photonics

Article

An Integrated Tunable-Focus Light Field Imaging System for 3D Seed Phenotyping: From Co-Optimized Optical Design to Computational Reconstruction

Jingrui Yang [1,2], Qinglei Zhao [1], Shuai Liu [1,2], Meihua Xia [3,4], Jing Guo [1,2], Yinghong Yu [3,4], Chao Li [5], Xiao Tang [3], Shuxin Wang [1], Qinglong Hu [1], Fengwei Guan [1], Qiang Liu [1], Mingdong Zhu [3,4,*] and Qi Song [6,*]

1 Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
2 University of Chinese Academy of Sciences, Beijing 100039, China
3 State Key Laboratory of Hybrid Rice, Hunan Academy of Agricultural Sciences, Changsha 410125, China
4 Yuelushan Laboratory, Changsha 410128, China
5 Qualcomm Inc., San Diego, CA 92121, USA
6 Soleilware Photonics LLC, San Diego, CA 92131, USA
* Correspondence: uhz_uhz@hotmail.com (M.Z.); songq@sibet.ac.cn (Q.S.)

Abstract

Three-dimensional seed phenotyping requires imaging systems capable of achieving micron-level resolution across a centimeter-level field of view (FOV), a goal constrained by the resolution–FOV trade-off in conventional light field architectures. This paper presents a hardware–software co-optimized framework that integrates a reconfigurable optical system with computational imaging pipelines to address this limitation. At the hardware level, we develop a tunable-focus lens module that enables flexible adjustment of the effective focal length, combined with a custom-designed microlens array (MLA). A mathematical model is established to analyze the interdependencies among FOV, lateral resolution, depth of field (DOF), and system configuration, guiding the design of individual optical components. On the computational side, we propose a hybrid aberration correction strategy: first, a co-calibration of lens and MLA aberrations based on line-feature detection; second, a conditional generative adversarial network (cGAN) with attention-guided residual learning to enhance sub-aperture images, achieving a PSNR of 34.63 dB and an SSIM of 0.9570 on seed datasets. Experimentally, the system achieves a resolution of 6.2 lp/mm at MTF50 over a 2–3 cm FOV, representing a 307% improvement over the initial configuration (1.52 lp/mm). The reconstruction pipeline combines epipolar plane image (EPI) analysis with multi-view consistency constraints to generate dense 3D point clouds at a density of approximately 1.5×10^4 points/cm^2 while preserving spectral and textural features. Validation on bitter melon and rice seeds demonstrates accurate 3D reconstruction and accurate extraction of morphological parameters across a large area. By integrating optical and computational design, this work establishes a reconfigurable imaging framework that overcomes the resolution–FOV limitations of conventional light field systems. The proposed architecture is also applicable to robotic vision and biomedical imaging.

Keywords: light field imaging; microlens array; depth estimation; aberration correction; 3D reconstruction

Received: 14 March 2026
Revised: 4 April 2026
Accepted: 14 April 2026
Published: 17 April 2026

1. Introduction

Three-dimensional seed phenotyping is a critical tool for digitizing germplasm resources and supporting functional genomics research [1–4]. Effective phenotyping requires imaging systems capable of resolving both macroscopic morphological parameters, including overall shape and volume, as well as microscopic surface features, including embryonic structures and surface defects, across seed varieties that differ substantially in physical scale (e.g., rice grains vs. bitter melon seeds). In addition, full 3D characterization is necessary for assessment of germination rates, seed vigor, varietal purity, and stress resistance [5–7]. These requirements translate directly into two imaging demands that must be met simultaneously: a wide FOV to cover the full seed surface, and high lateral resolution to resolve fine surface detail.

Achieving large FOV and high spatial resolution simultaneously is a long-standing problem in optical imaging [8], rooted in the fundamental trade-off between optical magnification and spatial coverage. Conventional approaches such as microscope image stitching or multi-camera arrays can partially circumvent this trade-off, but at the cost of mechanical scanning, extended acquisition times, or increased system complexity, all of which limit their applicability in high-throughput scenarios. Light field imaging offers a more compact alternative: by recording both spatial and angular information in a single exposure via an MLA, it enables computational refocusing and 3D reconstruction without mechanical movement [9–11]. Focused plenoptic cameras, in particular, provide a useful balance between lateral and angular resolution [12,13]. However, existing light field architectures are constrained by their static optical design. Standard plenoptic cameras suffer from low lateral resolution due to the fixed microlens-to-pixel mapping, while focused designs, though improved, still operate within a fixed parameter space that cannot adapt to objects of different physical scales [14–16]. The short baseline inherent to MLA-based systems further limits depth resolution, degrading EPI-based depth estimation for fine structural features [17,18], while wide-FOV configurations exacerbate off-axis aberrations such as distortion and field curvature [19]. As a result, the resolution–FOV trade-off remains unresolved in current light field systems [20].

