Optical Aberration Analysis of Light Field Displays: A Calibration Approach for Enhanced Performance

Qi Zhang, Yuta Miyanishi, Erdem Sahin, and Atanas Gotchev; Tampere University; Tampere, Finland

Abstract

Aberrations in the optical system of a light field (LF) display can degrade its quality and affect the focusing effects in the retinal image, formed by the superposition of multiperspective LF views. To address this problem, we propose a method for calibrating and correcting aberrated LF displays. We employ an LF display optical model to subsequently derive the retinal image formation with a given LF input. This enables us to efficiently measure actual viewpoint locations and the deformation of each perspective image, by capturing focal-stack images during the calibration process. We then use the calibration coefficients to pre-warp the input images so that the aberrations are compensated. We demonstrate the effectiveness of our method on a simulated model of an aberrated near-eye LF display and show that it improves the display's optical quality and the accuracy of the focusing effects.

Introduction

The LF formalizes the spatial and angular dimensions of light rays in space [1]. LF displays aim at replicating the LF of a 3D scene by rendering multiperspective views of the scene to the viewers. A key feature of LF displays is that they can enable viewers to correctly accommodate to the displayed 3D objects, as long as the display's angular resolution is high enough to deliver multiple views to the viewer's pupils [2].

However, this feature of LF displays relies on the assumption of ideal optics, meaning that the optical components of the display are perfectly aligned and free from aberrations to ensure that the emitted light rays are correctly propagated. In reality, physical lenses used in LF displays introduce various optical imperfections, which significantly affect the display performance and the quality of the created imagery, especially in the case of near-eye LF displays. Our previous study revealed a significant discrepancy between the ideal model and the actual performance of a near-eye LF display test bed in terms of optical performance, and we observed noticeable errors in the display's focusing effects [3]. We hypothesized that the aberrations and misalignments of the display optics might alter the ray behaviours unpredictably as the observer's eye adjusts focus, which could degrade the optical resolution for each view and also potentially hinder proper accommodation responses due to errors in the perceived focusing effects.

In this study, we develop a method for calibrating the optics and correcting the aberrations in an LF display. Our method consists of three steps: (1) Modelling the retinal image formation by specifically representing transverse ray aberrations [4] in an LF display; (2) estimating the aberration parameters in this model using focal-stack images; and (3) pre-processing the input images with the inverse distortion parameters to cancel out the aberrations. We validate our method with simulations based on
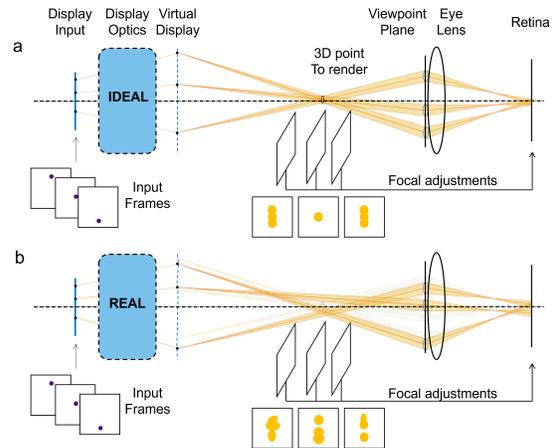

Figure 1. *Formation of retinal image by (a) ideal LF display and (b) aberrated LF display. Three perspective views (input frames) are present to generate the resulting superposition. Single activated pixels (in purple), are displayed in sequence within the corresponding views to render a 3D point in space, and combine into single retinal images in the form of three spots (in yellow) for varying accommodative states.*

the optical model of a table-top near-eye LF display setup.

Effects of Optical Aberrations on Observed Images

There are various possible implementations of LF displays, such as lens array based integral imaging, or super-multiview, or projector-based displays. Here we consider a time-multiplexing based display, especially suitable for near-eye use [3] [5]. As illustrated in Fig. 1, such a display consists of a display source plane (*display input*) that works in tandem with display optics to create the LF in the form of multiperspective images, corresponding to a given 3D scene. The time-multiplexing technique allows the input view images to be displayed consecutively in time through corresponding viewpoint locations with the help of its synchronized illumination/shuttering system (not plotted in the diagram). By 'viewpoints' we define a set of ideal pinholes corresponding to the angular (perspective) views, which in reality are finite-size sub-apertures parameterized on the *viewpoint plane*. Time multiplexing allows displaying view images with full spatial resolution and thus alleviates the spatio-angular resolution trade-off typical for lens-array based LF displays.

