



Kent Academic Repository

Khan, Ali, Hossain, Md. Moinul, Covaci, Alexandra, Sirlantzis, Konstantinos and Qi, Qi (2025) *Realistic object reconstruction under different depths through light field imaging for virtual reality*. IET Image Processing, 19 (1). ISSN 1751-9659.

Downloaded from

<https://kar.kent.ac.uk/109844/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1049/ipr2.70099>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts



If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

ORIGINAL RESEARCH OPEN ACCESS

Realistic Object Reconstruction Under Different Depths Through Light Field Imaging for Virtual Reality

Ali Khan¹  | Md. Moinul Hossain¹  | Alexandra Covaci¹ | Konstantinos Sirlantzis² | Qi Qi¹
¹School of Engineering, Mathematics and Physics, University of Kent, Canterbury, Kent, UK | ²School of Engineering, Technology and Design, Canterbury Christ Church University, Canterbury, Kent, UK

Correspondence: Md. Moinul Hossain (m.hossain@kent.ac.uk)

Received: 24 February 2025 | **Revised:** 16 April 2025 | **Accepted:** 24 April 2025

Funding: This research is supported by the European Regional Development Fund (Grant No. 2S05-038).

ABSTRACT

Virtual reality (VR) immerses users in digital environments and is used in various applications. VR content is created using either computer-generated or conventional imaging. However, conventional imaging captures only 2D spatial information, which limits the realism of VR content. Advanced technologies like light field (LF) imaging can overcome this limitation by capturing both 2D spatial and 2D angular information in 4D LF images. This paper proposes a depth reconstruction model through LF imaging to aid in creating realistic VR content. Comprehensive calibrations are performed, including adjustments for camera parameters, depth calibration, and field of view (FOV) estimation. Aberration corrections, like distortion and vignetting effect correction, are conducted to enhance the quality of the reconstruction. To achieve realistic scene reconstruction, experiments were conducted by setting up a scenario with multiple objects positioned at three different depths. Quality assessments were carried out to evaluate the reconstruction quality across these varying depths. The results demonstrate that depth reconstruction quality improves with the proposed method. It also indicates that the model reduces LF image size and processing time. The depth images reconstructed by the proposed model have the potential to generate realistic VR content and can also facilitate the integration of refocusing capabilities within VR environments.

1 | Introduction

The surge in the popularity of VR and its extensive applications across various domains, including education [1, 2] and industries such as manufacturing [3, 4], mining [5], entertainment [6], healthcare [7, 8], and training, [9, 10] has heightened the demand for high-quality VR content that provides comfortable and immersive experiences. The virtual environment can either be computer-generated or photorealistic. Presently, VR content is predominantly created by computer-generated [11, 12] or photographic [13, 14] content acquired through tradi-

tional photography methods such as 360° or omnidirectional video [15]. Computer-generated VR content offers navigation and interactivity capabilities but lacks photorealism. Conversely, traditional photography-based photorealistic VR content, derived from conventional photographic sources, can only be viewed from a fixed position, allowing the user only to rotate their head within the VR scene, while lacking depth perception due to insufficient captured data of the realistic scene.

For a realistic VR experience, perceiving depth in VR content is important. Integrating depth perception in the form of refocusing

Abbreviations: 3D, Three-dimensional; DOF, Degrees of freedom; BIQA, Blind image quality assessment; FOV, Field of view; FRIQA, Full-reference image quality assessment; GPU, Graphics processing unit; HMD, Head-mounted display; LF, Light field; LFC, Light field camera; LFI, Light field image; LM, Layered meshes; LoV, Laplacian of variance; MLA, Microlens array; MPI, Multi-plane image; MSI, Multi-sphere images; ROI, Region of interest; VR, Virtual reality.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

ability in VR content allows users to adjust their focus and refocus on different objects at different depths within the VR scene to replicate the human vision system effectively [16]. The refocusing ability is absent in the current photorealistic VR content [17, 18], which restricts the realism of VR content, deviating from the natural visual capabilities encountered in real-life scenarios [19, 20]. Consequently, the integration of refocusing capabilities becomes essential for providing a real-life experience [20, 21]. Achieving a lifelike VR experience necessitates imaging technology that can capture additional information along with spatial information of the real-world scene [22]. That additional information can be used for realistic scene reconstruction to create VR content to facilitate enhanced interaction, including refocusing capabilities.

LF imaging simultaneously captures 2D spatial information of a real-world scene along with the 2D angular information in a 4D LFI [23]. In contrast, conventional imaging systems capture only the intensities of light rays, which provide only spatial information in the form of 2D images [24]. The LF imaging to acquire additional directional information has garnered significant attention from researchers, leading to its utilization across different applications, ranging from entertainment [18, 25] and computer vision [26, 27] to industrial contexts [28, 29]. In the context of VR, LF imaging techniques offer advantages over traditional photography due to the inclusion of this additional directional information [30, 31]. This additional information aids in reconstructing scenes at various depths, which can provide a refocusing ability in VR content creation, resulting in enhanced VR content with novel interaction ability.

LF acquisition can be achieved through various methods such as camera array-based LF [32] and MLA-based LFC [33]. In the camera array-based approach, multiple cameras or image sensors [32, 34, 35] are used to capture LFIs. Camera array-based systems yield high spatial resolution images by utilizing multiple cameras or image sensors for spatial information acquisition. However, they offer limited angular information due to the limited number of cameras or image sensors in the acquisition setup. Conversely, the MLA-based LFC incorporates an MLA between the main lens and the image sensor, which captures directional information. MLA-based LFCs provide lower spatial resolution because only one image sensor is used in the acquisition. However, they offer superior resolution for angular information due to the MLA. The trade-off between angular information and spatial resolution in LF acquisition affects the reconstruction quality, as reconstruction always benefits from both high angular and spatial information. In the context of VR content creation, higher angular information aids in accurate depth reconstruction, while high spatial information helps in providing good-quality images for a better understanding of the scene.

Several studies have explored the application of LF imaging technology in the creation of VR content. For instance, Milliron et al. [36] introduced the lytro immerge system to capture LF data and reconstruct it for VR with six DoF. They utilized a large camera array for LF acquisition. Yu et al. [30] presented the Plex VR, which employs two types of camera systems to capture inward and outward LF. For outward acquisition, the DPCA system was used, which is composed of 140 cameras (80 static and 60 dynamic) capable of capturing LFIs of human-

sized objects at 60 frames per second. Subsequently, a dome was created to acquire the static objects' geometry and surface reflections. Overbeck et al. [17] proposed a VR system based on LF imaging technology and addressed the challenges, such as LF acquisition, compression, and associated quality. They employ per-view geometry and a disk-based blending field for reconstruction. The system was limited to still photography and produces ghost artifacts in dynamic scenes. Broxton et al. [18] introduced an immersive LF video system utilizing a camera array of 47, mounted on a 92 cm diameter hemisphere [37]. A DeepView interpolation algorithm [38] was used to replace the MPI and represent the scene with MSI. The MSI layers are further reduced to a fixed number of red, green, and blue (RGB) and depth layers without losing the visual quality. The resultant data is then compressed using texture aliasing and video compression techniques for rendering in VR. However, some parts of the reconstructed content [18] show ghosting artefacts, such as DOG [18], where the front area of the wire gauze looks blurred, such as GOATS [18], where ghosting artifacts occur on moving heads around the scene. Debevec et al. [39] proposed a system based on traditional cameras to capture LF for VR. This system employs a single Canon 5D Mark III DSLR camera with an 8 mm fisheye lens and a motorised pan/tilt head, achieving a 75 Hz frame rate. However, capturing images from different camera positions through a rotational mechanism is time-consuming and unsuitable for dynamic scenes due to the slow process, leading to motion artefacts if objects or lighting change.

Additionally, the system requires precise and complex motorized control, where any failure can compromise the accuracy and quality of the reconstruction. Although the above studies have achieved motion parallax in the form of 6DOF, they lack refocusing ability, which is important to provide depth perception in VR content [30]. Moreover, photorealistic VR content creation is based on camera array-based systems, which require a complex setup and calibration. Also, heavy data generation makes them computationally expensive. Additionally, camera array-based systems suffer from limited angular information due to the restricted number of image sensors.

Various studies also utilized point cloud reconstruction [40–42] for scene or object reconstruction. These methods typically involve the use of MLA-based cameras to capture LF data, followed by the generation of depth maps from central images. Subsequently, these depth maps are enhanced, and point clouds are generated from the data to reconstruct the scene or object. However, the reconstructed scenes do not meet the standards required for VR content (which demands high-quality reconstruction that accurately replicates realistic scenes). Point cloud-based reconstruction often suffers from poor quality [42].

Deep learning (DL) approaches are also employed in various applications, including depth reconstruction [43–47] and novel view synthesis [6]. These studies possess the potential to contribute to leveraging the advantages of LF imaging for VR content creation. However, these methods present several challenges. For instance, up-sampling angular dimensions by creating novel views [48] prior to depth reconstruction, can result in ghosting artifacts [49], leading to compromised accuracy and visual quality of the reconstructed images. Furthermore, DL models require extensive training on large-scale datasets, which may not always

be readily accessible for LF imaging. Additionally, these models demand substantial computational resources, rendering them impractical for real-time applications or resource-constrained environments. Notably, most of the DL methods for VR content creation are based on camera array-based LF acquisition systems, which encounter difficulties in reconstructing occluded regions due to their limited angular views.

To achieve high-quality VR content, it is crucial that the content is sufficiently realistic to be believable and possesses interactive capabilities, such as refocusing. Without these features, the VR experience may seem artificial and fail to deliver the desired level of realism and interactivity. Furthermore, significant calibrations are required to calculate the LFC intrinsic and extrinsic parameters and depth parameters for realistic scene reconstruction. To create a wider view of VR content creation, it is also important to calculate the camera's field of view (FOV) because the FOV of the Illum LFC is unknown under different focal lengths and depth settings.

This study proposes a model to reconstruct realistic objects under different depths based on LF imaging for VR content creation. To accurately implement the proposed model, several challenges related to LF imaging are investigated to reconstruct a realistic scene under different depths. For example,

- One of the challenges is vignetting effects, where certain angular images appear dark or with very low intensity. If the vignetting effects are associated with the images, it degrades the overall reconstruction quality and introduces visible artefacts and inconsistencies. This disrupts an immersive experience and reduces user engagement in VR. To address this, a vignetting correction technique is introduced.
- Another challenge is LF calibration, which is essential for estimating camera parameters, such as intrinsic and extrinsic parameters along with camera key parameters. To generate VR content, calibrating the camera's FOV is crucial, as it can be used to provide a wider FOV in the VR environment.
- The LF imaging system usually exhibits distortion, which can alter the size and shape of objects. This impacts the reconstruction quality. In a VR environment, such distortions disrupt realism and immersion, resulting in a less convincing and potentially uncomfortable user experience. To mitigate this issue, a distortion correction technique is also introduced to improve the depth reconstruction quality.
- Large data generated by LF imaging creates difficulty in managing and processing effectively. The proposed model mitigates this issue but also reduces the LFI size without sacrificing the original image quality. The proposed model is also optimized by minimizing dependencies, such as eliminating reliance on the pre-calibrated depth curve. The overall optimization reduces the size and processing time, contributing to more efficient and practical implementation in VR environments.

Additionally, an experimental scenario is established using various realistic objects at different depths and FOVs to evaluate the proposed depth reconstruction model. Assessment criteria, including BIQA and FRIQA, are used to measure the quality

of the reconstruction. The depth-reconstructed images are then combined according to the capture sequence and transformed into VR content using the Unity game engine. The proposed model is discussed in detail, and the results obtained from this study are presented and analysed.

Existing literature indicates that depth perception has the potential to enhance immersion [50–52] and reduce motion sickness [51, 53] in VR content. However, it is important to acknowledge that this study primarily focuses on the technical aspects of the proposed depth reconstruction model. The current work does not include experimental evaluations of user experiences, such as perceived immersion or motion sickness. As ongoing research, user experience assessment will be evaluated in the subsequent phase. This may involve deploying the reconstructed realistic scenes in practical VR applications such as virtual museum environments and conducting thorough user feedback to assess immersion levels and motion sickness effects.

2 | Methodology

Figure 1 presents the technical strategy of the proposed methodology for depth reconstruction which comprises: (1) LF acquisition and the decoding process involves pre-processing LF data, including transforming the initially captured 2D raw data into a 4D format; (2) Depth reconstruction utilizing the decoded 4D LFIs under different depths and FOVs; and (3) VR translation i.e., translating the depth-reconstructed scene into VR content.

2.1 | LF Imaging

LF imaging captures information of a 3D realistic scene simultaneously in the form of a 4D LFI (i.e., 2D angular and 2D spatial) [54, 55]. Initially, the captured LFI of the 3D scene exists in a 2D raw-format which can then be decoded to extract 4D information in the form of 4D LFI. In LFC, to capture the 2D angular information, the MLA samples the light rays before striking the image sensor. In conventional imaging, the light rays strike the image sensor after passing the main lens, and only capture spatial information of the light rays [56]. The difference between LF imaging and conventional imaging is shown in Figure 2.

LF can be represented by the 4D plenoptic function $L(u, v, s, t)$ [54], where (u, v) represents the directional resolution and (s, t) represents 2D spatial resolution. A two-plane parameterization (TPP) method [57, 58] is used to characterize the LF in the form of 4D. In TPP, the light rays traverse between two parallel planes such as the aperture and the image planes [59]. The light rays originating from the scene initially intersect the (u, v) plane before striking the (s, t) plane. Levoy et al. [54] explained that any point on the sensor plane (s, t) receives light from multiple rays passing through various points on the aperture plane (u, v) . The intensity at any given point (s, t) on the sensor plane is determined by the cumulative effect of all these rays. This can be mathematically expressed as:

$$E_f(x, y) = \frac{1}{f^2} \iint LF(u, v, s, t) A(u, v) \cos^4 \theta du dv \quad (1)$$

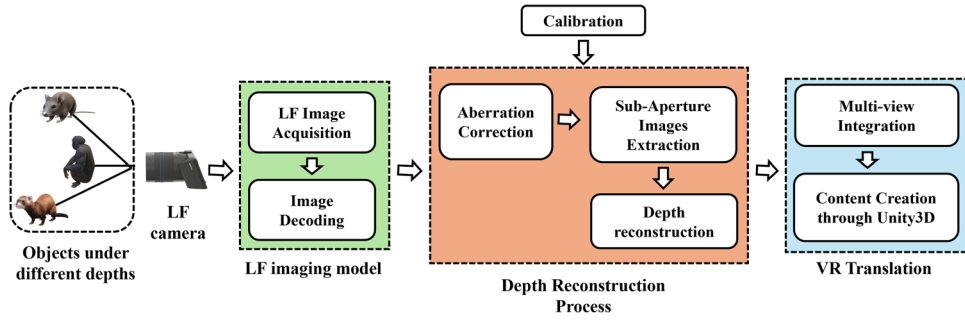


FIGURE 1 | Overview of the proposed technical strategy for depth reconstruction.

