131, 2021.

[209] Dong Liu, Lizhi Wang, Li Li, Zhiwei Xiong, Feng Wu, and Wenjun Zeng. Pseudo-sequence-based light field image compression. In *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–4. IEEE, 2016.

[210] Gaosheng Liu, Huanjing Yue, Jiamin Wu, and Jingyu Yang. Intra-inter view interaction network for light field image super-resolution. *IEEE Transactions on Multimedia*, 2021.

[211] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1651–1660, 2020.

[212] Elizabeth F. Loftus and Hunter G. Hoffman. Misinformation and memory: the creation of new memories. *Journal of experimental psychology: General*, 118(1):100–104, 1989.

[213] Andrew Lumsdaine and Todor Georgiev. The focused plenoptic camera. In *2009 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2009.

[214] Daniel SF. Magalhães, Rolando L. Serra, Andre L. Vannucci, Alfredo B. Moreno, and Li M. Li. Glasses-free 3D viewing systems for medical imaging. *Optics & Laser Technology*, 44(3):650–655, 2012.

[215] Marcus Magnor and Bernd Girod. Data compression for light-field rendering. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(3):338–343, 2000.

[216] Marcus A. Magnor, Andreas Endmann, and Bernd Girod. Progressive Compression and Rendering of Light Fields. In *VMV*, pages 199–204. Citeseer, 2000.

[217] Rafal Mantiuk, Scott J. Daly, Karol Myszkowski, and Hans-Peter Seidel. Predicting visible differences in high dynamic range images: model and its calibration. In *Human Vision and Electronic Imaging X*, volume 5666, pages 204–214. International Society for Optics and Photonics, 2005.

[218] Rafał Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)*, 30(4):1–14, 2011.

[219] Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. High Dynamic Range Imaging. In *Wiley Encyclopedia of Electrical and Electronics Engineering*. John Wiley Sons, Ltd, 2015.

[220] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *Computer Graphics Forum*, volume 37, pages 37–49. Wiley Online Library, 2018.

[221] Fabio Marton, Marco Agus, Enrico Gobbetti, Giovanni Pintore, and Marcos Balsa Rodriguez. Natural exploration of 3D massive models on large-scale light field

displays using the FOX proximal navigation technique. *Computers & Graphics*, 36(8):893–903, 2012.

[222] Fabio Marton, Marco Agus, Giovanni Pintore, and Enrico Gobbetti. FOX: The Focus Sliding Surface metaphor for natural exploration of massive models on large-scale light field displays. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 83–90, 2011.

[223] John Francis Mccullagh. Mastering the Subtle Techniques of the Zoom and the Punch In, 2019.

[224] Morgan McGuire. Computer Graphics Archive, 2017. `https://casual-effects.com/data`.

[225] John P. McIntire, Paul R. Havig, and Eric E. Geiselman. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays*, 35(1):18–26, 2014.

[226] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 39–46, 1995.

[227] Norman Medoff and Edward J. Fink. *Portable Video: Eng & Efp*. Routledge, 2013.

[228] Christopher A. Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1375–1385, 2020.

[229] Meysam Minoufekr, Pascal Schug, Pascal Zenker, and Peter W Plapper. Modelling of CNC Machine Tools for Augmented Reality Assistance Applications using Microsoft Hololens. In *ICINCO (2)*, pages 627–636, 2019.

[230] Kaushik Mitra and Ashok Veeraraghavan. Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 22–28. IEEE, 2012.

[231] Yu Mo, Yingqian Wang, Chao Xiao, Jungang Yang, and Wei An. Dense dual-attention network for light field image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7):4431–4443, 2021.

[232] Nuno Barroso Monteiro, Simao Marto, Joao Pedro Barreto, and José Gaspar. Depth range accuracy for plenoptic cameras. *Computer Vision and Image Understanding*, 168:104–117, 2018.

[233] Ricardo Monteiro, Luis Lucas, Caroline Conti, Paulo Nunes, Nuno Rodrigues, Sérgio Faria, Carla Pagliari, Eduardo Da Silva, and Luís Soares. Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction. In *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–4. IEEE, 2016.

[234] Sergio Moreschini, Filipe Gama, Robert Bregovic, and Atanas Gotchev. CIVIT datasets: Horizontal-parallax-only densely-sampled light-fields. In *Proc. Eur. Light Field Imag. Workshop*, volume 6, pages 1–4, 2019.