Fig. 1 illustrates also the retinal image formation for both an ideal LF display and a real (aberrated) LF display. Consider a 3D point in space to be rendered by the corresponding pixels
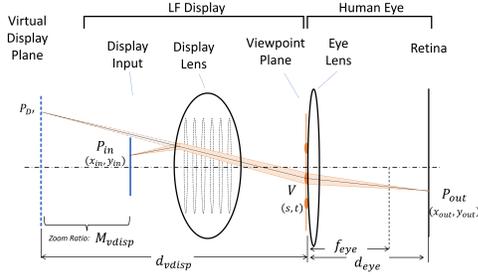
Figure 2. *Retinal image formation for a given display pixel, illustrating how the corresponding rays propagate through LF display optics onto the retina through a specific viewpoint.*

in three input view images. These pixels are lit by ray bundles (beams), which are projected onto the *virtual plane* and then directed toward corresponding viewpoints on the viewpoint plane, to eventually recombine on the retina of an ideal eye model. In the ideal case (Fig. 1 (a)), the beams from the three pixels converge at the 3D point in space. The retina captures different images of the scene by changing the eye lens focus, like slicing the scene at different depths and recording the lateral locations of the three spots, formed by intersection points of the three beams at these depths. When the eye changes its focus, the spots draw closer or diverge. The corresponding retinal images are sharpest when the eye's focus matches the depth of the rendered point, and gets blurry with excessive focusing powers. This is the expected focusing effect on the retinal image.

Optical aberrations disrupt the intended path of light rays due to variable refraction across different sections of the optics, leading to unpredicted deviations (Fig. 1 (b)). We pinpoint four key factors contributing to the difference between actual and ideal ray locations: *(i)* aberrated ray mapping from the display input to the virtual display plane; *(ii)* the virtual plane's non-planar geometry influencing distance to the viewpoint plane; *(iii)* chromatic dispersion arising from wavelength-dependent refraction; and *(iv)* variance in actual viewpoint locations compared to their intended locations. Collectively, these factors induce a lateral displacement of rays, known as *transverse ray aberrations*. These displacements prevent the anticipated pattern of converging, overlapping, and diverging of multiview rays' spots as the eye focus changes. Consequently, viewpoint images superimpose improperly, impairing focusing effects and overall optical performance.

Display–Retina Image Formation

To quantify the transverse ray aberrations that affect the viewer, we describe mathematically how rays are propagated from a display input pixel to the retina through a single viewpoint. Consider LF display optics formed by a stack of co-axial lenses, functioning as a magnifier that images the display plane to the virtual display plane, as illustrated in Fig. 2. In the ideal case, a pixel located at $P_{in} = (x_{in}, y_{in})$ on the display input plane generates a set of rays which goes through three steps to reach the retina. First, it is projected onto the virtual display plane with a magnification factor M_{vdisp}. Second, it is imaged through a viewpoint V located at (s,t). The distance between the virtual display plane and the viewpoint plane is denoted by d_{vdisp}. Third, it is formed as a spot pattern on the retina plane with its central loca-

tion $P_{out} = (x_{out}, y_{out})$. This location also depends on the current focal length of the eye lens f_{eye} and the lens-to-retina distance d_{eye}. The relationship between an input display pixel location (x_{in}, y_{in}) and the corresponding output retinal location (x_{out}, y_{out}) is described as:

$$\begin{bmatrix} x_{out} \\ y_{out} \end{bmatrix} = A_{ideal} \begin{bmatrix} x_{in} \\ y_{in} \end{bmatrix} + B(f_{eye}) \begin{bmatrix} s \\ t \end{bmatrix} + C_{ideal} \begin{bmatrix} s \\ t \end{bmatrix}, \quad (1)$$

where

$$A_{ideal} = -\frac{d_{eye} \cdot M_{vdisp}}{d_{vdisp}},$$

$$B(f_{eye}) = -\frac{d_{eye}}{f_{eye}}, \text{ and}$$

$$C_{ideal} = 1 + \frac{d_{eye}}{d_{vdisp}}.$$

Eq. (1) shows that the final retinal image is a magnified display input image offset by two central shift terms in relation to the viewpoint location. A_{ideal} is the magnification factor between the display input and the retinal image. The coefficients A_{ideal} and C_{ideal} are dependent on the display parameters, while the coefficient $B(f_{eye})$ does not, and varies only with the eye's focal adjustment f_{eye}.