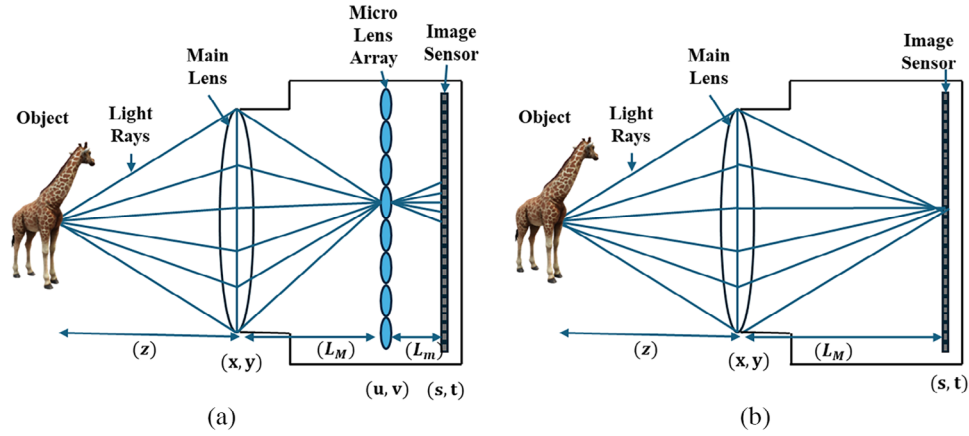


FIGURE 2 | Illustrates the difference between the (a) LFC equipped with a MLA; and (b) conventional camera.

where $E_f(x, y)$ represents the intensity at a point (s, t) on the sensor plane, f represents the separation between the aperture and the sensor planes. LF represents the ray traveling from the position (u, v) to (s, t) . $A(u, v)$ denotes the aperture and θ presents the incident ray angle. The term $\cos^4 \theta$ accounts for the natural vignetting effect of light rays [60]. To simplify Equation (1), assuming that $A(u, v) = 1$ by considering $E_f(x, y)$ as zero outside the aperture and neglecting the natural vignetting effect, can be written as

$$E_f(x, y) = \frac{1}{f^2} \iint LF(u, v, s, t) du dv \quad (2)$$

2.1.1 | LF Decoding

Figure 3 shows an example of a raw LFI and its microimage under each microlens [Figure 3(b)]. Once the raw-LFI is acquired, the raw LFIs need to be decoded to convert the 2D images into 4D LFIs. The decoding process interprets the information encoded by light rays on the image sensor in the form of 4D LF data $LF(u, v, s, t)$. The TPP can be used to parameterize the light rays that emerge from the 3D scene in the form of 4D information (u, v, s, t) . The block diagram of the LF decoding process is shown in Figure 3(d). Before initiating the decoding process, the grid parameters including the horizontal, vertical, and rotational offsets are determined. To determine these parameters, the centers of microimages can be identified by using a white image, as shown in Figure 3(c). Where the

brightest spot in each microimage approximates its center. The LF decoding process starts with demosaicing the raw LFI and then correcting the vignetting by dividing the white image, typically captured through a diffuser or from a white scene. In the raw LFI, the microimages are non-integer positioned and rotated relative to the pixel grid. To align the microimages, all of microimages are then resampled, rotated and scaled so that their centers align with the centers of the pixels. However, uniform scaling is not achievable due to the hexagonal structure of the microlens array (MLA) [61], where the microlenses are arranged in a honeycomb pattern. This arrangement results in unequal vertical and horizontal spacing. Converting this hexagonal grid into a square grid requires proper alignment of the vertical and horizontal axes. When scaling is non-uniform, the resulting grid becomes rectangular rather than square. Subsequently, the LFI image is sliced into identical segments to convert the 2D LFI into a 4D representation. A 1D interpolation is then performed along the spatial axis to de-hex the sampled LFI. Another 1D interpolation is performed along the angular axis to correct the rectangular pixels in each microimage to make them square. More detailed information about the decoding model can be found in [61].

2.1.2 | LF Refocusing

LF refocusing enables depth reconstruction [62] of realistic scenes under different depths. The refocusing can be achieved by capturing the angular information of light rays and facilitating the virtual adjustment of the focal plane [63]. The Ray Space model

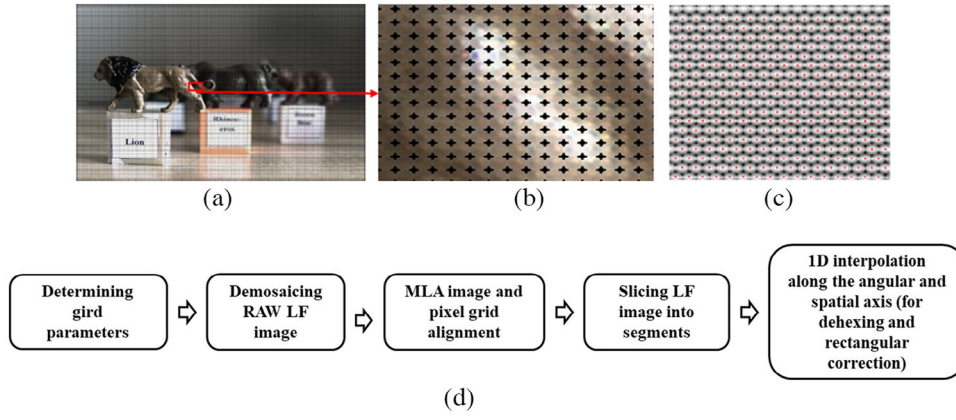


FIGURE 3 | LF decoding process (a) raw LFI; (b) cropped raw LFI showing microimage under each microlens; (c) white image with the estimated center position of each microimage; and (d) flow chart of the LF decoding process.

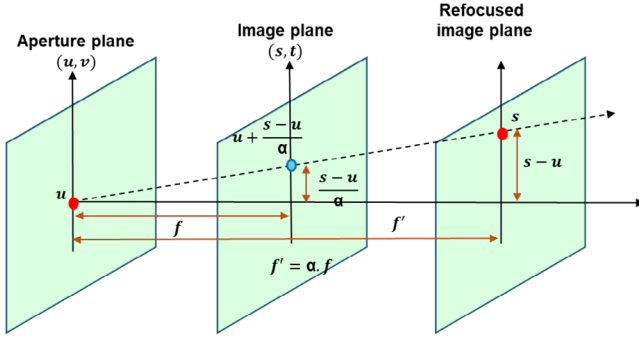


FIGURE 4 | Principle of the ray space model based on LF refocusing.

(as shown in Figure 4) provides the mathematical parameterization of light rays for LF refocusing by incorporating the aperture (u, v) and the image planes (s, t) [44]. Since all potential focus planes are parallel, the new focus plane f' (i.e., refocused image plane) can be computed by multiplying the refocusing parameter α with the existing image plane (sensor plane), which is located at a distance f from the aperture plane. Thus, the new focus plane f' becomes $f' = \alpha f$. By rearranging Equation (2), the LF refocusing can be obtained through:

$$E_f'(x, y) = \frac{1}{f'^2} \iint LF \left(u + \frac{s-u}{\alpha}, v + \frac{t-v}{\alpha}, u, v \right) dudv \quad (3)$$

where $E_f'(x, y)$ represents the refocused LFI in which the entire LFI is shifted by a factor of the refocusing parameter α .

2.2 | Depth Reconstruction Model

The proposed depth reconstruction model involves (a) aberration correction; (b) sub-aperture image extraction; and (c) depth reconstruction. Before performing the sub-aperture image extraction, it is crucial to correct the decoded LF 4D image to resolve aberrations such as distortion and vignetting effects. The detailed description of the depth reconstruction model is discussed below.

2.2.1 | Aberration Correction

The decoded LFI often contains aberrations such as distortion and vignetting effects, primarily caused by the arrangement of lenses and the sensor. One pertains to the distortion, whereby objects depicted in images appear slightly distorted. The other aberration is the vignetting effect, which impacts the sub-aperture images generated from the edge pixels of microimages. These aberrations affect the depth reconstruction quality and often create visual inaccuracies that can disrupt the realism of the VR content. Also, it leads to a less convincing and uncomfortable VR experience and reduces the overall effectiveness of the VR application. A distortion correction algorithm is utilized to rectify the distortions present in LFIs before they are used for the depth reconstruction process. Second-order radial distortion is applied based on the lens distortion model [64] and is defined as:

$$\begin{aligned} s_d &= s_{ud}(1 + k_1 r^2 + k_2 r^4) \\ t_d &= t_{ud}(1 + k_1 r^2 + k_2 r^4) \end{aligned} \quad (4)$$

where the s_d and t_d represent the distorted image coordinates that are formed by the projection of real-world points. The s_{ud} and t_{ud} represent the undistorted image coordinates and k_1, k_2 are the radial distortion coefficients. r is the radial distance, where $r^2 = \sqrt{(s_d - s_c)^2 + (t_d - t_c)^2}$ and s_c, t_c are the center of distortion [65].

After rearranging Equation (4), the distortion correction can be defined as:

$$\begin{aligned} s_{ud} &= \frac{s_d}{(1 + k_1 r^2 + k_2 r^4)} \\ t_{ud} &= \frac{t_d}{(1 + k_1 r^2 + k_2 r^4)} \end{aligned} \quad (5)$$

The vignetting effect is a common issue in LF imaging, where certain sub-aperture images appear darker or of lower intensity. In this study, a vignetting correction method is introduced for the LFC, which involves neglecting the sub-aperture images at the edges using Equation (6). This neglecting process removes the sub-aperture images which are in the first row and first column (where u and $v = 1$) the last row, and last column (where u and

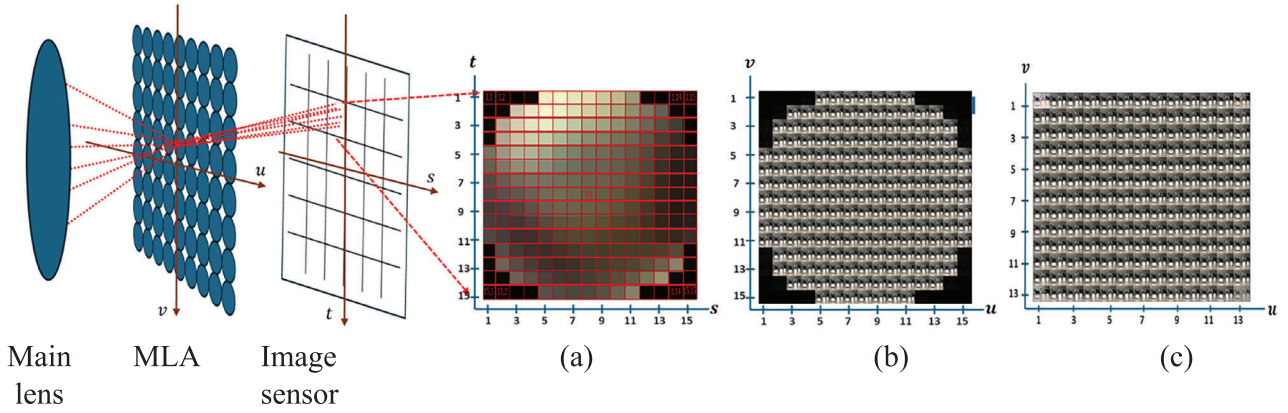


FIGURE 5 | Illustration of the sub-aperture image formation, an example of (a) microimages; (b) original sub-aperture images; and (c) corrected sub-aperture images.

$v = 15$), where most of the sub-aperture images are affected by vignetting, as shown in Figure 5(b). After neglecting the sub-aperture images at corners, there were still four sub-aperture images (sub-aperture at each corner can be seen in Figure 9(h)) that appeared with low intensity, which is corrected using a haze removal technique [66].

$$E_{f(u,v)}(x, y) = \frac{1}{f^2} \int_{u=2}^{14} \int_{v=2}^{14} LF(u, v, s, t) dudv \quad (6)$$

where $E_{f(u,v)}(x, y)$ represents the corrected LFI and u and v are the indices of the sub-aperture image.

2.2.2 | Sub-Aperture Images Extraction

Once the decoded LFIs are corrected, the sub-aperture images (also known as sub-aperture views) are extracted by:

$$I_{(u,v)}(x, y) = LF(u, v, s, t) \quad (7)$$

$I_{(u,v)}(x, y)$ is the sub-aperture image which is formed by collecting all the rays passing through the u^{th} and v^{th} position of each microlens. Figure 5 illustrates the process of sub-aperture image formation and an example of original and corrected sub-aperture images. Initially, the u and v of LFI are $u \in \{u | 1 \leq u \leq 15\}$ and $v \in \{v | 1 \leq v \leq 15\}$. Thus, a total of 15×15 rays can pass through each microlens to form a microimage on the sensor, which has a resolution of 15×15 pixels, as illustrated in Figure 5(a). Each pixel in the microimage represents a ray arriving from different angles. Each sub-aperture image is generated by integrating the pixels from each microimage, corresponding to light rays arriving from the same direction, forming a new image array using Equation (7). The resolution of each sub-aperture image corresponds to the number of microlenses in the MLA [67], while the number of sub-aperture images is equivalent to the number of pixels in each microimage [67]. These sub-aperture images provide views of the same scene from slightly different perspectives. The original extracted sub-aperture images possess distortion due to the main lens. The distortion correction discussed in Section 2.2.1 is applied to all sub-aperture images to correct the distortion. After correction, the sub-aperture images are reduced to 13×13 , as shown in Figure 5(c).

2.2.3 | Depth Reconstruction

In the depth reconstruction process, the corrected sub-aperture images are utilized to generate multiple depth images from a single exposure of LFC. During depth reconstruction, the extracted sub-aperture images are shifted to the desired depth plane and then summed together. This results in the depth reconstructed image of the desired depth plane. In this depth-reconstructed image, the objects placed at the desired depth plane can be focused, and others will be blurred. The technique of depth reconstruction is referred to as the shift-sum method, and it can be expressed by:

$$I(x, y) = \sum_{u=1}^n \sum_{v=1}^m I_{(u,v)}(x, y) \quad (8)$$

Where $I_{(u,v)}(x, y)$ represents the sub-aperture image. The n and m represent the total sub-aperture images extracted from the corrected LFI. In this study $n \times m = 13 \times 13$. $I(x, y)$ signifies the depth-reconstructed image corresponding to the depth plane at which the LFI was originally captured. However, to refocus the LFI at a specific depth plane, it is necessary to shift the sub-aperture images, as illustrated in the ray space model [Figure 4]. When the MLA is positioned x_o from the main lens, only the object located z_o will be focused, as illustrated in Figure 6. Here, z represents the distance between depth 2 and the camera lens, also referred to as the refocusing distance. As discussed in Section 2.1.2, the focus of the LFI can be altered, and objects can be reconstructed at different depths by virtually adjusting the sensor position by using Equation (3). To reconstruct the depth image at the desired depth plane, Equation (8) can be reformulated as [55]:

$$(x, y) = \sum_{u=1}^n \sum_{v=1}^m I_{(u,v)}(x, y) \quad (9)$$

Where $I(x, y; \alpha)$ represents the depth-reconstructed image at the desired depth plane, and α is the refocusing parameter. For example, an object placed at depth 2 in the scene is at a distance z from the camera lens, and the same object x represents the focal plane at which it is in focus. Based on the refocusing principle, as shown in Figure 4, the refocusing parameter α can be estimated

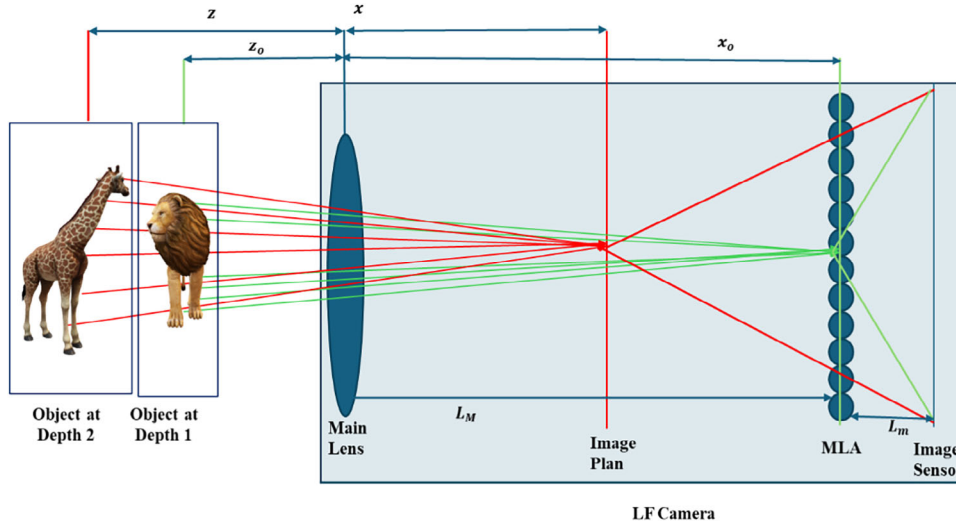


FIGURE 6 | Illustration of LF depth reconstruction, the captured LFI which is initially focused at the depth distance of z_o , can be virtually refocused to reconstruct the depth image at a distance of z by adjusting the image plane using the refocusing parameter (α).

for the object at depth 2 as:

$$x = \alpha x_o \quad (10)$$

Where $\alpha = x/x_o$ and when $\alpha = 0$, the refocusing distance remains the same as the original distance, $z = z_o$ implying that no refocusing occurs. Moreover, it is evident from Equations (9) and (10), that the refocusing distance z is contingent upon the value of α , where $\alpha > 0$, $z > z_o$ and $\alpha < 0$, $z < z_o$. To reconstruct the specific depth image within a realistic 3D scene, it is crucial to determine the optimal value α for that specific depth. The α value can then be utilized in Equation (9) to reconstruct the depth image focused on the desired depth plane. To estimate the α value for a depth image that is refocused at a specific object in the scene, focal stack images can be generated using Equation (9). In this study, the focal stack images are generated in a range of $\alpha = [-2 : 0.21 : 2]$. The range is considered arbitrary, and Equation (11) computes the α value of the focused object for a specific depth plane. The focal stack images comprise a series of depth images.