[235] Antoine Mousnier, Elif Vural, and Christine Guillemot. Partial light field tomographic reconstruction from a fixed-camera focal stack. *arXiv preprint arXiv:1503.01903*, 2015.

[236] Yoshitaka Narita, Shinsuke Tsukagoshi, Masahiro Suzuki, Yasuji Miyakita, Makoto Ohno, Hideyuki Arita, Yasuo Saito, Yoshiyuki Kokojima, Naofumi Watanabe, Noriyuki Moriyama, and Soichiro Shibui. Usefulness of a glass-free medical three-dimensional autostereoscopic display in neurosurgery. *International journal of computer assisted radiology and surgery*, 9(5):905–911, 2014.

[237] F. Perez Nava and JP. Luke. Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera. In *2009 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pages 1–4. IEEE, 2009.

[238] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, Pat Hanrahan, and Duval Design. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11, 2005.

[239] Shiro Ogata, Junya Ishida, and Tomohiko Sasano. Optical sensor array in an artificial compound eye. *Optical Engineering*, 33(11):3649–3655, 1994.

[240] Fumio Okano, Jun Arai, Haruo Hoshino, and Ichiro Yuyama. Three-dimensional video system based on integral photography. *Optical Engineering*, 38(6):1072–1077, 1999.

[241] Fumio Okura, Masayuki Kanbara, and Naokazu Yokoya. Full spherical high dynamic range imaging from the sky. In *2012 IEEE International Conference on Multimedia and Expo*, pages 325–332. IEEE, 2012.

[242] Soon-gi Park. Augmented and mixed reality optical see-through combiners based on plastic optics. *Information Display*, 37(4):6–11, 2021.

[243] Pradip Paudyal, Federica Battisti, Mårten Sjöström, Roger Olsson, and Marco Carli. Towards the perceptual quality evaluation of compressed light field images. *IEEE Transactions on Broadcasting*, 63(3):507–522, 2017.

[244] Zhao Pei, Yawen Li, Miao Ma, Jun Li, Chengcai Leng, Xiaoqiang Zhang, and Yanning Zhang. Occluded-object 3D reconstruction using camera array synthetic aperture imaging. *Sensors*, 19(3), 2019.

[245] Cristian Perra. Lossless plenoptic image compression using adaptive block differential prediction. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1231–1234. IEEE, 2015.

[246] David Perry and Rusel DeMaria. *David Perry on Game Design: A Brainstorming ToolBox*. Course Technology, 2009.

[247] Photoshop. High dynamic range images. `https://helpx.adobe.com/photoshop/using/high-dynamic-range-images.html`, 2020. [Online; accessed 20-05-2022].

[248] Carlos R. Ponce and Richard T. Born. Stereopsis. *Current Biology*, 18(18):R845–R850, 2008.

[249] Vladan Popovic, Hossein Afshari, Alexandre Schmid, and Yusuf Leblebici. Real-time implementation of Gaussian image blending in a spherical light field camera. In *2013 IEEE international conference on industrial technology (ICIT)*, pages 1173–1178. IEEE, 2013.

[250] Kari Pulli. 11-2: Invited paper: Meta 2: Immersive optical-see-through augmented reality. In *SID Symposium Digest of Technical Papers*, volume 48, pages 132–133. Wiley Online Library, 2017.

[251] G. Ramponi, A. Badano, S. Bonfiglio, L. Albani, and G. Guarnieri. An Application of HDR in Medical Imaging. In *High Dynamic Range Video*, pages 499–518. Elsevier, 2016.

[252] Bruce I. Reiner, Eliot L. Siegel, and Elizabeth A. Krupinski. Digital radiographic image presentation and display. *Advances in Digital Radiography: RSNA Categorical Course in Diagnostic Radiology Physics. Oak Brook: Radiological Society of North America, Inc*, pages 79–89, 2003.

[253] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting.* Morgan Kaufmann, 2010.

[254] Martin Rerabek and Touradj Ebrahimi. New light field image dataset. In *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.

[255] Martin Rerabek, Lin Yuan, Léonard Antoine Authier, and Touradj Ebrahimi. [ISO/IEC JTC 1/SC 29/WG1 contribution] EPFL Light-Field Image Dataset, 2015.

[256] Christian Reuter, Thomas Ludwig, and Patrick Mischur. RescueGlass: collaborative applications involving head-mounted displays for red cross rescue dog units. *Computer Supported Cooperative Work (CSCW)*, 28:209–246, 2019.