In an aberration-free system, the display-related parameters A_{ideal} and C_{ideal} are constants. In an actual, aberrated system however, M_{vdisp} and d_{vdisp} are affected by rays' locations, which are further dependent on the viewpoint location V, wavelength λ, and pixel location P_{in}. Additionally, the actual viewpoints, denoted as $(s^{V,\lambda}, t^{V,\lambda})$, may deviate from the designed locations and be wavelength-dependent. This consequently modifies Eq. (1), to include the aberration effects to the display-to-retina image formation:

$$\begin{bmatrix} x_{out} \\ y_{out} \end{bmatrix} = A^{V,\lambda}(P_{in}) \begin{bmatrix} x_{in} \\ y_{in} \end{bmatrix} + B(f_{eye}) \begin{bmatrix} s^{V,\lambda} \\ t^{V,\lambda} \end{bmatrix} + C^{V,\lambda}(P_{in}) \begin{bmatrix} s^{V,\lambda} \\ t^{V,\lambda} \end{bmatrix}. \quad (2)$$

Specifically, the parameters A_{ideal} and C_{ideal} in Eq. (1) are now denoted by $A^{V,\lambda}(P_{in})$ and $C^{V,\lambda}(P_{in})$ to highlight their dependencies, which might lead to potential deformations that not only occur in individual view images but also feature inconsistencies among the images from different views across wavelengths. These parameters are not known a priori and have to be measured or estimated along with the viewpoint locations $(s^{V,\lambda}, t^{V,\lambda})$.

Calibration Using Focal-stack Images

The calibration of the actual viewpoint location $(s^{V,\lambda}, t^{V,\lambda})$ and the parameters $A^{V,\lambda}(P_{in})$ and $C^{V,\lambda}(P_{in})$ are conducted utilizing focal-stack images. We use a camera to replace the eye retina and vary its focal length $f_i, i \in [1..N]$ to capture a stack of images for every viewpoint $V_k, k \in [1..K]$ and every principal wavelength $\lambda_j, j \in [1..M]$. A display input pixel P_{in} at (x_{in}, y_{in}) is imaged onto the camera sensor, generating a spot with central location $(x_{out}^{V_k,\lambda_j,f_i}, y_{out}^{V_k,\lambda_j,f_i})$. Varying the camera focal length f_i changes the camera-only parameter $B(f_i)$ that results in generating the desired stack of differently focused images. The viewpoint location of

V_k in the given wavelength λ_j is calculated by differentiating the stack coordinates along the focal length dimension:

$$\begin{bmatrix} s^{V_k,\lambda_j} \\ t^{V_k,\lambda_j} \end{bmatrix} = \frac{1}{N-1} \sum_{i=2}^{N} \frac{1}{B(f_i)-B(f_{i-1})} \begin{bmatrix} x_{\text{out}}^{V_k,\lambda_j,f_i} - x_{\text{out}}^{V_k,\lambda_j,f_{i-1}} \\ y_{\text{out}}^{V_k,\lambda_j,f_i} - y_{\text{out}}^{V_k,\lambda_j,f_{i-1}} \end{bmatrix}. \quad (3)$$

Having the location of V_k calibrated, the parameters $A^{V_k,\lambda_j}(P_{\text{in}})$ and $C^{V_k,\lambda_j}(P_{\text{in}})$ are subsequently calculated as:

$$A^{V_k,\lambda_j}(P_{\text{in}}) = \frac{1}{N} \sum_{i=1}^{N} \frac{x_{\text{out}}^{V_k,\lambda_j,f_i} \cdot t^{V_k,\lambda_j} - y_{\text{out}}^{V_k,\lambda_j,f_i} \cdot s^{V_k,\lambda_j}}{x_{\text{in}} \cdot t^{V_k,\lambda_j} - y_{\text{in}} \cdot s^{V_k,\lambda_j}} \quad \text{and} \quad (4)$$

$$C^{V_k,\lambda_j}(P_{\text{in}}) = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_{\text{out}}^{V_k,\lambda_j,f_i} \cdot x_{\text{in}} - x_{\text{out}}^{V_k,\lambda_j,f_i} \cdot y_{\text{in}}}{x_{\text{in}} \cdot t^{V_k,\lambda_j} - y_{\text{in}} \cdot s^{V_k,\lambda_j}} - B(f_i) \right). \quad (5)$$

Correction by Warping the Input Image

The transverse ray aberrations are corrected through warping (pre-distorting) the input image in three steps. (1) Find the corrected, non-uniformly shifted, input pixel positions $\hat{P}_{\text{in}}^{V_k,\lambda_j} = (\hat{x}_{\text{in}}^{V_k,\lambda_j}, \hat{y}_{\text{in}}^{V_k,\lambda_j})$ from their ideal locations $P_{\text{in}} = (x_{\text{in}}, y_{\text{in}})$ using the calibrated parameters. (2) Given the non-uniform grid of correctly-positioned input pixels, fit a least-squares cubic spline with knots at the initial uniform grid. (3) Sample the resulting spline at the knots (these are the input sources of light) [6].