Consider a scenario where three objects in the scene are placed at different depth levels. The $ROI_{i,j}$ is selected subsequently from each object in the scene. A greedy search method is used to find the optimal value α_j that reconstructs the refocused image $ROI_{i,j}$ with maximum focus $T_{i,j}$:

$$\alpha_j = \arg \max_a T_{i,j}(a) \quad (11)$$

where i represents the index of the image in the focal stack and j represents the index of the object. Once the α_j is obtained, it can be employed in Equation (9) to reconstruct the depth image with the j^{th} object refocused. To estimate the focus quality, the Tenengrad function [68], is used which is defined as:

$$T_{i,j} = \frac{1}{N} \sum_0^x \sum_0^y [G]^2 \quad (12)$$

Where $T_{i,j}$ represents the sharpness of the depth-reconstructed image at the ROI, N represents the number of pixels in the ROI and G represents the magnitude of the $ROI_{i,j}$.

Additionally, the α_j can be utilized to reconstruct the objects, wherein the objects are positioned at the refocusing distance z_j . The refocusing distance z_j corresponds to the distance of an object in the 3D scene from the camera's main lens. This distance can either be measured physically or obtained through a calibration technique [56]. Additionally, the α_j can be utilized to reconstruct the objects, wherein the objects are positioned at the refocusing distance z_j .

2.2.4 | Multi-View Integration

In VR, it is important to generate the content from a wider FOV, which can provide flexibility to the user in navigating a wide area. Additionally, it can enhance the VR user experience. In this study, to generate the content with a wider FOV, three angular views of the scene are captured by rotating the camera to the left and right from a straight direction, as discussed in Section 4. These three FOVs are determined based on the calibration. The depth images are reconstructed from each angular view. The reconstructed depth images are then combined based on the sequence of captured images for VR translation. The combination is performed in such a way that each depth of all three views provides a wider FOV, thus enabling the ability to refocus on different depth levels within the VR scene. The detailed procedure is discussed in Section 4.

2.3 | Quality Assessment

Two types of quality assessment methods are used to evaluate the reconstruction quality and of the proposed model. These are BIQA and FRIQA [69], and are described as follows:

2.3.1 | BIQA

The BIQA assesses the reconstruction quality without the reference [70]. The BIQA relies on focus estimation (in terms of sharpness and intensity) and incorporates four metrics, such as Tenengrad (T) [68], LoV (δ) [71], Gradient magnitude (G) [72] [these three measure the sharpness of the refocused object], and grey level pixel intensity (In) [measures the intensity of the refocused object]. These metrics are utilized to evaluate the quality of the refocused object within the reconstructed depth image and are compared with the original, uncorrected, and corrected LFIs. For each reconstructed depth image, the ROI corresponding to each target object (refocused) is extracted, and each metric is applied to the respective ROI. Once the four metric values are computed, Equation (13) is used to calculate the difference between the metric's values obtained from the original, corrected, and uncorrected LFIs. A positive $Diff$ indicates an improvement, whereas a negative $Diff$ indicates a degradation of reconstruction quality.

$$Diff (\%) = \frac{M_{corr_value} / M_{uncorr_value} - M_{orig_value}}{M_{orig_value}} \times 100 \quad (13)$$

Where $M_{corr_value} / M_{uncorr_value}$ represents the BIQA metric value estimated from the depth image reconstructed using corrected or uncorrected LFI, respectively. M_{orig_value} represents the BIQA metric value estimated from the depth image reconstructed using the original LFI.

2.3.2 | FRIQA

The FRIQA uses a peak signal-to-noise ratio (PSNR) [73] and a structural similarity index (SSIM) [74] to evaluate the quality of depth-reconstructed images [75]. These metrics are used to compare the performance of the proposed model with frequency-based depth reconstruction methods, such as 4D planar and 4D hyperfan filters [76]. The ground truth is a refocused image of a specific depth plane generated by shifting and averaging [62] using the originally captured LFI.

3 | System Calibration

Systems calibration is a crucial part of the proposed depth reconstruction and comprises several steps, such as camera calibration, which estimates the intrinsic and extrinsic parameters along with the camera's key parameters. The depth calibration estimates the minimum and maximum depth reconstruction range, along with refocusing intervals, to determine the number of depth images that can be reconstructed within the estimated depth range. Furthermore, (a) image pixel size; (b) FOV of the camera; and (c) angular difference between the sub-aperture images need to be estimated through calibrations. The FOV estimation is essential, as it helps to determine the angle at which the camera can be rotated to cover the maximum area for VR content. The details of each calibration process have been discussed below.

3.1 | Camera Calibration

The LFC calibration is carried out to estimate the distortion coefficients in addition, to the intrinsic and extrinsic parameters. For calibration, LFIs of a 9×7 checkerboard pattern were captured using a Lytro Illum camera within the depth range of 20 to 80 cm, with an interval of 1 cm. Figure 7 shows the block diagram of the overall camera calibration process. In the proposed depth reconstruction model, sub-aperture images have been used for depth reconstruction. Therefore, the central sub-aperture image has been utilized to obtain the calibration parameters. The central sub-aperture image is generated from the actual projection centers of the MLA. The traditional camera calibration approach applies to LF cameras because the central sub-aperture image can be treated as a conventional 2D image, captured through a reduced aperture [33, 55], similar to a traditional camera image. A pinhole camera model has been employed to establish the relationship between real-world 3D points and their projections onto the image sensor. Since the aperture does not affect the parameters of the pinhole projection model [55], a traditional camera calibration method [77] is applied to estimate the intrinsic and extrinsic matrix and distortion parameters of the main lens. This approach is widely adopted for calibrating LF cameras [55, 56, 78], ensuring accurate estimation of projection matrices.

The calibration is performed in two steps, based on Ref. [77]. The first step of transformation from the world coordinate systems $P = [x^w, y^w, z^w]^T$ to the camera coordinate systems $X^c = [x^c, y^c, z^c]^T$ is shown in Equation (14).

$$\begin{bmatrix} x^c \\ y^c \\ z^c \\ 1^c \end{bmatrix} = \begin{bmatrix} R_{3 \times 3} & T_{r 3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} x^w \\ y^w \\ z^w \\ 1^w \end{bmatrix} \quad (14)$$

where $R \in \mathbb{R}^{3 \times 3}$ is the rotation matrix that represents the orientation of the object in radians, and $T_r \in \mathbb{R}^{3 \times 1}$ is the translation matrix that represents the position of the object in millimeters (mm). In the second step, the camera coordinates X^c are transformed into image coordinates S , where $S = [s, t]^T$:

$$\begin{bmatrix} s \\ t \\ \omega \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x^c \\ y^c \\ z^c \\ 1^c \end{bmatrix} \quad (15)$$

where the f_x and f_y represent the focal length of the camera in terms of pixel dimensions [59] in the x and y directions. The c_x and c_y represent the principal point or optical center coordinates, where the ray pierces the image plane. By simplifying the Equations (14) and (15), the transformation from world coordinates point P to image coordinates point S can be written as (also known as the pinhole projection model):

$$\omega [S, 1]^T = KM[P, 1]^T \quad (16)$$

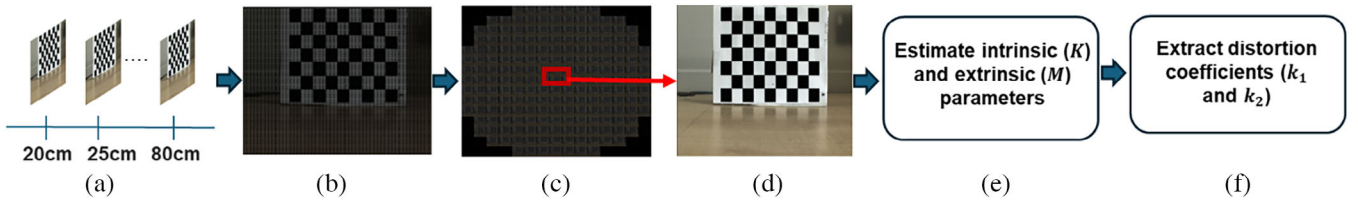


FIGURE 7 | Overview of the camera calibration process (a) example of LFIs under different depths; (b) raw LFI; (c) sub-aperture images; (d) central sub-aperture image; (e) estimation of camera parameters; and (f) obtaining distortion coefficients.

TABLE 1 | Intrinsic and extrinsic parameters of the camera calibration.

c_y (Pixels)	f_x, f_y (pixels)	k_1, k_2 (pixels)	R (rad)	T_r (mm)
322.77, 228.75	1126.3, 1122.6	-0.34, 0.64	$\begin{bmatrix} 0.949 & 0.0081 & -0.315 \\ 0.007 & 0.999 & 0.027 \\ 0.315 & -0.026 & 0.948 \end{bmatrix}$	$[-55.67 \ -102.21 \ 835.68]$

where ω is the constant, K is intrinsic and M is the extrinsic matrix and is defined as

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (17)$$

$$M = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix}$$

where $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$ and $T_r = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$ are the rotation and translation matrices that belong to the extrinsic matrix M . It represents the position of the object in the world coordinate system. The matrix K provides the intrinsic parameters of the camera. In this study, Brown's distortion model [79] is used to estimate the distortion coefficients k_1 and k_2 to correct the distortion effect. Table 1 presents the intrinsic and extrinsic parameters obtained through camera calibration.

In addition, the key unknown parameters such as L_M , D_M and f_m are also obtained through calibration using Equations (18)–(22). The overview of the known (f_M , z , D_m and L_m) and unknown parameters are shown in Table 2. The distance between the main lens and MLA, L_M can be calculated as follows [80]:

$$\frac{1}{f_M} = \frac{1}{z} + \frac{1}{L_M} \quad (18)$$

$$L_M = \frac{f_M \cdot z}{z - f_M} \quad (19)$$

The diameter of the main lens D_M can be calculated as [82]:

$$D_M = \frac{D_m}{L_m} (L_M + L_m) \quad (20)$$

The f-number of the microlens F_{MLA} is defined as the focal length of the microlens f_m divided by the diameter of the microlens D_m

TABLE 2 | Overview of the key LFC parameters.

Description	Symbol	Values	Unit
Focal length of the main lens	f_M	18.80	mm
The diameter of the main lens	D_M	8.20	mm
Distance between the camera and object	z	835.68	mm
Distance between the main lens and MLA	L_M	19.24	mm
Distance between the MLA and image sensor [64]	L_m	0.047	mm
The diameter of the microlens [81]	D_m	0.20	mm
Number of pixels covered by each microlens	N_p	15×15	Pixels
Number of microlenses	N_M	625 × 434	

[83]:

$$F_{MLA} = \frac{f_m}{D_m} \quad (21)$$

By simplifying Equation (21), f_m can be calculated as:

$$f_m = F_{MLA} \cdot D_m \quad (22)$$

The Lytro Illum has a constant f-number which is $F_{MLA} = 2$ [84]. Note that, the f-numbers of the microlens and the main lens are equal as stated in [33]. If the f-number of the main lens is higher, cropping occurs, resulting in many pixels being black and a consequent waste of resolution. Conversely, if the f-number of the main lens is lower, overlapping occurs between images formed under each microlens.

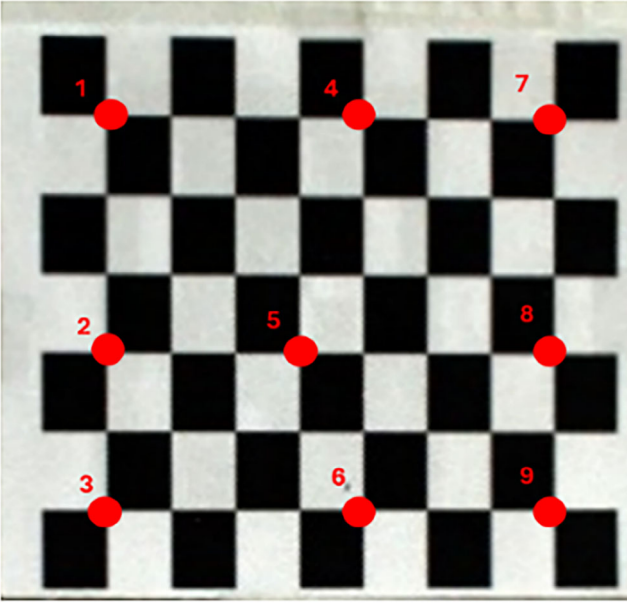


FIGURE 8 | Illustration of the measuring points utilized for the calculation of the radial distortion coefficient.

3.2 | Aberration Correction

In this study, the LF Lytro Illum camera utilized exhibits two aberrations. One pertains to distortion, whereby objects within images appear slightly distorted. The other aberration is the vignetting effect, which affects the sub-aperture images generated from the corners and the pixels that are located at the edge of microimages. These aberrations reduce the quality of depth reconstruction.

3.2.1 | Distortion Correction

A checkerboard is used to correct and assess the distortion between the coordinates of the distorted and undistorted images. The checkerboard images are captured at a depth of 30 cm from the camera. A total of nine measurement points are considered from the checkerboard image for analysis, as shown in Figure 8. The relative radial distortion R between the distorted and undistorted image is calculated for all nine measurement points by:

$$R = \left(\frac{\sqrt{(s_d - s_{ud})^2 + (t_d - t_{ud})^2}}{\sqrt{(s_c - s_{ud})^2 + (t_c - t_{ud})^2}} \right) \times 100 \quad (23)$$

Table 3 illustrates the overall distortion correction between the distorted and undistorted image at each measurement point. It demonstrates that the distortion is higher at the edges of the checkerboard while nearly zero at the center. It can be concluded that objects located near the edges of the image are significantly affected by distortion, whereas those positioned at the center have no distortion. It is also concluded that the objects near the camera occupy a larger portion of the image and are more likely to be closer to the edges, where distortion is more pronounced.