To address these limitations, this paper presents a hardware–software co-optimized light field imaging framework built around a tunable-focus lens module. On the hardware side, we replace the fixed main lens with a tunable-focus element rather than a conventional zoom lens. A conventional zoom lens achieves focal length variation by translating internal lens groups, which alters the overall optical path geometry and complicates the coupling with downstream MLA-based sampling. The tunable-lens approach adjusts the effective focal length by varying only the separation between two fixed lens elements, preserving the mechanical alignment between the main lens, MLA, and sensor throughout the adjustment range. This enables flexible adjustment of the effective focal length, allowing the system to match the geometric scale of different seed varieties while maintaining the same MLA and sensor. A mathematical model is established to quantify the interdependencies among FOV, lateral resolution, DOF, and system configuration, guiding the co-design of the main lens assembly and the custom MLA. Finite-difference time-domain (FDTD) simulations further inform MLA optimization by analyzing focusing efficiency and crosstalk. On the computational side, a two-stage aberration correction pipeline addresses the optical degradation introduced by the tunable optics: a co-calibration procedure based on line-feature detection first corrects geometric distortions across the lens and MLA, followed by a cGAN with attention-guided residual learning to suppress residual optical aberrations in sub-aperture images. The corrected sub-aperture images are then passed to a reconstruction pipeline that integrates EPI analysis with multi-view consistency constraints to generate dense 3D point clouds with preserved surface texture. Validation on bitter melon and rice seeds demonstrates that the proposed framework

https://doi.org/10.3390/photonics13040385

effectively mitigates the resolution–FOV trade-off across different imaging scales. The system is also applicable to robotic vision, biomedical imaging, and industrial inspection.

The remainder of this paper is organized as follows. Section 2 presents the design methodology, covering the tunable light field system architecture, mathematical modeling of optical parameters, and the image processing pipeline. Section 3 details experimental validation, including resolution characterization, deblurring performance evaluation, and 3D reconstruction results on bitter melon and rice seeds. Section 4 concludes with a summary of findings and directions for future work.

2. End-to-End Design Methodology for Tunable Light Field System

2.1. Architectural Design of Light Field System

This section focuses on the optical design and analysis of the light field system. In the context of 3D imaging for seed phenotyping, the system design must meet several critical requirements simultaneously. First, it should enable high-precision imaging over a relatively large field of view to record the diverse seed sizes and capture the details at the same time. Second, extended DOF is essential to ensure that the micro-bumps and cavities at the surface are captured and restored flawlessly. We construct a system model based on geometrical optics and analyze key design parameters, such as the lateral resolution, FOV, DOF, and depth resolution, in relation to lens parameters. However, the design of light field systems generally involves more trade-offs than that of conventional cameras. This work starts with a seed phenotyping case but is not exclusively limited to this application. We aim to provide a broader perspective on the system construction process that can be adapted to various scenarios.

2.1.1. Description of Light Field System Design

Serving as the mathematical foundation of light field imaging, the plenoptic function $P(\theta, \varphi, \lambda, t, x, \gamma, z)$ is described as the seven-dimensional information carried by light rays propagating in 3D space, including 3D spatial coordinates, 2D propagation directions, wavelength, and time [21]. A focused light field camera employs a synergistic design of an MLA with a high-resolution CMOS sensor to perform four-dimensional structured sampling of the plenoptic function. At the sensor plane, both the spatial intensity distribution (x, y) and angular propagation characteristics (u, v) of incident light rays are recorded, constructing a discretized light field matrix $L(x, y, u, v)$. The core principle lies in the wavefront sampling effect of microlenses on the object-side light field: each microlens unit projects light rays from an object point at different angles onto corresponding sub-aperture regions of the sensor, forming sub-images with spatial-angular correlations. For focused light field cameras, complex coupling and trade-offs exist among FOV, lateral resolution, depth resolution, and DOF.