[257] Mark S. Robinson, SM. Brylow, M. Tschimmel, D. Humm, SJ. Lawrence, PC. Thomas, Brett W. Denevi, E. Bowman-Cisneros, J. Zerr, MA. Ravine, et al. Lunar reconnaissance orbiter camera (LROC) instrument overview. *Space science reviews*, 150:81–124, 2010.

[258] Ségolène Rogge, Ionut Schiopu, and Adrian Munteanu. Depth Estimation for Light-Field Images Using Stereo Matching and Convolutional Neural Networks. *Sensors*, 20(21), 2020.

[259] Mario A. Rojas-Sánchez, Pedro R. Palos-Sánchez, and José A. Folgado-Fernández. Systematic literature review and bibliometric analysis on virtual reality and education. *Education and Information Technologies*, 28(1):155–192, 2023.

[260] Mattia Rossi, Mireille El Gheche, and Pascal Frossard. A nonsmooth graph-based approach to light field super-resolution. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2590–2594. IEEE, 2018.

[261] Mattia Rossi and Pascal Frossard. Graph-based light field super-resolution. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2017.

[262] Pawel Rotter. Why did the 3D revolution fail?: the present and future of stereoscopy [commentary]. *IEEE Technology and Society Magazine*, 36(1):81–85, 2017.

[263] Neus Sabater, Guillaume Boisson, Benoit Vandame, Paul Kerbiriou, Frederic Babon, Matthieu Hog, Remy Gendrot, Tristan Langlois, Olivier Bureller, Arno Schubert, and V. Allie. Dataset and pipeline for multi-view light-field video. In *Proceedings of the IEEE conference on computer vision and pattern recognition Workshops*, pages 30–40, 2017.

[264] Maximilian Schambach and Michael Heizmann. A Multispectral Light Field Dataset for Light Field Deep Learning. *IEEE access*, 8:193492–193502, 2020.

[265] Jesse Schell. *The Art of Game Design: A book of lenses.* CRC press, Taylor & Francis, 2008.

[266] Edris Shafiee and Maria G. Martini. Datasets for the quality assessment of light field imaging: Comparison and future directions. *IEEE Access*, 11:15014–15029, 2023.

[267] Liang Shan, Ping An, Deyang Liu, and Ran Ma. Subjective evaluation of light field images for quality assessment database. In *Digital TV and Wireless Multimedia Communication: 14th International Forum, IFTC 2017, Shanghai, China, November 8-9, 2017, Revised Selected Papers 14*, pages 267–276. Springer, 2018.

[268] Jianrui Shao, Enjian Bai, Xueqin Jiang, and Yun Wu. Light-Field Image Compression Based on a Two-Dimensional Prediction Coding Structure. *Information*, 15(6), 2024.

[269] Sumit Shekhar, Shida Kunz Beigpour, Matthias Ziegler, Michał Chwesiuk, Dawid Paleń, Karol Myszkowski, Joachim Keinert, Radosław Mantiuk, and Piotr Didyk. Light-field intrinsic dataset. In *British Machine Vision Conference 2018 (BMVC)*. British Machine Vision Association, 2018.

[270] Jinglei Shi, Yihong Xu, and Christine Guillemot. Learning Kernel-Modulated Neural Representation for Efficient Light Field Compression. *arXiv preprint arXiv:2307.06143*, 2023.

[271] Likun Shi, Shengyang Zhao, Wei Zhou, and Zhibo Chen. Perceptual evaluation of light field image. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 41–45. IEEE, 2018.

[272] Heung-Yeung Shum, Sing Bing Kang, and Shing-Chow Chan. Survey of image-based representations and compression techniques. *IEEE transactions on circuits and systems for video technology*, 13(11):1020–1037, 2003.

[273] Aniko Simon, Mary Guindy, Peter A. Kara, Tibor Balogh, and Laszlo Szy. Through a different lens: the perceived quality of light field visualization assessed by test participants with imperfect visual acuity and color blindness. In *Big Data IV: Learning, Analytics, and Applications*, volume 12097, pages 212–221. SPIE, 2022.

[274] Aniko Simon, Peter A. Kara, Mary Guindy, Xinyu Qiu, Laszlo Szy, and Tibor Balogh. One step closer to a better experience: Analysis of the suitable viewing distance ranges of light field visualization usage contexts for observers with reduced visual capabilities. In *Novel Optical Systems, Methods, and Applications XXV*, volume 12216, pages 133–143. SPIE, 2022.