To find the corrected input pixel positions, we first express any output pixel position for the specific viewpoint locations $(s^{V_k,\lambda_j}, v^{V_k,\lambda_j})$ and wavelengths λ_j using Eq. (1). We substitute it in the left-hand side of Eq. (2) for the real case and solve the equation for the sought (shifted) new input positions. This leads to

$$\begin{bmatrix} \hat{x}_{\text{in}}^{V_k,\lambda_j} \\ \hat{y}_{\text{in}}^{V_k,\lambda_j} \end{bmatrix} = \frac{A_{\text{ideal}}}{A^{V_k,\lambda_j}(P_{\text{in}})} \begin{bmatrix} x_{\text{in}} \\ y_{\text{in}} \end{bmatrix} + \frac{C_{\text{ideal}} - C^{V_k,\lambda_j}(P_{\text{in}})}{A^{V_k,\lambda_j}(P_{\text{in}})} \begin{bmatrix} s^{V_k,\lambda_j} \\ t^{V_k,\lambda_j} \end{bmatrix}. \quad (6)$$

In this way, the pre-distortion process aligns the aberrated display's results with those of an ideal display, achieving the same sensor (retina) output locations. Eq. (6) achieves precise correction results if the following approximation holds:

$$A^{V_k,\lambda_j}(P_{\text{in}}) \approx A^{V_k,\lambda_j}(\hat{P}_{\text{in}}^{V_k,\lambda_j}) \text{ and } C^{V_k,\lambda_j}(P_{\text{in}}) \approx C^{V_k,\lambda_j}(\hat{P}_{\text{in}}^{V_k,\lambda_j}).$$

Simulation Results

We validate our method by simulations using the Zemax OpticStudio optical simulation software, hereafter referred to as "ZEMAX". Specifically, we integrate the calibration and aberration correction phases into the model of a table-top near-eye LF test bed and analyze the results through simulations. Fig. 3 illustrates the workflow of our method. In the calibration phase, we model the LF display in the simulation environment and measure its output using a simulated varifocal camera. This gives us a set of differently focused images (focal stack) for all views and principal wavelengths the display supports. We apply the proposed calibration procedure on these focal-stacks to obtain the calibration parameters. In the correction phase, we use these parameters
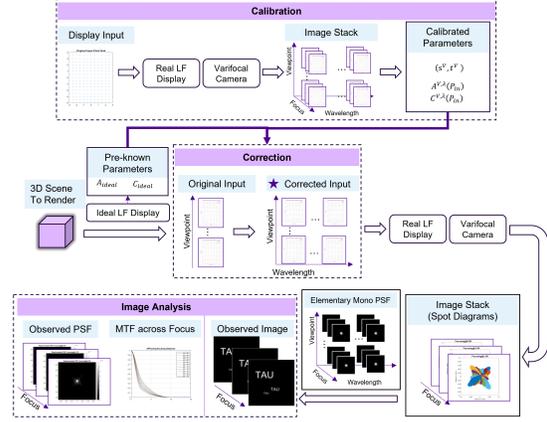

Figure 3. *Simulation workflow of an LF display viewing system. With calibrated parameters of a real LF display model, the display input can be corrected for better results, i.e., better image quality and proper focusing effects.*

to warp the input images for each view and each wavelength of the display, so that the output images are closer to the ideal ones.

With reference to Figure 3, below we discuss the key components and settings of the simulation model in more detail.

Real LF Display: The real LF display model closely resembles the actual experimental arrangement previously described. An effective input field of ± 2.5 mm horizontally and ± 4.0 mm vertically with 5 um pixel intervals is configured in the simulation, covering the central field of the display panel. For simplification, each pixel has an infinitesimal size. The LF optics consists of six practical lenses, functioning collectively as a microscopic system. It inversely magnifies the input display image, producing an enlarged, inverted image on a virtual display plane.

Camera Model: A perfect lens and an image receiver (sensor) in ZEMAX simulate the varifocal camera. During the simulation process, the distance between the lens and sensor is specified as $d_{cam} = 31.89$ mm, while the lens's focal length is changed to achieve different focal depths.