In a focused light field system, the distance between the MLA and the CMOS sensor is close but not equal to the focal length of the microlens. To maximize sensor pixel utilization, the F-numbers $(F/\#)$ of the main lens and microlens should be ideally matched. However, this constraint may not hold for a light field system with a tunable focus. According to the Nyquist sampling theorem, the object-side resolution must cover at least two pixels while balancing FOV requirements. The Nyquist theorem and MLA parameters apply the constraint on the object magnification β, where d_{pixel} is the pixel size and ϵ_{obj} is the desired object-side resolution. The lower bound for β is:

$$\beta \geq \frac{2d_{\mathrm{pixel}}}{\epsilon_{\mathrm{obj}}} \tag{1}$$

The practical magnification β' is calculated as:

$$\beta' = \frac{S_{sensor}}{\delta_{\mathrm{obj}}} = \frac{X_{MLA} \cdot \Delta N \cdot d_{\mathrm{pixel}}}{\delta_{\mathrm{obj}}} \tag{2}$$

X_{MLA} denotes the number of microlenses along the x-axis, ΔN is the number of pixels per macro-pixel shown in Figure 1, S_{sensor} denotes the physical length of the sensor along the corresponding axis, and δ_{obj} is the desired object-side FOV.

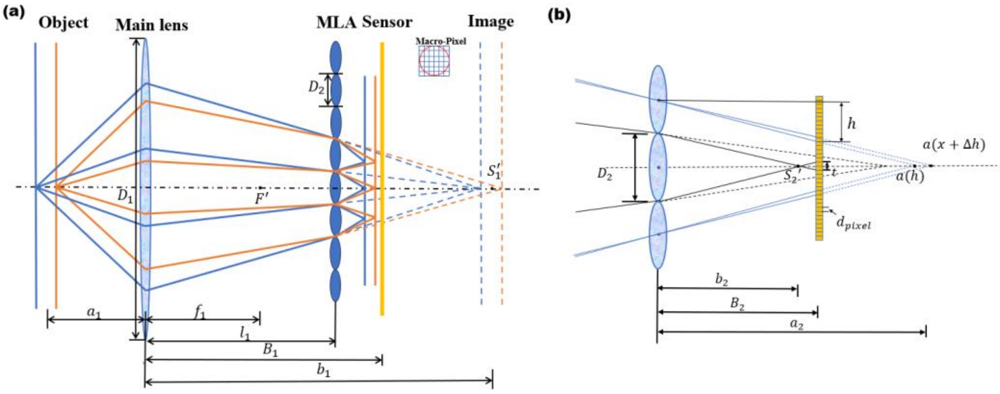

Figure 1. Optical path diagram. (**a**) Focused light field system (Galilean type) and schematic of macro-pixels. The orange, blue, and black rays represent light from objects at three different depth positions. (**b**) MLA defocuses on the sensor and depth of field in the focused light field camera.

2.1.2. Relationship Between FOV and Focal Length

The light field system is able to partially mitigate the inherent conflict between large FOV and high resolution through spatial-angular joint sampling and computational imaging algorithms. In traditional cameras, FOV is determined by the focal length f and CMOS plane size S:

$$FOV = 2\arctan\left(\frac{S}{2f}\right) \quad (3)$$

where increasing FOV requires reducing f or increasing S, which often amplifies off-axis aberrations (e.g., coma, distortion) at large incident angles. The light field system leverages the MLA to distribute light rays from different angles onto the sensors, effectively enabling multi-view joint sampling. FOV is primarily governed by the main lens; if the main lens image fully covers the MLA, the FOV matches that of a traditional camera. However, in focused light field designs, the F-number of the microlens must balance sub-image pixel resolution and DOF, and minimize the macro-pixel crosstalk, which will be discussed in detail later.

2.1.3. Relationship Between Lateral Resolution, DOF, and Depth Resolution

The core performance of a focused light field camera is governed by a tripartite coupling among lateral resolution, depth resolution, and DOF; therefore, we derive a unified mathematical model to locate the optimal operating region where the system maximizes information capture within the physical diffraction and sampling limits.

Lateral Resolution and DOF

According to Fraunhofer diffraction theory, an ideal aberration-free lens forms an Airy disk when imaging a point source. The angular radius θ of the first dark ring is:

$$\theta \approx 1.22\frac{\lambda}{D} \approx 1.22\frac{F/\# \cdot \lambda}{f} \quad (4)$$

Here, D is the entrance pupil diameter, and $F/\# = f/D$, with f as the focal length. Lateral frequency could be measured as the number of line pairs per millimeter (lp/mm).