[275] Tomas Sluka, Alexander Kvasov, and Tomas Kubes. Digital Light-Field. https://creal.com/app/uploads/2022/04/CREAL-White-Paper-Digital-Light-field.pdf, 2021. [Online; accessed 26-07-2024].

[276] Chris Solarski. *Interactive stories and video game art: A storytelling framework for game design*. CRC Press, 2017.

[277] Svend Erik Borre Sorensen, Per Skafte Hansen, and Nils Lykke Sorensen. Method for recording and viewing stereoscopic images in color using multichrome filters, 2004. US Patent 6,687,003.

[278] Ryan Southall. High Dynamic Range imaging in design education. *CEBE Transactions*, 8(2):50–62, 2011.

[279] Pratul P Srinivasan, Tongzhou Wang, Ashwin Sreelal, Ravi Ramamoorthi, and Ren Ng. Learning to synthesize a 4D RGBD light field from a single image. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2243–2251, 2017.

[280] Storyblocks. 7 Basic Camera Movements. https://blog.storyblocks.com/video-tutorials/7-basic-camera-movements/, 2019. [Online; accessed 20-05-2021].

[281] Jessi Stumpfel, Andrew Jones, Andreas Wenger, Chris Tchou, Tim Hawkins, and Paul Debevec. Direct HDR capture of the sun and sky. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06. Association for Computing Machinery, 2006.

[282] Kelvin Sung, Peter Shirley, and Steven Baer. *Essentials of interactive computer graphics: concepts and implementation*. CRC Press, 2008.

[283] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, pages 757–764. Association for Computing Machinery, 1968.

[284] Toqeer Ali Syed, Muhammad Shoaib Siddiqui, Hurria Binte Abdullah, Salman Jan, Abdallah Namoun, Ali Alzahrani, Adnan Nadeem, and Ahmad B Alkhodre. In-depth review of augmented reality: Tracking technologies, development tools, AR displays, collaborative AR, and security concerns. *Sensors*, 23(1), 2023.

[285] Yuichi Taguchi, Amit Agrawal, Srikumar Ramalingam, and Ashok Veeraraghavan. Axial light field for curved mirrors: Reflect your perspective, widen your view. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 499–506. IEEE, 2010.

[286] Yuichi Taguchi, Amit Agrawal, Ashok Veeraraghavan, Srikumar Ramalingam, and Ramesh Raskar. Axial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering. *ACM Trans. Graph.*, 29(6), 2010.

[287] Roopak R. Tamboli, Balasubramanyam Appina, Sumohana Channappayya, and Soumya Jana. Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display. *Signal Processing: Image Communication*, 47:42–55, 2016.

[288] Roopak R. Tamboli, Balasubramanyam Appina, Sumohana S. Channappayya, and Soumya Jana. Achieving high angular resolution via view synthesis: Quality assessment of 3D content on super multiview lightfield display. In *2017 International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2017.

[289] Roopak R. Tamboli, Aron Cserkaszky, Peter A. Kara, Attila Barsi, and Maria G. Martini. Objective quality evaluation of an angularly-continuous light-field format. In *2018 International Conference on 3D Immersion (IC3D)*, pages 1–8. IEEE, 2018.

[290] Roopak R. Tamboli, Peter A. Kara, Aron Cserkaszky, Attila Barsi, Maria G. Martini, and Soumya Jana. Canonical 3D object orientation for interactive light-field visualization. In *Applications of Digital Image Processing XLI*, volume 10752, pages 77–83. SPIE, 2018.

[291] Roopak R. Tamboli, M. Shanmukh Reddy, Peter A. Kara, Maria G. Martini, Sumohana S. Channappayya, and Soumya Jana. A high-angular-resolution turntable data-set for experiments on light field visualization quality. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2018.

[292] Roopak R. Tamboli, Kiran Kumar Vupparaboina, Jayanth Ready, Soumya Jana, and Sumohana Channappayya. A subjective evaluation of true 3D images. In *2014 International Conference on 3D Imaging (IC3D)*, pages 1–8. IEEE, 2014.

[293] Jun Tanida, Tomoya Kumagai, Kenji Yamada, Shigehiro Miyatake, Kouichi Ishida, Takashi Morimoto, Noriyuki Kondou, Daisuke Miyazaki, and Yoshiki Ichioka. Thin observation module by bound optics (TOMBO): concept and experimental verification. *Applied optics*, 40(11):1806–1813, 2001.

[294] Jun Tanida, Rui Shogenji, Yoshiro Kitamura, Kenji Yamada, Masaru Miyamoto, and Shigehiro Miyatake. Color imaging with an integrated compound imaging system. *Optics Express*, 11(18):2109–2117, 2003.