Ideal LF display: The Ideal LF display model represents our target display configuration, magnifying the display input plane to a virtual display plane through specified zoom ratio $M_{\text{vdisp}} = -84.5$ and distances $d_{\text{vdisp}} = 984.32$ mm relative to the viewpoint plane. Therefore, the parameters in the ideal model in Eq. (1) can be computed as $A_{\text{ideal}} = 2.74$, and $C_{\text{ideal}} = 1.03$.

Viewpoint Locations: In the case of a real LF display, due to imperfections in the illumination/shuttering system, the viewpoints are no longer distributed regularly but have offsets from ideal locations. To mimic this practical tolerance in simulation, viewpoint locations are arranged in an irregular grid. This configuration consists of $6 \times 6 = 36$ virtual viewpoints with an average spacing of approximately 0.5 mm, as shown in blue circles in Fig. 4(a). Considering practical states, each viewpoint is not a perfect pinhole but a sub-aperture with a diameter of 0.3 mm. The viewpoint locations and the sub-aperture size are set to be invariant to the wavelength.

Wavelengths: Three principal wavelengths are considered in the simulation: 470 nm, 550 nm, and 650 nm.

Ray Tracing Settings: In the sub-aperture of each view, a dense uniform grid is configured for sampling. Then a set of rays
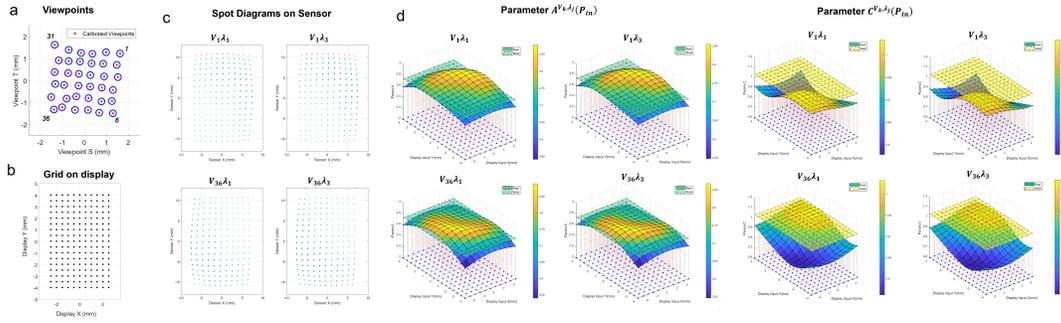
Figure 4. *Calibration input, output, and results. (a) Distribution of viewpoints. The centers of the blue circles represent the simulation-configured viewpoint locations, and the radius demonstrates sub-aperture size. Red crosses illustrate calibrated locations. (b) Display input grid. Each point represents the pixel location for calibration. (c) Spot diagrams on the camera sensor with the grid as the display input. The subtitle indicates the viewpoint and wavelength indexes considered in the diagram. All the diagrams are in the condition that the camera focuses at 1 diopter (1D). (d) Calibrated parameters A and C for different views and wavelengths. They are illustrated in curved surfaces. Note that within the same view, the behaviors of A and C vary across different wavelengths, despite their similar appearances. Dashed planes represent ideal values of A_{ideal} and C_{ideal} for comparison.*

is created, originating from one input pixel on the display plane to the points in the grid in the sub-aperture, just as the orange beam illustrated Fig. 2. In ZEMAX these rays are traced and their intersection locations on the camera sensor are calculated, which eventually form the corresponding output spot diagram. This process is employed repeatedly for each wavelength, each view, and across the camera's focal shifts, i.e., achieving sets of focal stack images.

Calibration

We utilize a 11×17 grid as the input "chart" for calibration, spanning the display input field, as depicted in Fig.4(b). The grid points are at intervals of 0.5 mm, which cover 100 pixels, and each grid point represents an activated pixel. We adjust the camera's focal length to focus on five distinct depths ranging from 0.6 diopters (0.6D) to 1.4D with incremental steps of 0.2D. For each focal depth, we record the spot diagrams on the camera sensor from all the points in the input grid for all 36 views and 3 wavelengths. Some examples of the resulting output grids are illustrated in Fig. 4(c). Each spot's central location is represented by its mass center coordinate.

The viewpoint locations, denoted now as (s^{V_k}, t^{V_k}) as simplified to be independent of wavelength, can be obtained by averaging the results calculated from each point in the grid in each wavelength utilizing Eq. (3). As demonstrated with red crosses in Fig. 4 (a), this process yields highly accurate calibration results for the viewpoint locations.