TABLE 3 | Overview of the distortion correction at each measuring point.

Measurement points	Distorted (pixels)	Undistorted (pixels)	Correction: R (%)
1	169.79	171.01	0.71
2	95.91	96.09	0.18
3	66.77	66.82	0.14
4	163.64	164.80	0.70
5	72.71	72.82	0.14
6	51.92	51.97	0.08
7	206.06	208.42	1.13
8	153.15	154.11	0.62
9	137.41	138.14	0.53

TABLE 4 | Comparative assessment of the vignetting correction method.

LFI	Refocused object (hippo)
Original and corrected	$T / G / \delta / In$ (%) 1.69 / 2.47 / 1.32 / 0.94
Original and uncorrected	-3.83 / -2.26 / -7.11 / -1.99
Corrected and uncorrected	5.25 / 4.84 / 9.08 / 3

3.2.2 | Vignetting Correction

To address the vignetting effect, an image of a realistic scene was captured [Figure 9(a)] using the same LFC, and each microimage has a resolution of 15×15 [Figure 9(c)]. Figure 9(b) presents an enlarged view of the ROI from Figure 9(a), highlighting the presence of black or low intensity pixels caused by the vignetting effect, resulting in some sub-aperture images with black or low intensity as illustrated in Figure 9(d). Equation (6) is used to exclude the dark or low-intensity pixels from the captured LFI. Figure 9(e) shows the obtained uncorrected LFI image, where the resolution of each microimage is reduced to 13×13 [Figure 9(g)], from which a total of 13×13 sub-aperture images are extracted [Figure 9(h)]. It can be seen in Figure 9(f and h) that the uncorrected LFI still possesses vignetting. Further reduction of the LFI was avoided, as this could harm the quality of depth reconstruction. The corner sub-aperture images [Figure 9(h)] are corrected using the haze removal technique [66], and the results can be seen in Figure 9(i), (j), (k) and (l), which presents the corrected LFI, enlarged view of the ROI, the microimage of the corrected LFI and sub-aperture images extracted from corrected LFI, respectively. Figure 9(l) demonstrates the corrected LFI.

For comparative study, the original, uncorrected and corrected images were used to reconstruct depth images of the object (Hippo). The BIQA was employed to evaluate the effectiveness of the vignetting correction method. Table 4 demonstrates the quality of the depth reconstructed image of the object (Hippo).

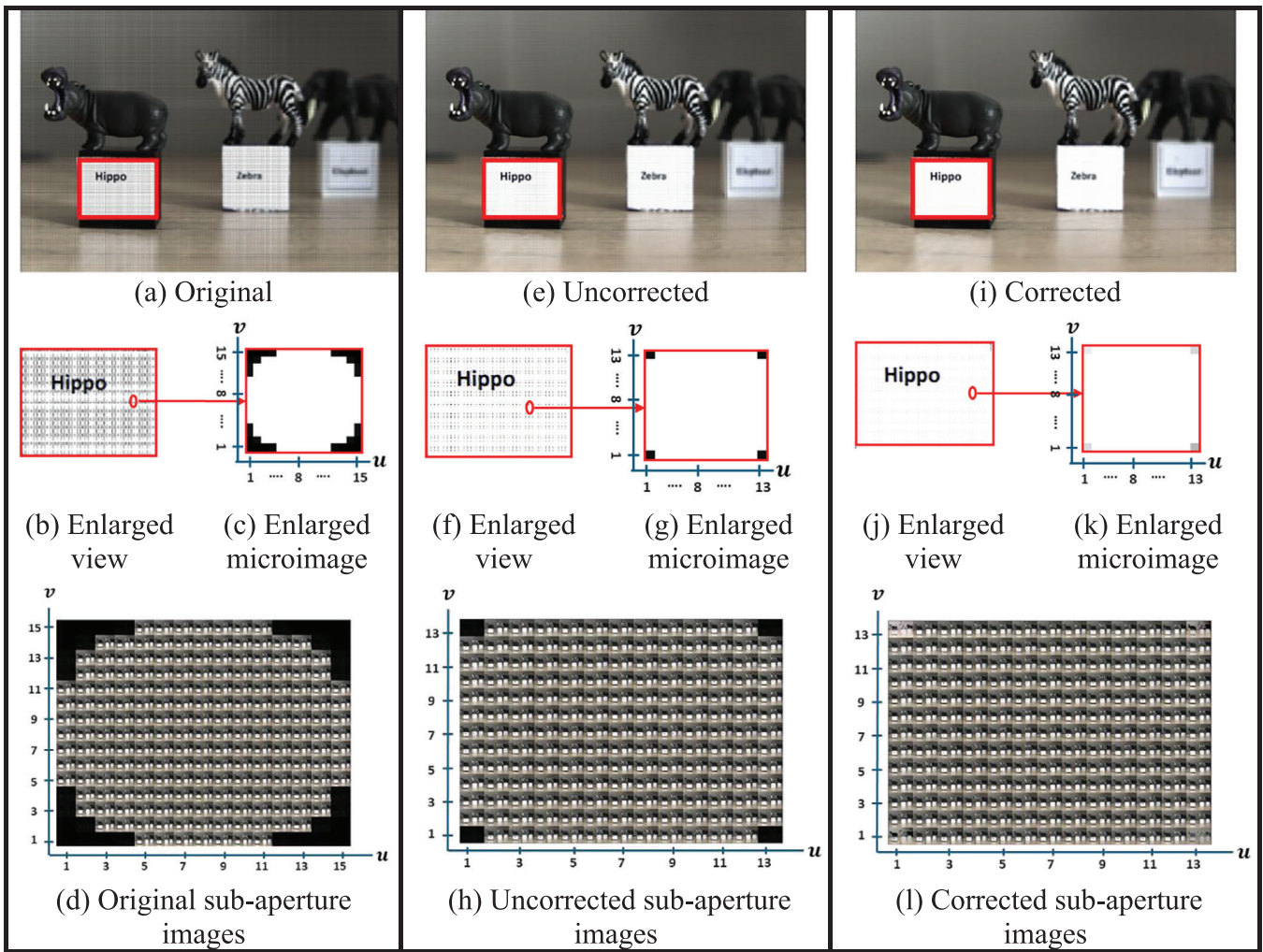


FIGURE 9 | Examples of vignetting correction approach. (a), (e) and (i) represent the original, uncorrected, and corrected LFI, and (d), (h), and (l) represent the sub-aperture images extracted from corresponding LFI.

The refocused quality is improved when the corrected LFI is used compared to both the original and uncorrected LFIs.

3.3 | Depth Calibration

To enable refocusing on each object within the scene, it is necessary to determine the permissible depth distance between the objects, allowing the algorithm to refocus on each object. Additionally, it is imperative to ascertain the maximum achievable depth resolution, which allows the depth reconstruction algorithm to refocus to the maximum depth level effectively. Consequently, depth calibration is performed to ascertain the depth resolution of the reconstruction model. This process involves estimating the maximum refocusing depth and determining the refocusing interval.

To estimate the refocusing intervals, multiple images of a checkerboard were captured under a depth range of 20 cm to 40 cm with a separation of 1 cm, as illustrated in Figure 10 along with the central sub-aperture images extracted from LFIs for each depth level. Subsequently, Equation (11) is used to estimate the refocusing parameter a at each depth level. Figure 11(b) shows the

relationship between the obtained a and T for the depth levels of 20 cm to 40 cm with an interval of 5 cm. In Figure 11(b), the maximum T value from the focal stack of images of each depth identifies the a value of a focused image. Figure 11(a) shows the relationship between the depth and the a . It reveals that objects closer to the camera exhibit narrower depth intervals than those farther away. For instance, objects within the depth of 20 cm to 28 cm demonstrate a shallower depth interval in contrast to objects between 29 cm and 40 cm. It shows that the objects' depth from 20 cm to 22 cm, 23 cm to 25 cm, and 26 cm to 28 cm have their distinct α values. Conversely, the objects within the depth level of 29 cm to 40 cm also exhibit three distinct α values and confirm that the same α can be used within the 5 cm range from 29 cm to 33 cm, 34 cm to 38 cm, as illustrated in Figure 11(a). It can be concluded that a total of six depth images can be reconstructed within the depth range of 20 cm to 40 cm.

Furthermore, to validate the quality of depth reconstructed images at different depth levels, a step-edge method [85] is applied to calculate the slope of the depth reconstructed images under the depth levels of 20 cm, 25 cm, 30 cm, 35 cm, and 40 cm, respectively. Figure 12(a) represents the normalized grayscale intensity of the

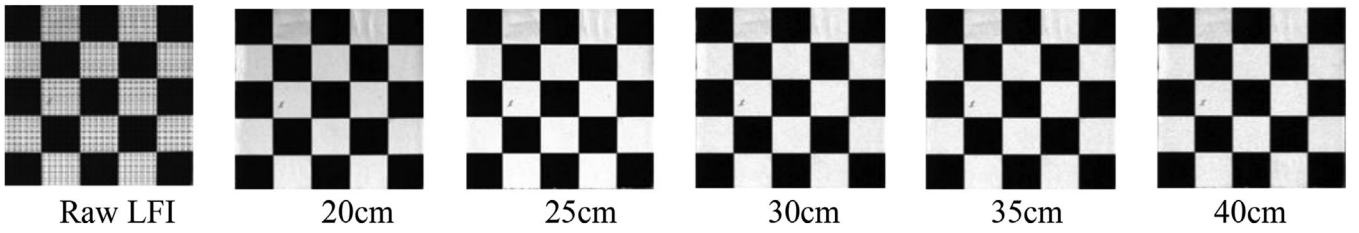


FIGURE 10 | Examples of the captured LFI along with the central sub-aperture images extracted from corresponding LFIs under various depths.

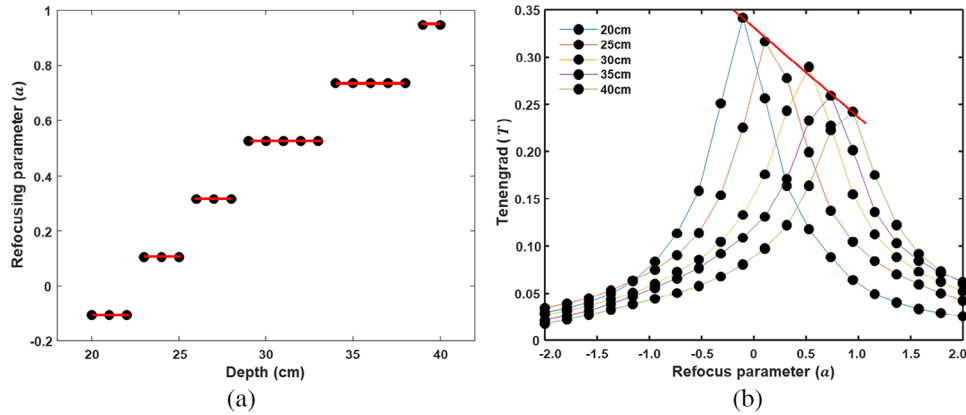


FIGURE 11 | Depth calibration involves establishing (a) a relationship between the depth distance and α and; (b) the relationship between α and T and, estimating α value for each depth.

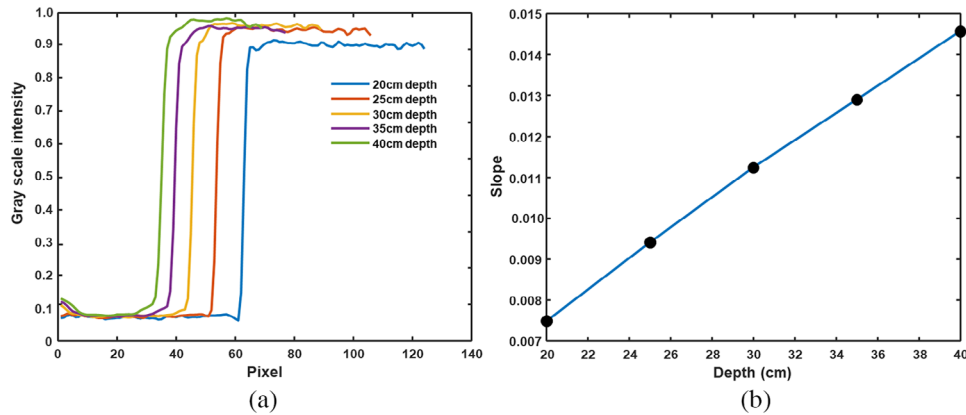


FIGURE 12 | Step-edge-based refocus quality under different depths, (a) the normalized grayscale of the central row of each depth, where 0 represents black pixels and 1 represents white; and (b) the relationship between the depths and slope.

center row of the corresponding depth images. Figure 12(b) shows the relationship between the depth and the slope which reveals that the quality of depth reconstructed images decreases with an increase in depth distance. A depth of 20 cm shows the lowest slope value, and the slope value increases with increasing depth. The minimum slope value represents the sharp image while the maximum slope value represents the more blurred image.

It is crucial to investigate the maximum achievable depth range of the proposed depth-reconstruction algorithm. For this purpose, an experiment was conducted. Multiple images of a square object containing a label with descriptive text [Figure 14 (a)] are captured within the depth range of 20 to 40 cm with an

interval of 1 cm, as illustrated in Figure 13(a). Each captured image is refocused by using the α values obtained through the depth calibration. Subsequently, an ROI comprising descriptive text is cropped from each depth reconstructed image, as demonstrated in Figure 13(b). Equation (13) is then used on all cropped images to evaluate the quality of refocused images (in terms of sharpness) at each depth level. The refocused quality is degraded when the depth increases. The maximum sharpness is observed at a depth of 20 cm, and the text is easily readable. In contrast, at a depth of 40 cm, sharpness decreases significantly, resulting in blur, and the descriptive text becomes unreadable, as depicted in Figure 14. It shows that the better depth resolution of the refocused image can be achieved at depths of 20cm to 30 cm. Beyond this depth, the descriptive text becomes blurred significantly, and it is difficult