Diffraction limit: Normally the diffraction limit sets the upper bound for system resolution. For the main lens and microlens, the diffraction-limited resolution is:

$$\nu_{diff} = \frac{1}{1.22 \cdot \lambda \cdot F/\#} \quad (5)$$

CMOS sensor: As the final stage in the imaging system, the sampling frequency follows the Nyquist criterion:

$$\nu_{Nyquist} = \frac{1}{2 \cdot d_{pixel}} \quad (6)$$

MLA sampling frequency: The spatial sampling frequency of an MLA is defined as the reciprocal of the microlens pitch p_{MLA} with system magnification M, which is the ratio of focal length f_1 to object distance a_1. It quantifies the number of microlenses per unit length, determining the Nyquist limit for spatial information capture in the object or image plane.

$$\nu_{MLA} = \frac{M}{2 \cdot p_{MLA}} \quad (7)$$

Thus, the total system resolution is jointly managed by the lowest sampling frequency across all stages:

$$\nu_{\text{total}} = \min\left(\nu_{diff}, \nu_{\text{Nyquist}}, \nu_{MLA}\right) \quad (8)$$

Taking the Galilean configuration (Figure 1) as an example, the minimum resolvable spot diameter s is determined by the sensor pixel size d_{pixel} and the microlens numerical aperture. To reconstruct depth information, the intermediate image must be collected at least by two microlenses, which defines the important theoretical detection limit for the depth estimation. Here we need to make the assumption that the MLA is simplified as a thin lens, where D_2 is the microlens diameter, B_2 is the distance between the microlens and sensor plane, f_2 is the focal length, a_2 is the distance from the intermediate image point to the microlens and d_{pixel} is the pixel size.

The lateral resolution and DOF are a pair of design trade-offs in most optical imaging systems. Here we try to deliver a general analysis of the effective resolution ratio (ERR) and DOF according to plenoptic lens parameters.

First, S_1 and S_2 are defined as the spot size of the main lens and MLA at the image plane and are defined as

$$S_1 = D_1 \cdot \left|\frac{B_1 - b_1}{b_1}\right| \quad (9)$$

$$S_2 = D_2 \cdot \left|\frac{B_2 - b_2}{b_2}\right| \quad (10)$$

where $b_1 = \frac{f_1 a_1}{a_1 - f_1}$ and $b_2 = \frac{f_2 a_2}{a_2 - f_2}$ are derived from the well-known thin lens equation. The effective resolution of the imaging system is jointly constrained by three physical mechanisms, each imposing a lower bound on the minimum resolvable spot size at the sensor plane. The first is the geometric blur spot s_i, which arises from defocus of the lens and scales with the aperture diameter and the deviation between the actual and ideal image distances. The second is the pixel pitch d_{pixel} of the CMOS sensor, which sets the Nyquist sampling limit: even if the optical system delivers a point spread function smaller than a single pixel, the discrete sampling grid cannot resolve features below this scale. The third is the diffraction-limited spot size $s_0 = 1/\nu_{\text{total}}$, determined by the Airy disk radius of the lens under the given F-number and wavelength. These three constraints originate from geometrical optics, sampling theory, and wave optics, respectively, and they act as independent bottlenecks on the imaging chain. The actual minimum resolvable spot diameter is therefore governed by whichever mechanism produces the largest blur contribution. Accordingly, we define the effective resolution ratio ERR_i using the maximum of the three quantities as the denominator, so that ERR_i reflects the true resolving capability of the system under all operating conditions:

$$ERR_1 = \frac{d_{pixel}}{\max\left[s_1, d_{pixel}, s_0\right]} = \frac{d_{pixel}}{\max\left[\left|D_1\left(\frac{B_1}{b_1} - 1\right)\right|, d_{pixel}, \frac{1}{\nu_{\text{total}}}\right]} \tag{11}$$

$$ERR_2 = \frac{1}{\nu}\frac{d_{pixel}}{\max\left[s_2, d_{pixel}, s_0\right]} = \frac{1}{\nu}\frac{d_{pixel}}{\max\left[\left|D_2\left(\frac{B_2}{f_2} - \frac{1}{\nu} - 1\right)\right|, d_{pixel}, \frac{1}{\nu_{\text{total}}}\right]} \tag{12}$$

Here, ERR_1 and ERR_2 are utilized to define the spot projection and effective resolution ratio from main lens and MLA, respectively. The axial distance from the image formed by the main lens to the microlens array is denoted as a_2, and the distance from the microlens array to the sensor is denoted as B_2. The ratio of these two distances is defined as the virtual depth, expressed as $\nu = \frac{a_2}{B_2}$. The virtual depth can be regarded as the projection of a real main lens image point onto the camera's image plane by a microlens centered on the main lens's optical axis. Accordingly, the lateral resolution on the object side can be expressed as follows:

$$LR = \frac{1}{\nu}\frac{1}{\max\left[s_2, d_{pixel}, s_0\right]} \cdot \frac{f_1}{f_1 + a_1} = \frac{1}{\nu}\frac{1}{\max\left[\left|D_2\left(\frac{B_2}{f_2} - \frac{1}{\nu} - 1\right)\right|, d_{pixel}, \frac{1}{\nu_{\text{total}}}\right]} \cdot \frac{f_1}{f_1 - a_1} \tag{13}$$