Subsequently, the parameters A and C corresponding to the sparse sampling locations of the 11×17 input grid are derived from the calibrated viewpoint locations using Eq. (4) and Eq. (5). The values of A and C for the original-resolution (dense) input grid are found through cubic interpolation from the sparse ones. This approach gives a good approximation with less than 0.1% error between the true dense parameters and their interpolated values, corresponding to ± 2 um of tolerance in the predicted output locations $P_{\text{out}} = (x_{\text{out}}, y_{\text{out}})$ by Eq. (2).

Fig. 4(d) illustrates how parameters A and C fluctuate in relation to three factors: the spatial location of the input pixel, the

viewpoint, and the wavelengths. In comparison, A_{ideal} and C_{ideal} in the ideal model, plotted as a plane with dashed outlines, are constants and unaffected by these factors.

The variations of the realistic A and C across the viewpoints and wavelengths elucidate the reason for the variation in the output grid's shape on the sensor in Fig. 4 (c). It also reveals that images from different wavelengths and views will potentially exhibit inconsistent deformations in both shape and location, culminating in the observed superposition challenge.

Correction

After calibrating the parameters, the input images can be corrected to enhance the output quality. This section showcases three examples where different scenes are rendered to demonstrate the effectiveness of our correction approach. Initially, we render a 3D point in space. The simulation is then expanded in the second case to include rendering 3D points at varying depths and locations. The final simulation depicts the rendering of three planar patterns in space.

CASE1: Superposition of spot diagrams and observed MTFs due to an on-axis point at 1D

In the first example, a 3D point is rendered on the optical axis and positioned at a distance of 1D from the observing camera. A single point source is activated in the input image of each view and wavelength. Ray tracing with simulation yields superposition of spot diagrams for 36 views and 3 wavelengths on the sensor for one specific focal distance.

A focal stack containing 5 images, ranging from 0.0D to 2.0D with increments of 0.5D, captures these spot diagrams, as shown in the top row of Fig. 5(a). As a comparison, the second row in the figure depicts the diagrams without the input correction. In the corrected case, the spot diagrams from various views and wavelengths exhibit a coherent trajectory, converging and then diverging. The spots' centers are aligned at the 1D focus depth corresponding to the rendered point's depth. Conversely, the spots in the uncorrected case do not hold a clear focusing trend. Moreover, they are distributed in a larger field, potentially leading to a worse image quality.
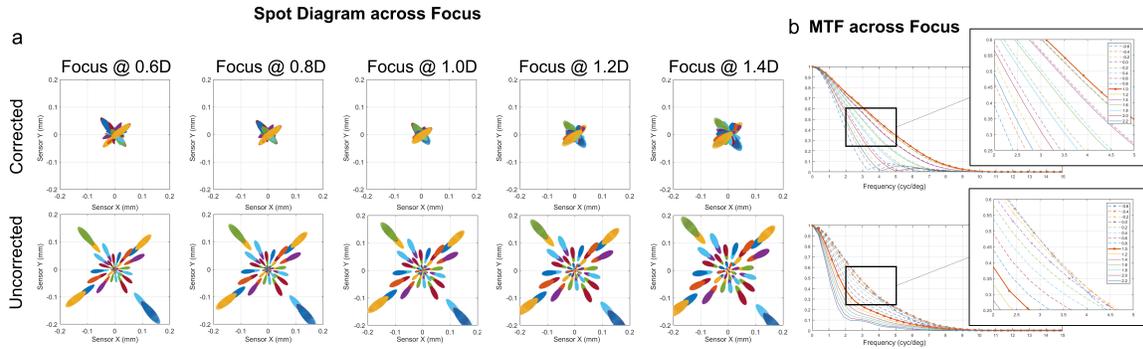
Spot Diagram across Focus

Figure 5. *Simulation outcomes for CASE 1: The first row displays results with corrected input, and the second row features the uncorrected case. (a) Superposition of spot diagrams on the sensor across focus. Every image contains spots plotted in different colors from 36 views in 3 wavelengths. The rendered point in space is located at 1D. (b) Observed MTF curves across focus. Cross-solid curves represent the result when the camera focuses at 1D. Best MTF performances in uncorrected cases are plotted in cross-Dashed curves.*

In addition to spot diagrams, we assess the image quality through the elementary monochromatic PSF, which is generated on the sensor from an input pixel, in each view, each wavelength, and each camera's focusing depth. We extract these PSFs in ZE-MAX, position them on the mass centers of corresponding spots, and then aggregate them to achieve the final observed (retinal) PSF on the sensor. Eventually, the optical transfer function (OTF) and MTF for each camera's focusing depth are obtained from the Fourier transform of the observed PSF.