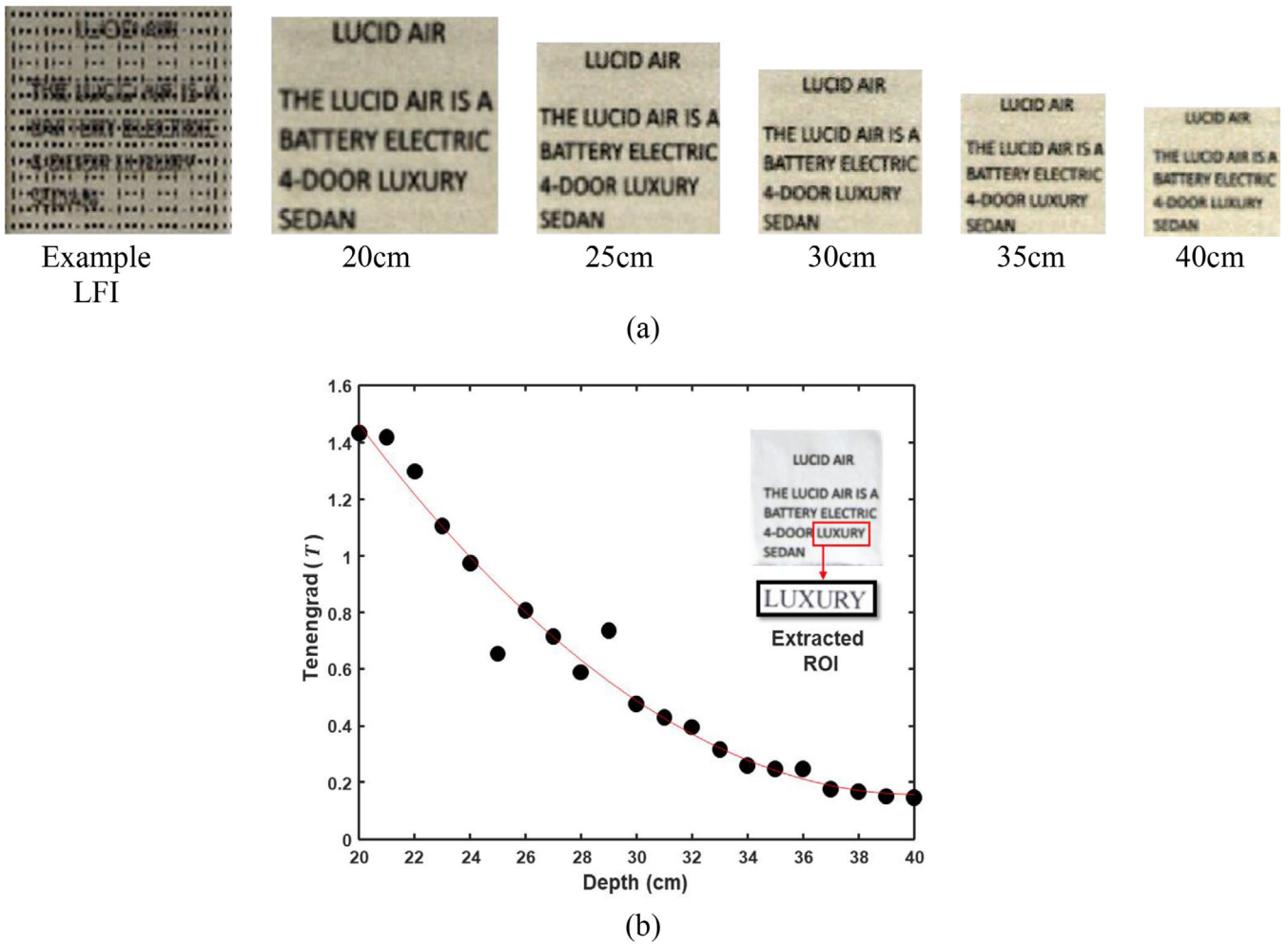


FIGURE 13 | Example of the LFI along with the central sub-aperture images with descriptive text extracted from corresponding depths LFIs (a) and (b) relationship between the refocused distances and the T values.

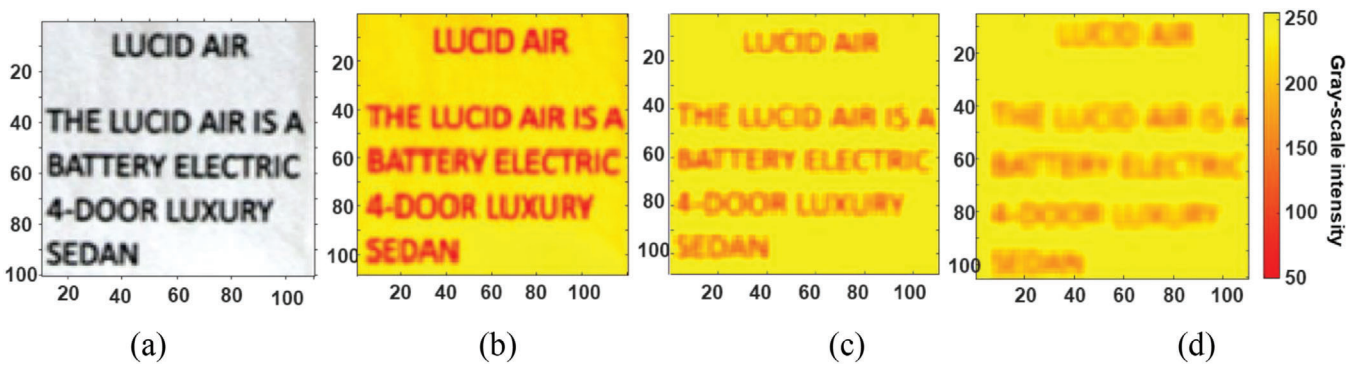


FIGURE 14 | Example of (a) reference image and (b-d) depth reconstructed images under the depths of 20 cm, 30 cm and 40 cm [color is used for visualization only].

to read the text. It is also observed that the depths of 20 cm to 22 cm show the same α value. Different sharpness can be seen for each depth level within the range of 20 cm to 22 cm, as shown in Figure 13(b) while using the same α value [Figure 11(a)]. This is because the difference in the distance of the refocused object from the camera and the reduction in the pixels of the reconstructed depth object led to each refocused object within the same range displaying a different sharpness value.

It is also crucial to ascertain that the objects placed within the same depth plane exhibit consistent focus quality (sharpness). An experiment was conducted where three objects were placed at 20 cm with a separation of 2.5 cm (horizontally). The LFC is used to capture images. Each object is refocused individually through the depth reconstruction process discussed in section 2.2.3 and the corresponding α values are recorded. Figure 15(a) shows an example of refocused images obtained at 20 cm. This process is

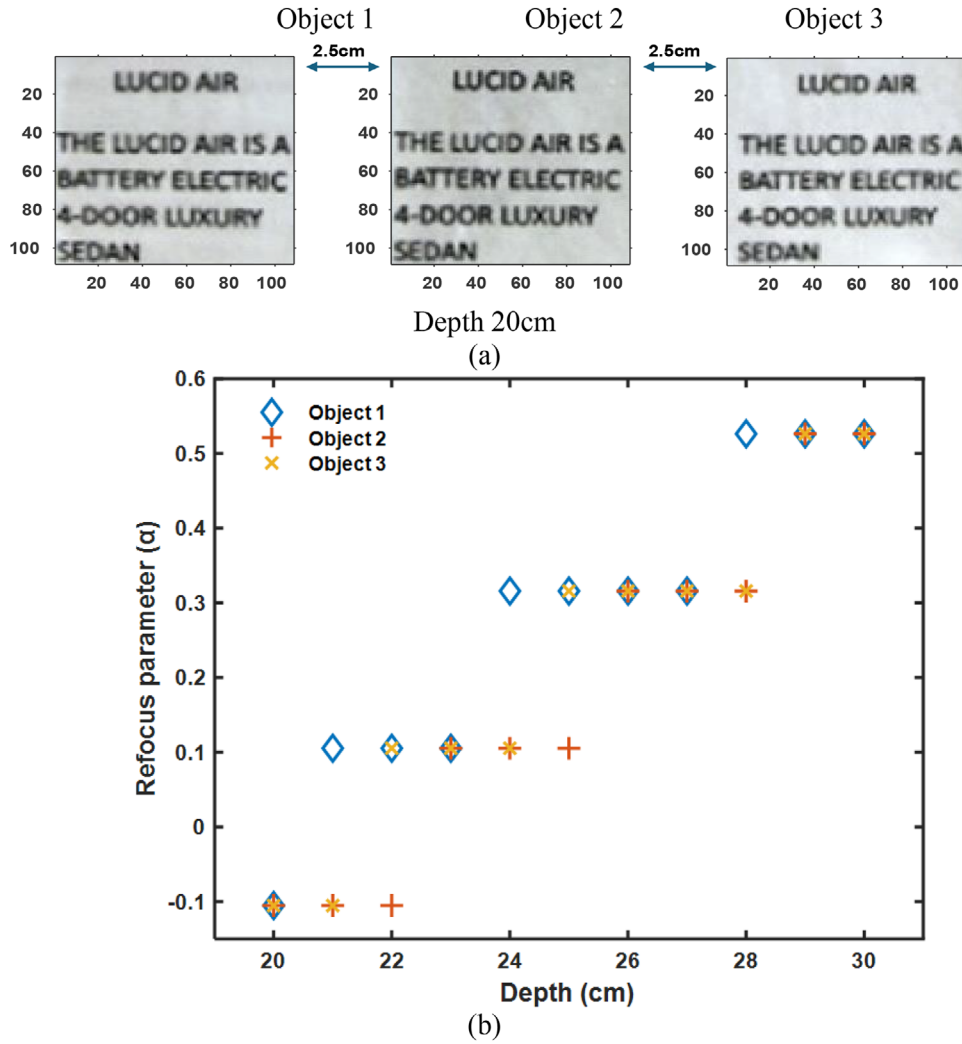


FIGURE 15 | Example of the refocused image at 20 cm depth, illustrating three objects sharing the same depth plane (a) and (b) relationship between the α and the refocused depths for three target objects.

repeated for other depths within the range of 21 cm to 40 cm. The relationship between refocus depth and α for each object is illustrated in Figure 15(b). It reveals that multiple objects placed at the same depth plane have different α values and do not maintain the same focus quality. This discrepancy arises due to variations in the distance of each object from the center of the image sensor. Although each object is equidistant from the camera under the same depth plane, slight angular differences in their positions, and changes in distances from the camera, result in different refocus quality.

3.4 | Pixel Size and FOV Calibration

FOV estimation is crucial for achieving a wider FOV in VR. It can provide the angle information of the camera to capture multiple views of the scene. Before calculating the FOV, it is necessary to calculate the pixel size, which can be further used for the FOV estimation. For the pixel size calculation, a checkerboard image was used to capture the image at a depth of 40 cm from the camera, as shown in Figure 16. The pixel size is estimated by:

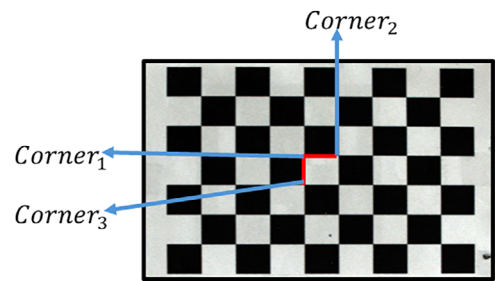


FIGURE 16 | Illustrates the measurement points for pixel size estimation (i.e., *Corners 1, 2, and 3*).

$$Ps_L_x = \frac{Ss_L_x}{Ss_p_x} \quad (24)$$

$$Ps_L_y = \frac{Ss_L_y}{Ss_p_y} \quad (25)$$

where Ps_L_x and Ps_L_y represent each pixel size in the x and y directions, Ss_L_x and Ss_L_y represents the actual square size of the checkerboard in horizontal and vertical directions. In this

TABLE 5 | Assessment of depth quality in terms of T under a different FOV.

f_m (mm)	FOV (°)	$T(a.u)$
9.40	76.26	0.056
18.80	36.67	0.061
28.21	24.39	0.087

study, which is 20 mm in both directions. Ss_{p_x} and Ss_{p_y} represents the size in pixels of each square of the captured checkerboard in horizontal and vertical directions, respectively, and they can be calculated by:

$$Ss_{p_x} = Corner_2 - Corner_1 \quad (26)$$

$$Ss_{p_y} = Corner_3 - Corner_1 \quad (27)$$

The FOV is estimated by:

$$FOV = 2\arctan^{-1} \left(\frac{d}{2z} \right) \quad (28)$$

where z represents the distance between the camera and the object, which is estimated through the extrinsic parameters of the camera calibration. d represents the diagonal length of the image, which can be calculated as:

$$d = \sqrt{W^2 + H^2} \quad (29)$$

where W and H represent the width and height of the image, which are calculated by:

$$W = Ps_{L_x} \times N_x \quad (30)$$

$$H = Ps_{L_y} \times N_y \quad (31)$$

where N_x and N_y represent the resolution of the central sub-aperture image in the x (width) and y -axis (height). In this study, $N_x \times N_y$ is equal to 625×434 pixels, which is the resolution of each sub-aperture image. The resolution of each sub-aperture image depends on the total number of microlenses [67]. By using Equations (24), (25), and (28), the estimated pixel size in mm is 0.41×0.41 , and the FOV is 36.67° . The angle of camera rotation should be equal to the FOV. Thus, the camera can be rotated to the left and right to acquire additional views of the scene. This provides the maximum FOV for wider VR content creation. Also, the LFC used in this study provides a variable focal length, f_M . Therefore, it is important to obtain the FOV and focus values under the different f_M . Table 5 presents depth quality in terms of T under the different FOVs. As shown in Table 5, the T value increases with the increase of f_M and decreases with the increase in FOVs. In this study, the focal length $f_M = 18.80$ mm is used, as it provides a better FOV compared to $f_M = 28.21$ mm. Although $f_M = 28.21$ mm provides a better focus level, it requires capturing more views of the scene. This increases the complexity of the VR translation process, limits the refocusing ability for objects located at the extreme left or right positions, and negatively impacts the VR user experience. Furthermore, capturing additional images makes the depth reconstruction and

translation processes computationally intensive, increasing the processing time. On the other hand, $f_M = 9.40$ allows fewer images to be captured to achieve a wider FOV in VR but compromises depth reconstruction quality. Considering these trade-offs, $f_M = 9.40$ was chosen for LF acquisition to ensure an optimal balance between the FOV, focus level, and computational efficiency.

The calculated pixel size is subsequently used to determine the size of the acquired checkerboard, which is then compared with the actual size of the checkerboard to validate the accuracy of the estimated pixel size. The Ps_{L_x} and Ps_{L_y} are used to calculate the width and height of the checkerboard. Initially, the number of pixels across the width and height of the checkerboard is estimated. Subsequently, these pixel counts across width and height are multiplied by the pixel size to calculate the checkerboard width and height. The results are presented in Table 6. They indicate that the pixel size is accurately estimated, as the calculated dimensions of the checkerboard closely match its actual dimensions.

4 | Reconstruction of a Realistic Scene

To reconstruct the real objects under different depths, experiments were conducted by setting up a scenario where multiple objects were placed at different depths. Figure 17 shows the schematic illustration of the experimental scenario and its physical setup. The proposed depth reconstruction algorithm is used to reconstruct the object under different depths. The Lytro Illum camera is used to capture the objects under three different views (FOVs) by rotating the camera from a straight direction to the left and right as shown in Figure 17(a). The rotation of the camera is determined through FOV calibration. The camera covers 36.67° of FOV. To capture the left (FOV 2) and right (FOV 3) views, the camera can be rotated 36.67° to the left and right from the central position. The central position was considered 0° . Figure 18 shows examples of raw LFIs captured under the three FOVs.

After decoding, the raw LFIs are corrected through the aberration correction algorithm proposed in Section 3.2. Figure 19 shows the corrected and uncorrected LFIs. The quality of the LFIs improved compared to the captured [Figure 18] and uncorrected [Figure 19(a)] LFIs. The corrected LFIs were then used to reconstruct the depth images for each object under different depths. The depth levels of 20 cm, 25 cm, and 30 cm are considered. The depth of 30 cm is identified as the maximum achievable depth resolution through the calibration. Examples of depth reconstructed images for each scene are shown in Figure 20.

The BIQA and FRIQA, as discussed in Section 2.3, are used to evaluate the quality of the reconstructed image. For BIQA, the depth reconstructed images of FOV 1, and for FRIQA, the depth reconstructed images of FOV 1 and FOV 3 are used. Three distinct objects were placed at a depth of 20 cm, 25cm, and 30 cm, respectively, in each FOV. Figure 22 shows the α values obtained for each refocused object under different depths for both FOV 1 and FOV 3.