From Figure 1, the object distance to MLA can be expressed as $a_2 = b_1 - B_1 + B_2$, and could be calculated in terms of physical lens parameters and object distance at main lens side a_1. In a one-dimensional case, when the virtual depth is $v = -1$, it means that every image point formed by the main lens within this plane is re-imaged by at least one microlens. This implies that full coverage of the main lens image can be achieved at $v = -1$. To enable depth estimation, the image points from the main lens must be re-imaged by at least two microlenses. Therefore, the depth estimation becomes feasible when the virtual depth starts from $|\nu| > 2$.

Another impactful parameter of MLA is B. The case $f = B$ is usually the standard plenoptic camera described by Ng [22] and the case $f \neq B$ is called the plenoptic 2.0 or the focused plenoptic camera. The theory developed for the processing of light field data can be easily applied to images generated with the standard plenoptic camera. However, the standard plenoptic camera is limited by low lateral resolution, as the number of resolvable spatial points equals the number of microlenses rather than the number of sensor pixels. In the plenoptic 2.0, by optimizing the image plane position, we can adjust the B/f for different ERR and DOF in the confined area of virtual depth, which is defined as $|\nu| \in [1, D_2/s_0]$ and could be solved from Equation (14). The case is shown in Figure 2 and will be discussed below. Moreover, the hexagonal array arrangement exhibits a coverage factor $\kappa_h \approx 1.15$, which is superior to orthogonal arrays ($\kappa_o \approx 1.41$) and is discussed in ref [13]. This achieves the full coverage at a lower $|\nu|$, reducing the minimum $|\nu|$ required for overlapping projections.

Figure 2 indicates how ERR_2 varies with the virtual depth ν under different microlens focal lengths of 65 μm, 72 μm, 80 μm, 100 μm, 130 μm, 160 μm and 220 μm when $B_2 = 100$ μm, which makes a series of B/f ratios. When $B = f$, the case falls into a standard plenoptic camera in the range of $1 \leq |\nu| \leq D_2/s_0$ and the ERR is constant as d_{pixel}/D_2. This gives the bottom line for all $B \neq f$ cases and all f values departing from B are above this line, which demonstrates the better performance of the focused light field camera compared with the standard light field camera but will require more sophisticated optical engineering efforts.

$$ERR_2 = \frac{1}{\nu}\frac{d_{pixel}}{max\left[|D_2/\nu|, d_{pixel}, \frac{1}{\nu_{\text{total}}}\right]} = \min\left[\frac{d_{pixel}}{D_2}, \frac{1}{|\nu|}, \frac{d_{pixel} \cdot \nu_{\text{total}}}{\nu}\right] \tag{14}$$

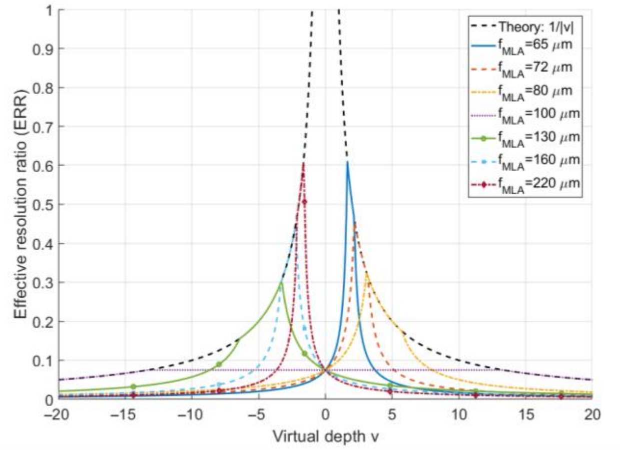

Figure 2. The relationship of ERR2 with virtual depth ν for various focal lengths.

In general, when the large focal length f is closer to the image distance B, DOF increases, and this comes at the cost of reduced *ERR*. Conversely, using a shorter focal length improves the lateral resolution and reduces DOF, which means that adjusting the B/f ratio allows for flexible control of the ERR and DOF shown in Figure 3. The ideal ERR = $\frac{1}{\nu}$ curve in Figure 2 shows a limitation over all possible ERR curves with different *B*/*f* ratios. The results show that the focused plenoptic camera provides more degrees of freedom for balancing lateral resolution and DOF than a standard light field camera. To effectively establish the depth estimation, the virtual depth should start from $|\nu| > 2$ to receive light from at least two microlens. This also lowers the upper bound of ERR values and makes it hard for a plenoptic camera to obtain high resolution.