Fig. 5 (b) illustrates the MTF curves across varying focal depths from -0.6D to 2.2D, comparing the corrected and uncorrected inputs (negative diopters means "beyond infinity", theoretically indicating that the camera's focal depth is virtually behind the viewpoint plane rather than in front of it). Consistent with the findings from spot diagrams, the observed MTFs with the corrected inputs demonstrate a pronounced focusing effect, peaking at the depth of the rendered point, while those with the uncorrected inputs exhibit a focusing trend not in line with the intended design, i.e., the best MTF is not obtained at the depth of the rendered point, but at the depth of "further" than 0D. Furthermore, there is also a notable improvement in the corrected case in terms of best MTF magnitudes, which directly translates to better observed image quality.

CASE2: Observed cut-off frequencies of rendering points at different positions

We validate our approach by rendering scenes at various locations in space. Specifically, we render 6 points: either on-axis or off-axis and at one of the three depths of 0.2D, 1D, or 2D. In scenarios where each point is individually rendered, spot diagrams are recorded as the camera's focus is adjusted within a range of -1.2D to +1.2D relative to the depth of each rendered point; for example, for points at 0.2D, the focus range covers 0D to 1.4D. Subsequently, the observed MTFs across these focal adjustments are computed. Fig. 6 shows the cut-off frequencies of the MTFs as a function of the camera's focus for each scenario for better visibility of the focusing effects. The threshold for deriving the cut-off frequencies is set to 0.1.

In all corrected cases, the observed image achieves the best performance when the camera focuses on the exact depth of the rendered point, while image quality worsens as the focal depth
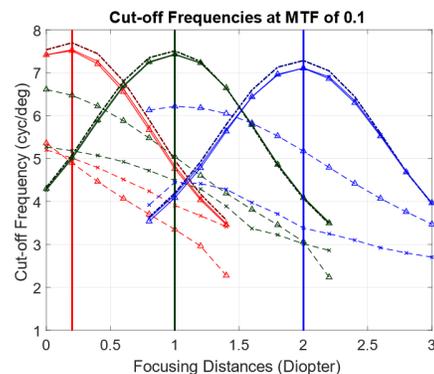

Figure 6. *CASE2 results: Cut-off frequencies at value of 0.1 in the observed MTFs. Vertical lines in red, green, and blue denote depths of 0.2D, 1.0D, and 2D, at which each point is rendered. Curves with the same color as the vertical lines represent cross-focus cut-off frequencies corresponding to the rendered point at that depth. With reference to different simulation conditions: triangle-solid curves depict cut-off frequency variation across focus for corrected input when rendering an on-axis point, cross-solid curves represent corrected but off-axis scenario. The corrected curves for the on- and off-axis points rendered at identical depths nearly overlap. The uncorrected case is represented by triangle-dashed and cross-dashed curves, indicating on-axis and off-axis rendering, respectively. Dash-dotted curves represent the ideal scenarios that only consider the diffraction limitation due to the viewpoint size.*

moves far away. The results confirm the efficacy of our method; utilizing the pre-distorted display inputs leads to the proper focusing effects. The focusing trends in uncorrected inputs, however, become less pronounced and vary depending on the locations of the rendered targets. Furthermore, at the best focal depth, the image quality (characterized by the cut-off frequency) in the corrected scenarios significantly surpasses that of the uncorrected ones, demonstrating comparable parity with those ideal results.

CASE3: Observed images of rendering a 3D scene

For a visual assessment of the proposed method, we render a 3D scene with three planar targets, set at 0.2D, 1.0D, and 1.7D, respectively, as illustrated in Fig. 7 (a).

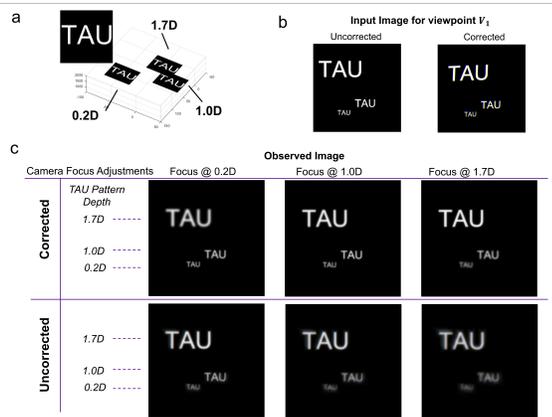**Figure 7.** *CASE3 simulation. (a) The "TAU" image, and the arrangement of the three image copies in the rendered 3D scene. (b) Input images of view V_1 showcase the differences between original (uncorrected) image and the corrected one. (c) Observed images across the camera's focus, with the first row in the corrected cases and the second row in the uncorrected cases.*