[295] Kedeng Tong, Xin Jin, Chen Wang, and Fan Jiang. SADN: learned light field image compression with spatial-angular decorrelation. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1870–1874. IEEE, 2022.

[296] Alan Mathison Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[297] Kensuke Ueda, Takafumi Koike, Keita Takahashi, and Takeshi Naemura. Adaptive integral photography imaging with variable-focus lens array. In *Stereoscopic Displays and Applications XIX*, volume 6803, pages 443–451. SPIE, 2008.

[298] Kensuke Ueda, Dongha Lee, Takafumi Koike, Keita Takahashi, and Takeshi Naemura. Multi-focal compound eye: liquid lens array for computational photography. In *ACM SIGGRAPH 2008 New Tech Demos*, SIGGRAPH '08, New York, NY, USA, 2008. Association for Computing Machinery.

[299] Jonas Unger, Andreas Wenger, Tim Hawkins, Andrew Gardner, and Paul E. Debevec. Capturing and Rendering with Incident Light Fields. *Rendering Techniques*, pages 1–10, 2003.

[300] Hakan Urey, Kishore V Chellappan, Erdem Erden, and Phil Surman. State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE*, 99(4):540–555, 2011.

[301] Vaibhav Vaish and Andrew Adams. The (new) stanford light field archive. *Computer Graphics Laboratory, Stanford University*, 6(7):3, 2008.

[302] Maurice HPH van Beurden, Gert Van Hoey, Haralambos Hatzakis, and Wijnand A. Ijsselsteijn. Stereoscopic displays in medical domains: a review of perception and performance effects. *Human Vision and Electronic Imaging XIV*, 7240:74–88, 2009.

[303] Vinh Van Duong, Thuc Nguyen Huu, Jonghoon Yim, and Byeungwoo Jeon. Light field image super-resolution network via joint spatial-angular and epipolar information. *IEEE Transactions on Computational Imaging*, 9:350–366, 2023.

[304] Ashok Veeraraghavan, Amit Agrawal, Ramesh Raskar, Ankit Mohan, and Jack Tumblin. Non-refractive modulators for encoding and capturing scene appearance and depth. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[305] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3), 2007.

[306] Alexandre Vieira, Helder Duarte, Cristian Perra, Luis Tavora, and Pedro Assuncao. Data formats for high efficiency coding of Lytro-Illum light fields. In *2015 international conference on image processing theory, tools and applications (IPTA)*, pages 494–497. IEEE, 2015.

[307] Irene Viola and Touradj Ebrahimi. VALID: Visual quality assessment for light field images dataset. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2018.

[308] H. Kenneth Walker, W. Dallas Hall, and J. Willis Hurst. Clinical methods: the history, physical, and laboratory examinations. 1990.

[309] Ting-Chun Wang, Jun-Yan Zhu, Ebi Hiroaki, Manmohan Chandraker, Alexei A Efros, and Ravi Ramamoorthi. A 4D light-field dataset and CNN architectures for material recognition. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 121–138. Springer, 2016.

[310] Wei Wang, Meihui Zhang, Gang Chen, HV Jagadish, Beng Chin Ooi, and Kian-Lee Tan. Database meets deep learning: Challenges and opportunities. *ACM SIGMOD Record*, 45(2):17–22, 2016.

[311] Xiang Wang, Lin Li, and GuangQi Hou. High-resolution light field reconstruction using a hybrid imaging system. *Applied optics*, 55(10):2580–2593, 2016.

[312] Yingqian Wang, Longguang Wang, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Spatial-angular interaction for light field image super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, pages 290–308. Springer, 2020.

[313] Yingqian Wang, Jungang Yang, Longguang Wang, Xinyi Ying, Tianhao Wu, Wei An, and Yulan Guo. Light field image super-resolution using deformable convolution. *IEEE Transactions on Image Processing*, 30:1057–1071, 2020.

[314] Yunlong Wang, Fei Liu, Kunbo Zhang, Guangqi Hou, Zhenan Sun, and Tieniu Tan. LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. *IEEE Transactions on Image Processing*, 27(9):4274–4286, 2018.

[315] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE transactions on pattern analysis and machine intelligence*, 36(3):606–619, 2013.

[316] Sven Wanner, Stephan Meister, and Bastian Goldluecke. Datasets and benchmarks for densely sampled 4D light fields. In *VMV*, volume 13, pages 225–226, 2013.