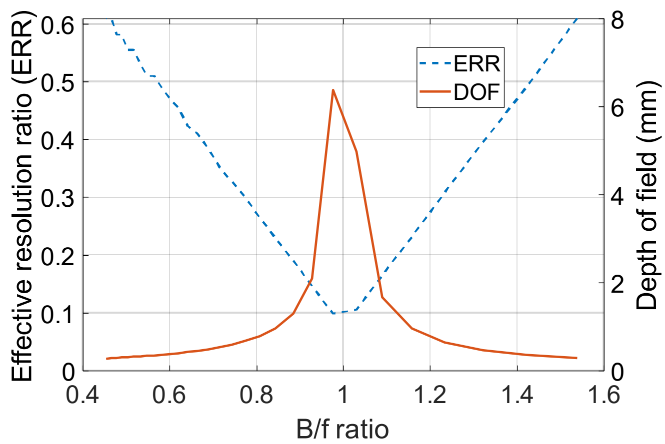

Figure 3. Trade-off between effective resolution (ERR2) and depth of field with respect to *B*/*f* ratio.

As discussed before, the MLA has a much smaller aperture than the main lens, which limits the resolution. Since $\nu = \frac{a_2}{B_2}$ is the ratio of object distance a_2 and the distance from MLA to the image plane B_2, ERR_2 could be connected to the object distance of the main lens a_1 with $b_1 = \frac{f_1 a_1}{a_1 - f_1}$, $b_2 = \frac{f_2 a_2}{a_2 - f_2}$ and $a_2 = b_1 - B_1 + B_2$. The DOF of the main lens at the object side could be obtained by cascading the lens parameters for an even more complicated lens group. This suggests that the DOF on the object side of the main lens can be engineered by tuning the image distance b_1 in conjunction with different focal lengths f_1.

Moreover, Figure 3 also indicates the important fact that the combination of multiple focal-point configurations could possibly cover the broader range of DOF, but the effective pixel number per microlens is reduced as the "trade-off". This is another important way for the light field camera to extend DOF but involves a more complicated design process, which is not discussed in this work.

Since depth sensing is the most common use case for the light field system, achieving the greatest possible depth range becomes critical for 3D reconstruction. In the following section we try to obtain the DOF at the object side of the main lens. Let x denote the axial distance from the intermediate image point formed by the main lens to the MLA plane, which serves as the object distance for the microlens. The sensor's minimum blur circle diameter is defined as $t \leq s$, and the additional defocus from microlens imaging is given by:

$$D_2 \cdot \left[B_2\left(\frac{1}{f_2}+\frac{1}{x}\right)-1\right] \leq d_{pixel} \tag{15}$$

Simplified as:

$$-\frac{d_{pixel}}{D_2} \leq B_2\left(\frac{1}{f_2}+\frac{1}{x}\right)-1 \leq \frac{d_{pixel}}{D_2} \tag{16}$$

For the virtual intermediate image behind the microlens plane ($x < -B_2$), we solve for the far defocus boundary x^- by taking the lower limit of equation:

$$x^- = [\frac{1}{f_2} - \frac{\left(D_2 + d_{pixel}\right)}{D_2 \cdot B_2}]^{-1} \tag{17}$$

The acceptable depth range of the microlens array extends from $a_2 = x^-$ to $a_2 = -B_2$. Under practical conditions where the pixel size is much smaller than the microlens diameter ($d_{pixel} \ll D_2$) and the ERR varies approximately linearly within the DOF range, the relationship between DOF and the effective resolution ratio can be expressed as:

$$DOF = \frac{1}{ERR_2}\frac{2d_{pixel}F/\#}{\left|\frac{B_2}{f_2}-1\right|} \tag{18}$$

Figure 3 illustrates the relationship between ERR and DOF with respect to the B/f ratio. As the B/f ratio approaches 1, the ERR decreases linearly while the DOF exhibits a sharp increase with a distinct peak near the ratio of 1. When the B/f ratio exceeds 1, the ERR begins to rise linearly. This observation aligns with the well-known inverse relationship between resolution and depth of field, where higher depth of field typically corresponds to reduced resolution capability, and vice versa. It should be noted that the ERR and DOF analysis presented above is derived under paraxial and aberration-free assumptions. In practice, when the microlens focal length f is small and the ratio B/f deviates significantly from unity, the large cone angle of the converging beam increases the contribution of Seidel aberrations, in particular spherical aberration and coma. These aberrations enlarge the effective spot size beyond the geometric prediction of Equation (10), which reduces the actual ERR below the values shown in Figure 3. A complete analytical treatment incorporating aberration terms into the ERR formulation is beyond the scope of this work but represents a worthwhile direction for future investigation.