Table 7 represents the overview of reconstruction quality obtained through BIQA. It indicates that the depth-reconstructed images

TABLE 6 | Comparison between actual and calculated sizes of checkerboard.

f_m (mm)	$Ps_{L_x} \times Ps_{L_y}$ (mm)	Actual (mm)		Calculated (mm)		Relative error (%)	
		Hight	Width	Hight	Width	Hight	Width
18.80	0.41×0.41	140	180	140.01	180	0.01	0

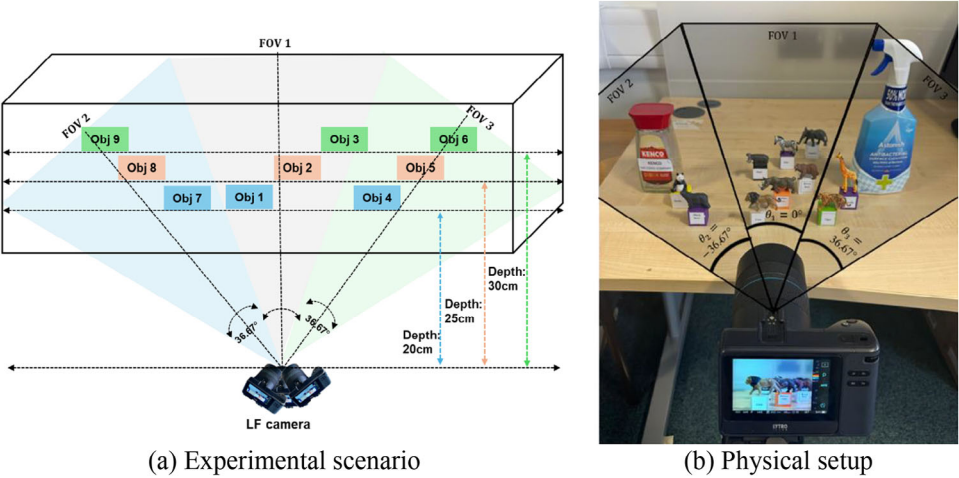


FIGURE 17 | Schematic illustration of the experimental scenario and its physical setup.

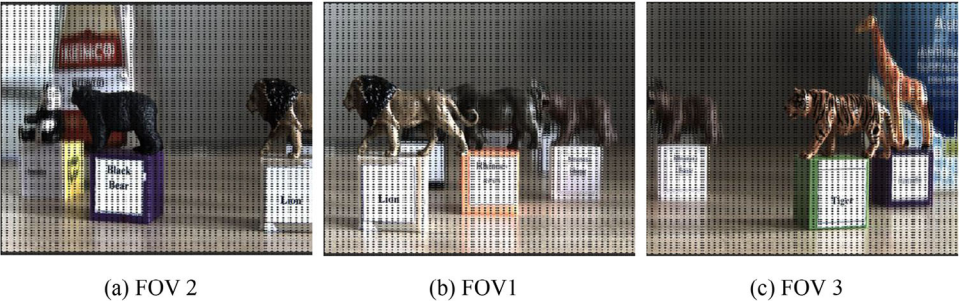


FIGURE 18 | Examples of original LFIs captured under different FOVs.

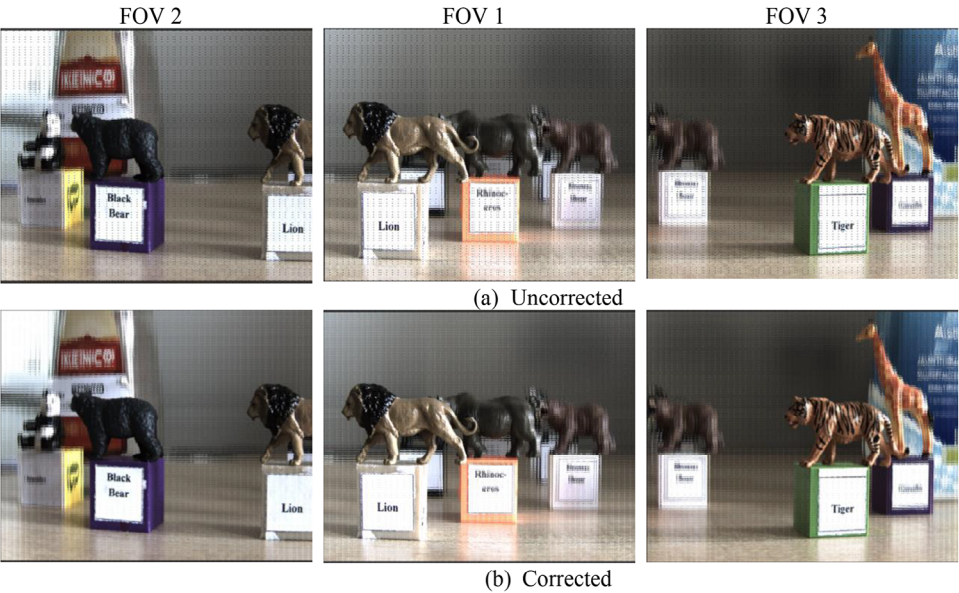


FIGURE 19 | Examples of corrected and uncorrected LF images under different FOVs.

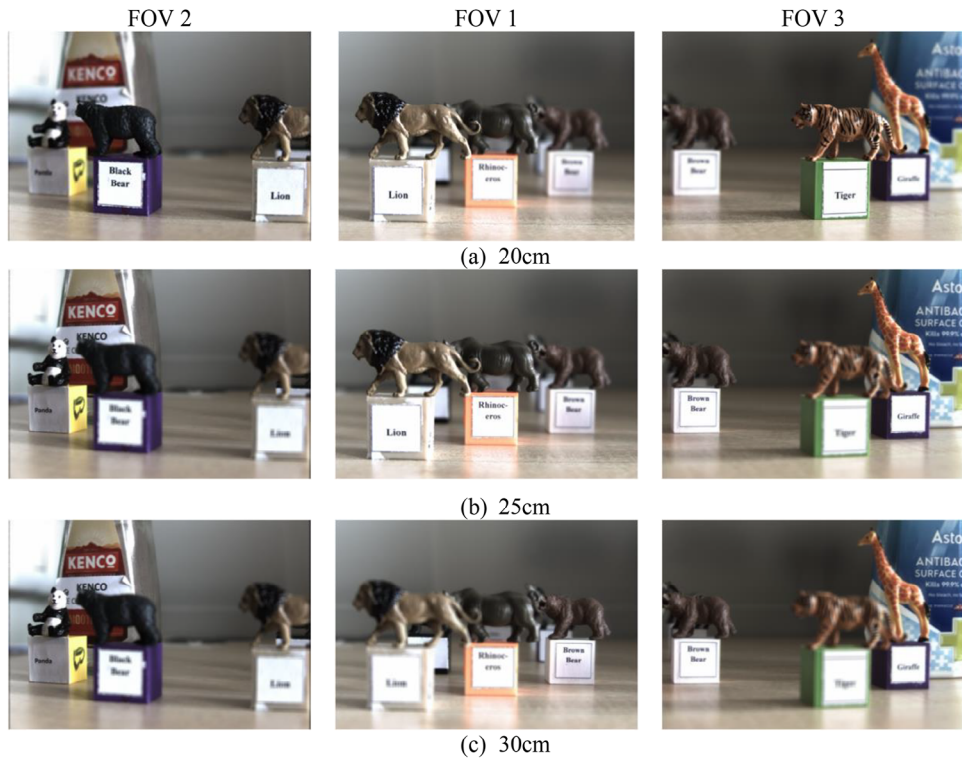


FIGURE 20 | Examples of reconstructed images under different FOVs and depth levels.

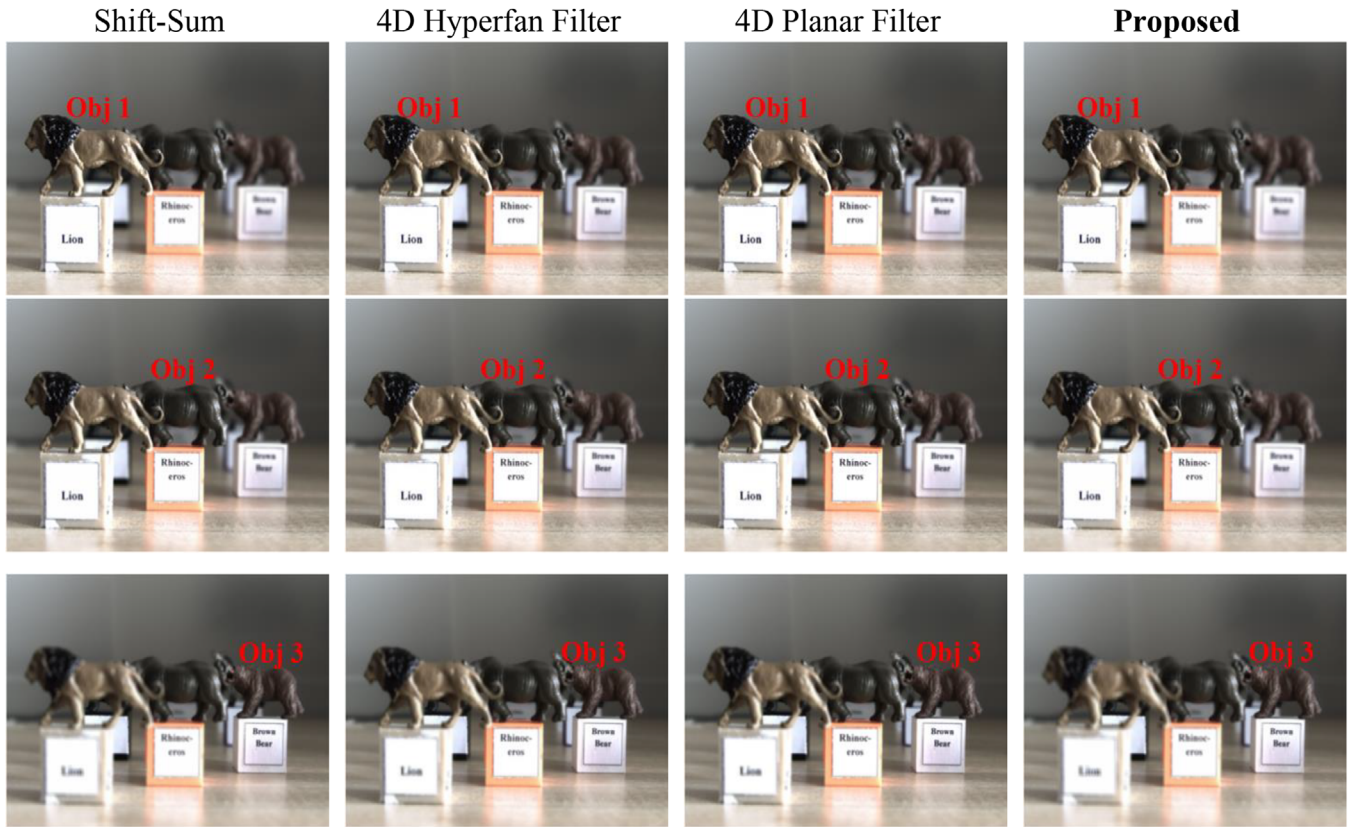
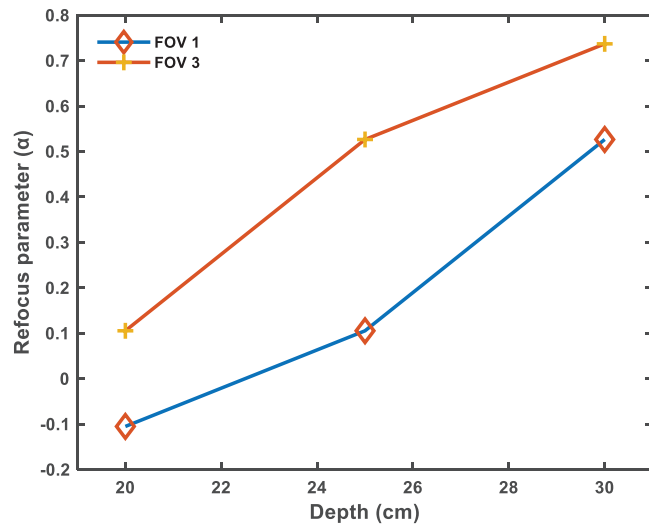


FIGURE 21 | Reconstructed images under different depths for FOV 1 achieved through Shift-sum, 4D hyperfan filter, 4D planar filter, and the proposed model.

TABLE 7 | Overview of reconstruction quality achieved using tenengrad T , gradient magnitude G , LOV δ , and intensity In at FOV 1.

LF images	Obj 1 (lion)	Obj 2 (rhino)	Obj 3 (brown bear)
	$T/G/\delta/In$ (%)	$T/G/\delta/In$ (%)	$T/G/\delta/In$ (%)
Original and corrected	0.16 / 0.84 / -0.64 / 0.80	1.09 / 2.06 / 2.08 / 0.26	-1.78 / -0.26 / -3.83 / 0.64
Original and uncorrected	-3.16 / -2.28 / -6.3 / -2	-2.45 / -1.97 / -4.97 / -1.82	-3.73 / -2.83 / -7.50 / -1.44
Corrected and uncorrected	3.44 / 3.19 / 6.12 / 2.70	2.63 / 4.11 / 7.42 / 2.13	2.02 / 2.63 / 3.97 / 2.12

**FIGURE 22** | Illustration of α values under different depths for FOV 1 and FOV 3.

derived from corrected LFIs show better reconstruction quality for objects placed at 20 cm (Obj 1) and 25 cm (Obj 2) compared to those from the original LFIs, except for the Obj 1 sharpness in terms of δ . For an object placed at 30 cm (Obj 3), the corrected LFIs show an improvement in intensity; however, a decline in sharpness is observed. Conversely, comparing the depth reconstructed images derived from the original LFIs with those from the uncorrected LFIs shows reduced reconstruction quality due to the low intensity of sub-aperture images (caused by vignetting) in the uncorrected LFIs. A better reconstruction quality is achieved when comparing the depth-reconstructed images from corrected LFIs with those from uncorrected LFIs. This indicates that the vignetting effect reduced the reconstruction quality.

The proposed depth reconstruction model is compared with other frequency-domain-based depth reconstruction methods, such as 4D planar filtering, along with a 4D frequency hyperfan filter [76, 86]. Figure 21 illustrates the reconstructed images under different depths for FOV 1 achieved through frequency domain-based depth reconstruction methods, and the proposed model Table 8 represents the SSIM and PSNR for depth reconstructed images achieved by the proposed model and the frequency-based depth reconstruction methods. It shows that the proposed model provides better depth reconstruction quality than the frequency-based methods. Therefore, it demonstrates that the proposed model reconstructs the depth images in better quality in terms of sharpness, pixel intensity, and focus quality.

As stated in the introduction, camera array-based [25, 30] and point cloud-based [40, 41] approaches are used for depth

reconstruction in various applications. These approaches differ fundamentally from the proposed depth reconstruction model in terms of data requirements, system complexity, and reconstruction pipelines. Due to these inherent differences, a direct quantitative comparison is not feasible. The reconstruction performed using camera array-based systems creates ghosting artifacts, which commonly manifest in occluded regions due to their limited angular resolution. For instance, Figure 23 presents a qualitative analysis through visual inspection. It is evident that scenes reconstructed using camera array-based systems exhibit ghosting artifacts [see Figure 23(a)], likely attributed to the limited angular resolution. Furthermore, in Figure 23(b), which depicts an object reconstructed using point cloud reconstruction, the object lacks realism and is unsuitable for VR content creation. Figure 23(c) illustrates the depth images reconstructed using the proposed depth reconstruction model. Notably, this model is devoid of ghosting artifacts, resulting in a realistic depiction of the reconstructed scene.