[317] Gregory J. Ward. The RADIANCE lighting simulation and rendering system. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, pages 459–472. Association for Computing Machinery, 1994.

[318] Colin Ware, Kevin Arthur, and Kellogg S Booth. Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, CHI '93, pages 37–42. Association for Computing Machinery, 1993.

[319] Donald Watson. Architecture, technology, and environment. *Journal of Architectural Education*, 51(2):119–126, 1997.

[320] Gordon Wetzstein. *Computational plenoptic image acquisition and display.* PhD thesis, University of British Columbia, 2011.

[321] Gordon Wetzstein, D. Lanman, M. Hirsch, and Ramesh Raskar. Real-time Image Generation for Compressive Light Field Displays. In *Journal of Physics: Conference Series*, volume 415, pages 1–9. IOP Publishing, 2013.

[322] Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, 2017.

[323] Po-Jung Wu, Kuang-Tsu Shih, and Homer Chen. Dual-camera HDR synthesis guided by long-exposure image. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–4. IEEE, 2016.

[324] Yichao Xu, Kazuki Maeno, Hajime Nagahara, and Rin-ichiro Taniguchi. Camera array calibration for light field acquisition. *Frontiers of Computer Science*, 9(5):691–702, 2015.

[325] J. Yang, C. Lee, A. Isaksen, and L. McMillan. A Low-Cost Portable Light Field Capture Device. In *Siggraph Conference Abstracts and Applications*, 2000.

[326] Jason C. Yang, Matthew Everett, Chris Buehler, and Leonard McMillan. A real-time distributed light field camera. *Rendering Techniques*, 2002(77-86), 2002.

[327] Leonid Yaroslavsky and N. S. Merzlyakov. *Methods of digital holography.* Consultants Bureau, New York, 1980.

[328] Hojatollah Yeganeh, Zhou Wang, and Edward R. Vrscay. Adaptive windowing for optimal visualization of medical images based on a structural fidelity measure. In *International Conference Image Analysis and Recognition*, pages 321–330. Springer, 2012.

[329] Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, and In So Kweon. Light-field image super-resolution using convolutional neural network. *IEEE Signal Processing Letters*, 24(6):848–852, 2017.

[330] Haiyang Yu, Xiaoyu Jiang, Xingpeng Yan, Zhiqiang Yan, Chenqing Wang, Zhan Yan, and Fenghao Wang. Research Summary on Light Field Display Technology Based on Projection. In *Journal of Physics: Conference Series*, volume 1682. IOP Publishing, 2020.

[331] Dingcheng Yue, Muhammad Shahzeb Khan Gul, Michel Bätz, Joachim Keinert, and Rafał Mantiuk. A benchmark of light field view interpolation methods. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2020.

[332] Faezeh Sadat Zakeri, Ahmed Durmush, Matthias Ziegler, Michel Bätz, and Joachim Keinert. Non-planar inside-out dense light-field dataset and reconstruction pipeline. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1059–1063. IEEE, 2019.

[333] Cha Zhang and Tsuhan Chen. Light field capturing with lensless cameras. In *IEEE International Conference on Image Processing 2005*, volume 3, pages III–792. IEEE, 2005.

[334] Shuo Zhang, Song Chang, and Youfang Lin. End-to-end light field spatial super-resolution network using multiple epipolar geometry. *IEEE Transactions on Image Processing*, 30:5956–5968, 2021.

[335] Shuo Zhang, Youfang Lin, and Hao Sheng. Residual networks for light field image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11046–11055, 2019.

[336] Xujing Zhang, Sean Braley, Calvin Rubens, Timothy Merritt, and Roel Vertegaal. LightBee: A self-levitating light field display for hologrammatic telepresence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2019.

[337] Zhenghui Zhao, Shanshe Wang, Chuanmin Jia, Xinfeng Zhang, Siwei Ma, and Jiansheng Yang. Light field image compression based on deep learning. In *2018 IEEE International conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2018.

[338] Feng Zhou, Henry Been-Lirn Duh, and Mark Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 193–202. IEEE, 2008.

[339] Xiaoqing Zhu, Anne Aaron, and Bernd Girod. Distributed compression for large camera arrays. In *IEEE Workshop on Statistical Signal Processing, 2003*, pages 30–33. IEEE, 2003.

[340] Adam Zizien and Karel Fliegel. LFDD: Light field image dataset for performance evaluation of objective quality metrics. In *Applications of Digital Image Processing XLIII*, volume 11510, pages 671–683. SPIE, 2020.