Depth Resolution

Here, $a(h)$ is the distance from the imaging point of the main lens to the MLA; the height from an image point to the center of the microlens is h. The distance from the intermediate image to the MLA is given by:

$$a(h) = \frac{D_2 \cdot B_2}{h} \tag{19}$$

Transforming the depth relationship of the intermediate image plane to the object space of the main lens yields the object depth. Therefore, the object at distance $a_l(h)$ from the main lens with focal length f_1 is:

$$a_l(h) = \left[\frac{1}{f_1} - \frac{1}{l_1 + a(h)}\right]^{-1} \tag{20}$$

The minimum resolvable depth difference $DR = a_l(h) - a_l(x + \Delta h)$ in the object space is defined as:

$$DR = 2\frac{{f_1}^2}{D_2 \cdot B_2} \cdot d_{pixel} \tag{21}$$

Therefore, the depth resolution is proportional to the square of the main lens focal length and depth range and inversely proportional to MLA lens diameter D_2 and distance B_2 between the MLA and the sensor plane, which is in line with the resolution and DOF discussion above.

Relationship Between MLA Microlens Radius and Main Lens

The physical properties of the MLA (period p or radius r) and the total number of array elements N are the critical trade-offs affecting imaging performance. Based on the diffraction-based modulation transfer function (MTF) theory, the MLA's resolution can be expressed as:

$$\nu_{\text{MLA}} = \min\left(\frac{\text{M}}{2 \cdot d_{pixel}}, \frac{1}{1.22 \cdot \lambda \cdot F/\#}\right) \tag{22}$$

Equation (22) indicates that increasing the microlens radius D_2 enhances the high-frequency cutoff capability of the MTF response, thereby improving the lateral resolution. However, large aperture reduces the overall sampling rate of the MLA and causes serious degradation of the total optical resolution. Additionally, the number of sensor pixels k covered by a single microlens scales with r^2 and $k = \pi\, r^2/\, s^2$, where s is the pixel size. While a larger r improves angular resolution by providing denser ray sampling for multi-view reconstruction, the total number of microlenses N decreases inversely with r^2 for a fixed sensor area, reducing the lateral sampling density of macro-pixels and thus compromising the lateral resolution.

Here we define the lateral resolution $R_s = \sqrt{N} \cdot \nu_{MLA}$ and the angular resolution $R_a = k \cdot \Delta\theta$ (where $\Delta\theta$ is field of view per macro-pixel). The relationship becomes:

$$R_s \propto \frac{\nu_{\text{MLA}}}{r}, R_a \propto r^2 \tag{23}$$

This confirms that lateral resolution R_s is inversely proportional to r, while angular resolution R_a is proportional to r^2, which aligns with empirical observations. Consequently, the MLA design strikes a balance between microlens radius and array element density to optimize lateral resolution and angular resolution for seed phenotyping applications.

2.1.4. MLA Performance Analysis

The MLA is a pivotal component in light field systems, primarily defining critical parameters for 3D reconstruction, including spatial, angular, and depth resolutions. While the geometric models above define the macroscopic system limits, the microscopic performance is dictated by wave optics; thus, we employ Finite-Difference Time-Domain (FDTD) simulations to rigorously analyze the focusing efficiency and crosstalk of the microlens array, ensuring the chosen parameters minimize artifacts at the sensor plane.

Figure 4 shows the intensity profile with a plane source and a single-channel profile from a Gaussian source. The MLA is simulated with a microlens configuration of radius R = 10 μm and focal length F = 50 μm, arranged in a 3 × 3 array. The intensity distribution in the x–y plane demonstrates tight spatial confinement. To quantitatively assess optical crosstalk, we calculated the CMOS sensor response under single Gaussian source illumination, yielding a crosstalk ratio of approximately 0.30% between central area and adjacent area. Such a low level of signal leakage confirms that inter-focal interference is effectively suppressed. However, this high degree of optical confinement corresponds to a limited DOF of 20 μm, which may restrict the capture of depth features in high-relief samples.