Table 9 also provides a comprehensive analysis in terms of the advantages and limitations of the camera array-based and point cloud reconstruction methods in comparison to the proposed depth reconstruction model. It demonstrates that the proposed model incorporates more angular information compared to camera array-based reconstruction and generates a more realistic scene than point cloud-based reconstructions. The proposed model employs a single-sensor LF system that captures the utmost angular information, which is subsequently utilized for precise depth reconstruction.

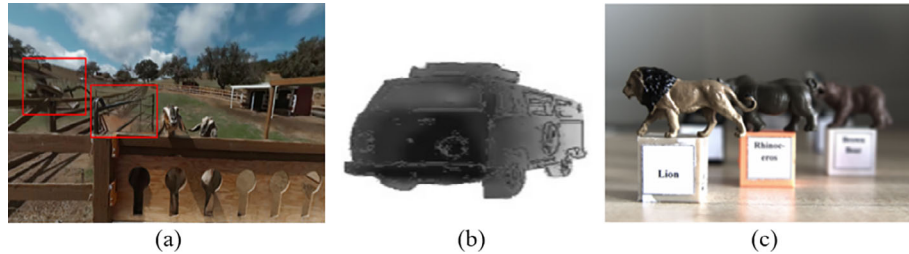
Table 10 presents a comparison of the LF image size and processing time between the original and corrected LFIs. The processing time of each LF image is measured from loading the LF image into MATLAB to depth reconstruction. All processing was conducted using MATLAB R2021a and an Intel(R) Core (TM) i5-4460 CPU operating at 3.20 GHz. It is evident that the proposed model reduces the processing time by 31.79%, and the LFI size is reduced by 43.42% after correction.

5 | Translation to VR

Once the objects are reconstructed under different depths and FOVs, the final step is to translate them into VR content. For translation purposes, the Unity game engine has been employed. In Unity, a scene has been meticulously crafted by integrating all three FOVs to create a broader VR scene. To facilitate user interaction within this scene, the Oculus Integration package has been integrated. Instead of a joystick, a hand-tracking mechanism is employed to track users' hands for interaction within the scene.

TABLE 8 | FRIQA-based reconstruction quality assessment for FOVs 1 and 3.

Algorithm	Obj 1 PSNR/SSIM		Obj 2 PSNR/SSIM		Obj 3 PSNR/SSIM	
	FOV 1	FOV 3	FOV 1	FOV 3	FOV 1	FOV 3
4D Planar filter [86]	44.22/0.997	36.29/0.991	40.32/0.986	36.00/0.997	45.53/0.996	36.75/0.993
4D Hyperfan filter [76]	44.82/0.998	38.54/0.994	43.78/0.993	38.18/0.998	47.96/0.998	39.87/0.996
Proposed model	48.47/0.999	48.78/0.999	50.17/0.999	40.79/0.999	48.15/0.999	42.60/0.998

**FIGURE 23** | Comparison between (a) reconstructed using a camera array-based method; (b) reconstructed object through point cloud reconstruction; and (c) depth reconstructed image through the proposed model.**TABLE 9** | Qualitative analysis of the camera array-based and point cloud depth reconstruction techniques along with the proposed depth reconstruction model.

Method	Advantages	Limitations
Camera array-based systems [25, 30]	<ul style="list-style-type: none"> • Captures limited angular information. • Allows for viewpoint changes. 	<ul style="list-style-type: none"> • Requires multiple cameras, making it expensive. • Generates a large amount of raw data. • Needs complex calibration for geometric alignment. • High computational demand. • Limited angular information.
Point cloud-based reconstruction [40, 41]	<ul style="list-style-type: none"> • Can provide 3D structural information of the scene. • Cost-effective. • Low computational cost. • Supports some viewpoint changes. 	<ul style="list-style-type: none"> • LF camera-based point cloud reconstruction struggles with realistic scene reconstruction. • Generates discrete 3D points leading to artefacts. • Lacks refocusing capability, affecting depth perception.
Proposed method (single-sensor LF camera-based VR content creation)	<ul style="list-style-type: none"> • A low-cost system using a single LF camera. • Reduces LF image size. • Low computational cost. • Provides more angular information. • Easy to set up and portable. 	<ul style="list-style-type: none"> • Lower spatial resolution.

TABLE 10 | Overview of the comparison of LF image size and processing time of the depth reconstruction.

	Size (MB)	Difference (%)	Processing time (seconds)	Difference (%)
Original LFI (13 × 13) sub-aperture images	397.22	43.42	216.98	31.79
Corrected LFI (13 × 13) sub-aperture images	224.72		147.99	

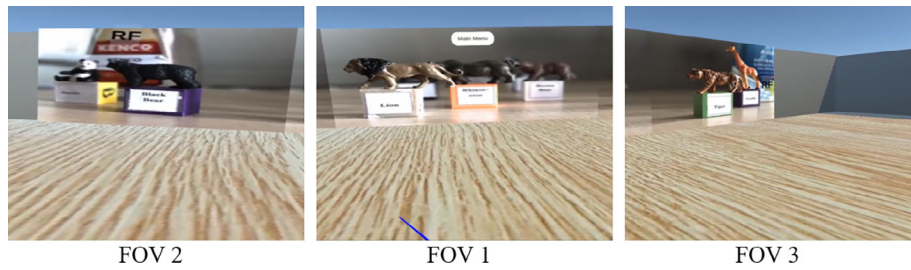


FIGURE 24 | Illustration of the translated VR scene.

Interaction is enabled by pinching the user's fingers rather than pressing joystick buttons. This capability allows users to maintain interaction within the scene, providing a more lifelike experience. In the created VR scene, users can refocus on each object by clicking the label of each object within the scene, and only the refocused object will appear as focused, while others will appear blurred. The Meta Quest2 headset has been used to experience the translated VR content.

Figure 24 illustrates the translated, reconstructed LFIs presented at various depths within VR content. Users can click on the labels of each object to refocus on that specific object. This phenomenon aligns with human vision systems, which allow users to perceive only one object in complete focus at a time. As part of ongoing research, user experience will be evaluated through user trials and feedback surveys. During the user trial, participants will be exposed to two types of VR content: computer-generated and realistic VR content generated by the proposed model. After each VR experience, user feedback will be collected using two types of questionnaires provided in [75] and [76].

6 | Conclusion

In this paper, a depth reconstruction model based on LF imaging has been proposed to generate data for VR content creation. The primary objective of this model is to enable refocusing, a feature that is currently absent in realistic virtual reality content. The proposed model initially corrects the distortion and vignetting effects present in the captured LFIs. Subsequently, sub-aperture images are extracted from the corrected LFIs. Following this, depth images are reconstructed using these extracted sub-aperture images. To accurately reconstruct depth images, depth calibrations were performed.

The key findings and contributions of this work are summarized below:

- It has been revealed that distortion impacts the reconstruction quality. Without distortion correction, objects within the reconstructed depth images may exhibit deviations in shape and size, resulting in reduced reconstruction quality.
- By addressing vignetting effects that cause certain angular images to have low intensity or appear black, the inclusion of poor-quality sub-aperture images is prevented in the reconstruction process. The results demonstrate that including sub-aperture images with low-quality or completely black pixels in the reconstruction model results in reduced reconstruction quality.

- The proposed model reduces the data size without compromising reconstruction quality. This is achieved by removing unwanted sub-aperture images (which have low intensity or completely dark pixels) through vignetting effect correction, thereby reducing data processing time and storage demands.
- The quality of reconstructed images is enhanced compared to the original LFIs, ensuring that the final output remains free from artifacts caused by vignetting and distortion. This improvement in reconstruction quality is achieved by addressing most of the challenges associated with LF imaging.
- The proposed model's reconstruction quality is validated by two quality assessments. The results demonstrate that the proposed depth reconstruction model shows superior performance.

However, the proposed model is constrained by the spatial resolution of the lytro illum camera, resulting in a reduction in the output image quality. Furthermore, the limited baseline of angular images restricts the accurate reconstruction of objects within a restricted range. Additionally, the parameters defined in this study for the reconstruction model are specific to the lytro illum camera. This model can be adapted for other LFCs by modifying the model parameters.

For future enhancements, modifications to the acquisition system could improve spatial resolution and provide adequate angular information for improved reconstruction. Furthermore, consideration of the baseline (the distance between sub-aperture images or different viewpoints) during camera selection for LF acquisition in VR is crucial, as it directly influences the reconstruction quality of distant objects. Regarding the translation of this model to VR content, an eye-tracking mechanism can be integrated into VR content to allow users to adjust their focus based on eye tracking, thereby enhancing the realism of the experience.

Furthermore, it is envisaged that applying the realistic scene to an experimental scenario, such as a virtual animal museum, and conducting a user study to evaluate immersion and the level of motion sickness in comparison with reference VR content will be considered.

Author Contributions

Ali Khan: writing – original draft, investigation, collection of materials, references. **Md. Moinul Hossain:** technical discussion, re-writing, review, editing, supervision. **Konstantinos Sirlantzis:** technical

discussion, funding acquisition. **Alexandra Covaci**: technical discussion, analysis, review. **Qi Qi**: analysis, review.

Acknowledgements

The authors acknowledge receipt of the following financial support for the research, authorship, and publication of this article: Interreg 2 Seas programme 2014–2020 co-funded by the European Regional Development Fund under subsidy contract No. 2S05-038 (MOTION project).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. R. Schroeder, "Learning From Virtual Reality Applications in Education," *Virtual Real* 1 (1995): 33–39, <https://doi.org/10.1007/BF02009711>.
2. D. Kamińska, T. Sapiński, S. Wiak, et al., "Virtual Reality and Its Applications in Education: Survey," *Information* 10, no. 10 (2019): 318, <https://doi.org/10.3390/info10100318>.
3. L. P. Berg and J. M. Vance, "Industry Use of Virtual Reality in Product Design and Manufacturing: A Survey," *Virtual Real* 21 (2017): 1–17, <https://doi.org/10.1007/s10055-016-0293-9>.
4. S. Choi, K. Jung, and S. D. Noh, "Virtual Reality Applications in Manufacturing Industries: Past Research, Present Findings, and Future Directions," *Concurrent Engineering* 23, no. 1 (2015): 40–63, <https://doi.org/10.1177/1063293214568814>.
5. E. van Wyk and R. de Villiers, "Virtual Reality Training Applications for the Mining Industry," in *Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa* (2009), 53–63, <https://doi.org/10.1145/1503454.1503465>.
6. D. A. Guttentag, "Virtual Reality: Applications and Implications for Tourism," *Tour Management* 31, no. 5 (2010): 637–651, <https://doi.org/10.1016/j.tourman.2009.07.003>.
7. B. van Twillert, M. Bremer, and A. W. Faber, "Computer-Generated Virtual Reality to Control Pain and Anxiety in Pediatric and Adult Burn Patients during Wound Dressing Changes," *Journal of Burn Care and Research* 28, no. 5 (2007): 694–702, <https://doi.org/10.1097/BCR.0B013E318148C96F>.
8. M. Javaid and A. Haleem, "Virtual Reality Applications Toward Medical Field," *Clinical Epidemiology and Global Health* 8, no. 2 (2020): 600–605, <https://doi.org/10.1016/j.cegh.2019.12.010>.
9. D. T. Nicholson, C. Chalk, W. R. J. Funnell, and S. J. Daniel, "Can Virtual Reality Improve Anatomy Education? A Randomised Controlled Study of a Computer-Generated Three-Dimensional Anatomical Ear Model," *Medical Education* 40, no. 11 (2006): 1081–1087, <https://doi.org/10.1111/j.1365-2929.2006.02611.x>.
10. M. Lin, L. San, and Y. Ding, "Construction of Robotic Virtual Laboratory System Based on Unity3D," *IOP Conference Series: Materials Science and Engineering* 768, no. 7 (2020): 072084, <https://doi.org/10.1088/1757-899X/768/7/072084>.
11. G. Gabajová, M. Krajčovič, M. Matys, B. Furmannová, and N. Burganova, "Designing Virtual Workplace Using Unity 3D Game Engine," *Acta Technologia* 7, no. 1 (2021): 35–39, [10.22306/atec.v7i1.101](https://doi.org/10.22306/atec.v7i1.101).
12. S. Wang, Z. Mao, C. Zeng, H. Gong, S. Li, and B. Chen, "A New Method of Virtual Reality Based on Unity3D," in *2010 18th International Conference on Geoinformatics* (2010), 1–5, <https://doi.org/10.1109/GEOINFORMATICS.2010.5567608>.
13. E. Abdul Rahim, A. Duenser, M. Billingham, et al., "A Desktop Virtual Reality Application for Chemical and Process Engineering Education," in *Proceedings of the 24th Australian Computer-Human Interaction Conference* (2012), 1–8, <https://doi.org/10.1145/2414536.2414537>.
14. H.-H. Tsai, X.-Y. Hou, C.-T. Chang, C.-Y. Tsai, P.-T. Yu, and J.-S. Roan, "Interactive Contents With 360-Degree Panorama Virtual Reality for Soil and Water Conservation Outdoor Classroom," in *2020 International Symposium on Educational Technology (ISET)* (2020), 78–82, <https://doi.org/10.1109/ISET49818.2020.00026>.
15. P. Kuna, A. Hašková, and L. Borza, "SWOT Analysis of Virtual Reality Creation Equipments," *RE-SOURCE* (2022), <https://doi.org/10.53349/resource.2022.iS24.a1106>.
16. A. Khan, M. M. Hossain, A. Covaci, K. Sirlantzis, and C. Xu, "Light Field Imaging Technology for Virtual Reality Content Creation: A Review," *IET Image Processing* 18, no. 11 (2024): 2817–2837, <https://doi.org/10.1049/ipr2.13144>.
17. R. S. Overbeck, D. Erickson, D. Evangelakos, M. Pharr, and P. Debevec, "A System for Acquiring, Processing, and Rendering Panoramic Light Field Stills for Virtual Reality," *ACM Transactions on Graphics* 37, no. 6 (2018): 1–15, <https://doi.org/10.1145/3272127.3275031>.
18. M. Broxton, J. Flynn, R. Overbeck, et al., "Immersive Light Field Video With a Layered Mesh Representation," *ACM Transactions on Graphics* 39, no. 4 (2020): 1–86, <https://doi.org/10.1145/3386569.3392485>.
19. M. van Gisbergen, M. Kovacs, F. Campos, M. van der Heeft, and V. Vugts, "What We Don't Know. The Effect of Realism in Virtual Reality on Experience and Behaviour," in *Augmented Reality and Virtual Reality: The Power of AR and VR for Business*, Edited by M. C. tom Dieck and T. Jung (2019): 45–57, https://doi.org/10.1007/978-3-030-06246-0_4.
20. M. Balcerak Jackson and B. Balcerak Jackson, "Immersive Experience and Virtual Reality," *Philosophy & Technology* 37, no. 19 (2024): 19, <https://doi.org/10.1007/s13347-024-00707-1>.
21. R. Lavoie, K. Main, C. King, and D. King, "Virtual Experience, Real Consequences: The Potential Negative Emotional Consequences of Virtual Reality Gameplay," *Virtual Reality* 25, (2021): 69–81, <https://doi.org/10.1007/s10055-020-00440-y>.
22. F.-C. Huang, D. P. Luebke, and G. Wetzstein, "The Light Field Stereoscope," in *SIGGRAPH'15: ACM SIGGRAPH 2015 Emerging Technologies* (2015): 24.
23. I. Ihrke, J. Restrepo, and L. Mignard-Debise, "Principles of Light Field Imaging: Briefly Revisiting 25 Years of Research," *IEEE Signal Processing Magazine* 33, no. 5 (2016): 59–69, <https://doi.org/10.1109/MSP.2016.2582220>.
24. M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light Field Microscopy," in *Acm Siggraph 2006 Papers* (2006): 924–934.
25. R. S. Overbeck, D. Erickson, D. Evangelakos, and P. Debevec, "The Making of Welcome to Light Fields VR," in *ACM SIGGRAPH 2018 Talks* (2018): 1–2, <https://doi.org/10.1145/3214745.3214811>.
26. M. Broxton, L. Grosenick, S. Yang, et al., "Wave Optics Theory and 3-D Deconvolution for the Light Field Microscope," *Optics Express* 21, no. 21 (2013): 25418–25439, <https://doi.org/10.1364/OE.21.025418>.
27. H. Sheng, S. Deng, S. Zhang, C. Li, and Z. Xiong, "Segmentation of Light Field Image With the Structure Tensor," in *2016 IEEE International Conference on Image Processing (ICIP)* (2016): 1469–1473, <https://doi.org/10.1109/ICIP.2016.7532602>.
28. Q. Qi, M. M. Hossain, J.-J. Li, B. Zhang, J. Li, and C.-L. Xu, "Approach to Reduce Light Field Sampling Redundancy for Flame Temperature Reconstruction," *Optics Express* 29, no. 9 (2021): 13094–13114, <https://doi.org/10.1364/OE.424112>.
29. J. Sun, J. Sun, C. Xu, et al., "Three-Dimensional Temperature Field Measurement of Flame Using a Single Light Field Camera," *Optics Express* 24, no. 2 (2016): 1118–1132, <https://doi.org/10.1364/OE.24.001118>.
30. J. Yu, "A Light-Field Journey to Virtual Reality," *IEEE Multimed* 24, no. 2 (2017): 104–112, <https://doi.org/10.1109/MMUL.2017.24>.