The input (uncorrected) images for each viewpoint are warped to their corrected versions, as shown in Fig. 7(b) for viewpoint 1. The latter exhibits pre-distortion effects such as deformed shape and color fringes. An elementary monochromatic PSF represents the output of an on-axis input point through a specific sub-aperture and for a particular wavelength. It is instrumental in computing the viewpoint-specific sub-image on the sensor taking into account the diffraction effects in each wavelength channel. Experimental verification confirmed that the elementary PSF shows negligible deviation from the actual PSF corresponding to off-axis input pixels. Therefore, the viewpoint sub-image is formed as a superposition of the PSFs, centred according to Eq (2) and weighted by the corresponding input pixels. The final retinal image is obtained as a superposition of all viewpoint sub-images across all wavelength channels.

Fig. 7(c) illustrates the results when the camera focuses at 0.2D, 1.0D, and 1.7D, respectively. With the corrected display inputs (first row), the simulated sensor image achieves the sharpest "TAU" pattern when the camera focuses on the exact depth of that rendered target. The images of the targets at the other depths show proper defocus blur. In contrast, focusing effects are less pronounced in the cases without the input correction (second row). For all objects, the corrected display inputs provide noticeably better observed image quality (sharpness) at the focused state. These observations are in accordance with the previous experimental results, demonstrating the potential of the proposed method for enhancing the performance of LF displays.

Conclusion

This study illustrated that the transverse ray aberration in an LF display is one of the important factors affecting the quality of observed images in terms of both the accommodative response and the optical resolution. Through a model for the retinal image formation and utilizing focal-stack images, we introduced a novel approach for calibrating these aberrations and compensating for their effects through the pre-distortion of display inputs. The efficacy of the proposed method was substantiated through the simulations. Specifically, the simulated MTFs and observed images of a 3D scene indicated that the image quality and focusing effects are improved by correcting the aberrations with the proposed method. One of the primary objectives in the future is to test this method in an actual LF display for further refinement of the approach.

Acknowledgments

References

[1] A. Gershun, "The light field," *Journal of Mathematics and Physics*, vol. 18, no. 1-4, pp. 51–151, 1939.

[2] Y. Kajiki, H. Yoshikawa, and T. Honda, "Hologramlike video images by 45-view stereoscopic display," in *SPIE Electronic Imaging*, S. S. Fisher, J. O. Merritt, and M. T. Bolas, Eds., vol. 3012. International Society for Optics and Photonics, May 1997, p. 154.

[3] Y. Miyanishi, Q. Zhang, E. Sahin, and A. Gotchev, "Observed optical resolution of light field display: Empirical and simulation results," in *Electronic Imaging*. Society for Imaging Science and Technology, Jan. 2023, pp. 1–6.

[4] J. Sasián, *Ray aberrations*. Cambridge University Press, 2012, p. 100–118.

[5] T. Sluka, "Near-eye sequential light-field projector with correct monocular depth cues," Patent WO2 018 091 984A1, May 24, 2018.

[6] M. Unser, "Splines: a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.

Author Biography

Qi Zhang received his B.E. degree in information engineering and M.S. degree in optical engineering from Zhejiang University in 2011 and 2014, and he is now a Ph.D. student at Tampere University in Finland. His research focuses on the implementation of 3D displays.

Yuta Miyanishi received his B.E. degree in computer science and M.S. degree in human centered science and biomedical engineering from Tokyo Institute of Technology in 2016 and 2018, and he is currently pursuing his doctoral degree at Tampere University in Finland. After his study on stereopsis in a psychophysical approach in Tokyo Tech, he is now involved in theoretical and empirical research on light field displays and human observers.

Erdem Sahin received the Ph.D. degree from the Electrical and Electronics Engineering Department, Bilkent University in 2013. In 2014, he joined the 3D Media Group at Tampere University, as Marie Curie experienced Researcher, where he has been Senior Research Fellow since 2019. His current research interests include development of computational light field and holographic imaging algorithms and methods for next-generation 3D cameras and displays.

Atanas Gotchev received the M.Sc. degrees in radio and television engineering (1990) and applied mathematics (1992), the Ph.D. degree in telecommunications from Technical University of Sofia (1996), and the D.Sc.(Tech.) degree in information technologies from Tampere University of Technology (2003). He is currently Professor of Signal Processing with Tampere University. His recent work concentrates on developing methods for multi-sensor 3D scene capture, and light field imaging and display.