31. C. Gao, Y. Peng, R. Wang, Z. Zhang, H. Li, and X. Liu, "Foveated Light-Field Display and Real-Time Rendering for Virtual Reality," *Applied Optics* 60, no. 28 (2021): 8634–8643.
32. C. Zhang and T. Chen, "A Self-Reconfigurable Camera Array," in *ACM SIGGRAPH 2004 Sketches on—SIGGRAPH '04* (2004), 151, <https://doi.org/10.1145/1186223.1186412>.
33. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography With a Hand-Held Plenoptic Camera," Stanford University, Computer Science Tech. 2005, Research Report CSTR 2005-02, Apr. 2005. Accessed: Jul. 26, 2023. [Online]. Available: <https://hal.science/hal-02551481>.
34. B. Wilburn, N. Joshi, V. Vaish, et al., "High Performance Imaging Using Large Camera Arrays," in *ACM SIGGRAPH 2005 Papers* (2005): 765–776, <https://doi.org/10.1145/1186822.1073259>.
35. K. Venkataraman, D. Lelescu, J. Duparré, et al., "PiCam: An Ultra-Thin High Performance Monolithic Camera Array," *ACM Transactions on Graphics* 32, no. 6 (2013): 1–13, <https://doi.org/10.1145/2508363.2508390>.
36. T. Milliron, C. Szczupak, and O. Green, "Hallelujah: The World's First Lytro VR Experience," in *ACM SIGGRAPH 2017 VR Village* (2017): 1–2, <https://doi.org/10.1145/3089269.3089283>.
37. M. Broxton, J. Busch, J. Dourgarian, et al., "A Low Cost Multi-Camera Array for Panoramic Light Field Video Capture," in *SIGGRAPH Asia 2019 Posters* (2019): 1–2, <https://doi.org/10.1145/3355056.3364593>.
38. J. Flynn, M. Broxton, P. Debevec, et al., "DeepView: View Synthesis With Learned Gradient Descent," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019): 2362–2371, <https://doi.org/10.1109/CVPR.2019.00247>.
39. P. Debevec, G. Downing, M. Bolas, H.-Y. Peng, and J. Urbach, "Spherical Light Field Environment Capture for Virtual Reality Using a Motorized Pan/Tilt Head and Offset Camera," in *ACM SIGGRAPH 2015 Posters* (2015): 1–1, <https://doi.org/10.1145/2787626.2787648>.
40. H. Farhood, S. Perry, E. Cheng, and J. Kim, "3D point Cloud Reconstruction From a Single 4D Light Field Image," in *Optics, Photonics and Digital Technologies for Imaging Applications VI* 11353, <https://doi.org/10.1117/12.2555292>.
41. F. Murgia, C. Perra, and D. Giusto, "3D Point Cloud Reconstruction From Single Plenoptic Image," *Telfor Journal* 8, no. 1 (2016): 26–31, <https://doi.org/10.5937/telfor1601026M>.
42. H. Farhood, S. Perry, E. Cheng, and J. Kim, "Enhanced 3D Point Cloud From a Light Field Image," *Remote Sensing* 12, no. 7 (2020): 1125, <https://doi.org/10.3390/rs12071125>.
43. S. Rogge, I. Schiopu, and A. Munteanu, "Depth Estimation for Light-Field Images Using Stereo Matching and Convolutional Neural Networks," *Sensors* 20, no. 21 (2020): 6188, <https://doi.org/10.3390/s20216188>.
44. S. B. Dayan, D. Mendlovic, and R. Giryes, "Deep Sparse Light Field Refocusing," *CoRR*, abs/2009.02582, (2020), <https://doi.org/10.48550/arXiv.2009.02582>.
45. Z. Huang, J. A. Fessler, T. B. Norris, and I. Y. Chun, "Light-Field Reconstruction and Depth Estimation From Focal Stack Images Using Convolutional Neural Networks," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2020): 8648–8652, <https://doi.org/10.1109/ICASSP40776.2020.9053586>.
46. H. W. F. Yeung, J. Hou, J. Chen, Y. Y. Chung, and X. Chen, "Fast Light Field Reconstruction With Deep Coarse-to-Fine Modeling of Spatial-Angular Clues," in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018): 137–152.
47. M. Gupta, A. Jauhari, K. Kulkarni, S. Jayasuriya, A. Molnar, and P. Turaga, "Compressive Light Field Reconstructions Using Deep Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017): 11–20.
48. N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-Based View Synthesis for Light Field Cameras," *ACM Transactions on Graphics* 35, no. 6 (2016): 1–10, <https://doi.org/10.1145/2980179.2980251>.
49. G. Wu, Y. Liu, L. Fang, Q. Dai, and T. Chai, "Light Field Reconstruction Using Convolutional Network on EPI and Extended Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, no. 7 (2019): 1681–1694, <https://doi.org/10.1109/TPAMI.2018.2845393>.
50. K. Y. Tang, G. Yu, J. Wang, Y. He, S.-Z. Xu, and S.-H. Zhang, "Strategies for Reducing Motion Sickness in Virtual Reality Through Improved Handheld Controller Movements," *Graph Models* 138 (2025): 101254.
51. K. Carnegie and T. Rhee, "Reducing Visual Discomfort With HMDs Using Dynamic Depth of Field," *IEEE Computer Graphics and Applications* 35, no. 5 (2015): 34–41, <https://doi.org/10.1109/MCG.2015.98>.
52. A. Rasla and M. Beyeler, "The Relative Importance of Depth Cues and Semantic Edges for Indoor Mobility Using Simulated Prosthetic Vision in Immersive Virtual Reality," in *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology* (2022): 1–11, <https://doi.org/10.1145/3562939.3565620>.
53. R. Hussain, M. Chessa, and F. Solari, "Mitigating Cybersickness in Virtual Reality Systems Through Foveated Depth-of-Field Blur," *Sensors* 21, no. 12 (2021): 4006, <https://doi.org/10.3390/s21124006>.
54. M. Levoy and P. Hanrahan, "Light Field Rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* (1996): 31–42, <https://doi.org/10.1145/237170.237199>.
55. S. Pertuz, E. Pulido-Herrera, and J.-K. Kamarainen, "Focus Model for Metric Depth Estimation in Standard Plenoptic Cameras," *ISPRS Journal of Photogrammetry and Remote Sensing* 144 (2018): 38–47, <https://doi.org/10.1016/j.isprsjprs.2018.06.020>.
56. A. Khan, M. M. Hossain, K. Sirlantzis, A. Covaci, and W. S. Chowdhury, "Depth Estimation and Validation of Plenoptic Light Field Camera," in *2023 IEEE International Conference on Imaging Systems and Techniques (IST)* (2023): 1–6, <https://doi.org/10.1109/IST59124.2023.10355697>.
57. A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically Reparameterized Light Fields," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH '00* (2000): 297–306, <https://doi.org/10.1145/344779.344929>.
58. X. Gu, S. J. Gortler, and M. F. Cohen, "Polyhedral Geometry and the Two-Plane Parameterization," *Rendering Techniques '97* (1997): 1–12, https://doi.org/10.1007/978-3-7091-6858-5_1.
59. Y. Ji and J. Wu, "Calibration Method of Light-Field Camera for Photogrammetry Application—ScienceDirect," *Measurement* 148 (2019): 106943.
60. C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Refocusing Distance of a Standard Plenoptic Camera," *Optics Express* 24, no. 19 (2016): 21521, <https://doi.org/10.1364/OE.24.021521>.
61. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition* (2013): 1027–1034, <https://doi.org/10.1109/CVPR.2013.137>.
62. M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic Aperture Confocal Imaging," *ACM Transactions on Graphics* 23, no. 3 (2004): 825–834, <https://doi.org/10.1145/1015706.1015806>.
63. W. Fu, X. Tong, C. Shan, S. Zhu, and B. Chen, "Implementing Light Field Image Refocusing Algorithm," in *2015 2nd International Conference on Opto-Electronics and Applied Optics (IEM OPTRONIX)* (2015): 1–8, <https://doi.org/10.1109/OPTRONIX.2015.7345517>.
64. P. Zhou, W. Cai, Y. Yu, Y. Zhang, and G. Zhou, "A Two-Step Calibration Method of Lenslet-Based Light Field Cameras," *Optics and Lasers in Engineering* 115 (2019): 190–196, <https://doi.org/10.1016/j.optlaseng.2018.11.024>.

65. J. P. de Villiers, F. W. Leuschner, and R. Geldenhuys, "Centi-Pixel Accurate Real-Time Inverse Distortion Correction," *Optomechatronic Technologies* 2008. (2008): 320–327, <https://doi.org/10.1117/12.804771>.
66. K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, no. 12 (2011): 2341–2353, <https://doi.org/10.1109/TPAMI.2010.168>.
67. W. Darwish, Q. Bolsee, and A. Munteanu, "Plenoptic Camera Calibration Based on Microlens Distortion Modelling," *ISPRS Journal of Photogrammetry and Remote Sensing* 158 (2019): 146–154, <https://doi.org/10.1016/j.isprsjprs.2019.10.008>.
68. S. Gao, M. Han, and X. Cheng, "The Fast Iris Image Clarity Evaluation Based on Tenengrad and ROI Selection," in *Ninth International Conference on Graphic and Image Processing (ICGIP 2017)* (2018): 1391–1396, <https://doi.org/10.1117/12.2302509>.
69. Y. Liu, K. Gu, G. Zhai, X. Liu, D. Zhao, and W. Gao, "Quality Assessment for Real Out-of-Focus Blurred Images," *Journal of Visual Communication and Image Representation* 46 (2017): 70–80, <https://doi.org/10.1016/j.jvcir.2017.03.007>.
70. Q. Wu, Z. Wang, and H. Li, "A Highly Efficient Method for Blind Image Quality Assessment," in *2015 IEEE International Conference on Image Processing (ICIP)*, (2015): 339–343, <https://doi.org/10.1109/ICIP.2015.7350816>.
71. R. Bansal, G. Raj, and T. Choudhury, "Blur Image Detection Using Laplacian Operator and Open-CV," in *2016 International Conference System Modeling and Advancement in Research Trends (SMART)* (2016), 63–67, <https://doi.org/10.1109/SYSMAST.2016.7894491>.
72. W. Kong, J. Chen, Y. Song, Z. Fang, X. Yang, and H. Zhang, "Sobel Edge Detection Algorithm With Adaptive Threshold Based on Improved Genetic Algorithm for Image Processing," *International Journal of Advanced Computer Science and Applications* 14, no. 2 (2023), <https://doi.org/10.14569/IJACSA.2023.0140266>.
73. U. Sara, M. Akter, and M. S. Uddin, "Image Quality Assessment Through FSIM, SSIM, MSE and PSNR—A Comparative Study," *Journal of Computer Communication* 7, no. 3 (2019): 8–18, <https://doi.org/10.4236/jcc.2019.73002>.
74. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing* 13, no. 4 (2004): 600–612, <https://doi.org/10.1109/TIP.2003.819861>.
75. L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A Comprehensive Evaluation of Full Reference Image Quality Assessment Algorithms," in *2012 19th IEEE International Conference on Image Processing* (2012): 1477–1480, <https://doi.org/10.1109/ICIP.2012.6467150>.
76. D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear Volumetric Focus for Light Field Cameras," *ACM Transactions on Graphics* 34, no. 2 (2015): 1–20, <https://doi.org/10.1145/2665074>.
77. Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, no. 11 (2000): 1330–1334, <https://doi.org/10.1109/34.888718>.
78. Y. Liu, F. Mo, M. Aleksandrov, S. Zlatanova, and P. Tao, "Accurate Calibration of Standard Plenoptic Cameras Using Corner Features From Raw Images," *Optics Express* 29, no. 1 (2020): 158.
79. D. Brown, "Decentering Distortion of Lenses," *Photogrammetric Engineering* 32, no. 3 (1996): 444–462.
80. T. Georgiev and A. Lumsdaine, "Resolution in Plenoptic Cameras," in *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics and Photonics Technical Digest* (2009) (2009), <https://doi.org/10.1364/COSI.2009.CTuB3>.
81. Y. Liu, M. Zhu, T. Wang, et al., "Spatial Resolution of Light Field Sectioning Pyrometry for Flame Temperature Measurement," *Optics and Lasers in Engineering* 140 (2021): 106545, <https://doi.org/10.1016/j.optlaseng.2021.106545>.
82. J. Sun, M. M. Hossain, C. Xu, and B. Zhang, "Investigation of Flame Radiation Sampling and Temperature Measurement Through Light Field Camera," *International Journal of Heat and Mass Transfer* 121 (2018): 1281–1296, <https://doi.org/10.1016/j.ijheatmasstransfer.2018.01.083>.
83. C. Li, X. Jin, J. Li, and Q. Dai, "F-Number Adaptation for Maximizing the Sensor Usage of Light Field Cameras," in *2019 IEEE International Conference on Multimedia and Expo (ICME)* (2019): 109–114, <https://doi.org/10.1109/ICME.2019.00027>.
84. 'Illum Is Version 2.0 of the Lytro Camera You Can Refocus after Shooting, Video Might Be Next | no Film School', Joe Marine Apr 22, 2014. Accessed: Jul. 5, 2024. [Online]. Available: <https://nofilmschool.com/2014/04/lytro-illum-camera-refocus-images-video>.
85. Z. Liu, H. Hong, Z. Gan, J. Wang, and Y. Chen, "An Improved Method for Evaluating Image Sharpness Based on Edge Information," *Applied Sciences* 12, no. 13 (2022): 6712, <https://doi.org/10.3390/app12136712>.
86. D. G. Dansereau, D. L. Bongiorno, O. Pizarro, and S. B. Williams, "Light Field Image Denoising Using a Linear 4D Frequency-Hyperfan All-in-Focus Filter," in *Computational Imaging XI* (2013): 176–189.