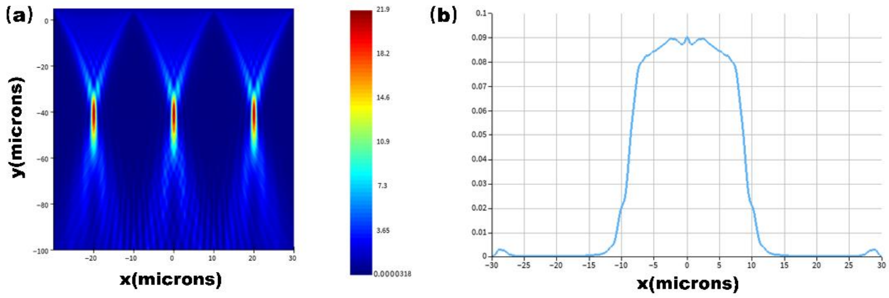

Figure 4. (**a**) Intensity profile along the x–y plane of the MLA with radius R = 10 μm and focal length F = 50 μm from a plane source, (**b**) beam profile from a single Gaussian source.

Figure 5 presents the results for the microlens configuration with R = 10 μm and F = 100 μm. The simulation demonstrates a crosstalk ratio of 0.65%, which remains within a robust range for signal separation. Notably, the focusing efficiency remains high, with the peak pixel response being comparable to the F = 50 μm case, indicating that optical energy is well-preserved even with a longer focal length. Crucially, this configuration provides a beneficial trade-off by extending the DOF from 20 μm to approximately 50 μm without compromising signal strength.

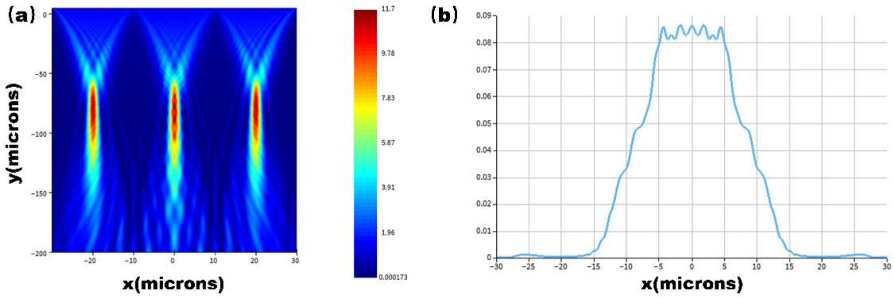

Figure 5. (**a**) Intensity profile along the x–y plane of the MLA with radius (R) = 10 μm and focal length F = 100 μm from a plane source, (**b**) beam profile from a single Gaussian source.

The configuration of R = 10 μm and F = 200 μm is analyzed in Figure 6. In this scenario, the crosstalk ratio rises to approximately 1.57% due to increased beam expansion from adjacent microlenses. Owing to the poor focal spot confinement, the focal efficiency almost drops to 45%, and the focal spot can no longer maintain a good shape.

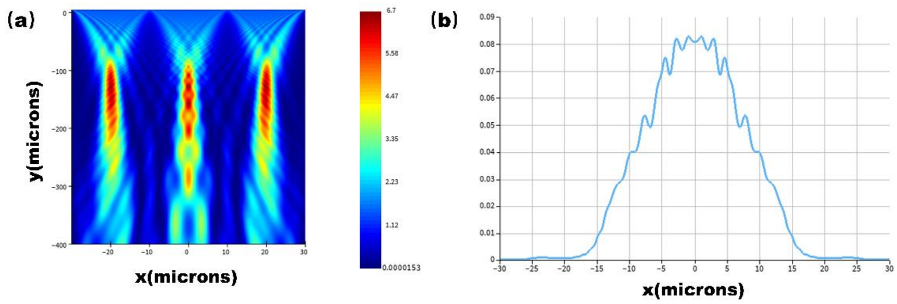

Figure 6. (**a**) Intensity profile along the x–y plane of the MLA with radius R = 10 μm and focal length F = 200 μm from a plane source; (**b**) beam profile from a single Gaussian source.

The simulation results indicate an inverse relationship between focal efficiency and DOF with respect to focal length. Given that the target application environment typically features adequate illumination, DOF performance is prioritized over focal efficiency in this context. Consequently, the MLA configuration with an F-number of 100 μm is selected as the optimal design. Figure 7 shows the MTF calculated from the PSF profile, which is

https://doi.org/10.3390/photonics13040385

converted to the spatial frequency domain and exhibits a typical low-pass filter characteristic. Quantitatively, the system achieves a spatial frequency of 10.2 lp/mm at the MTF50 threshold, which is close to the optimized design parameters discussed in the following section.

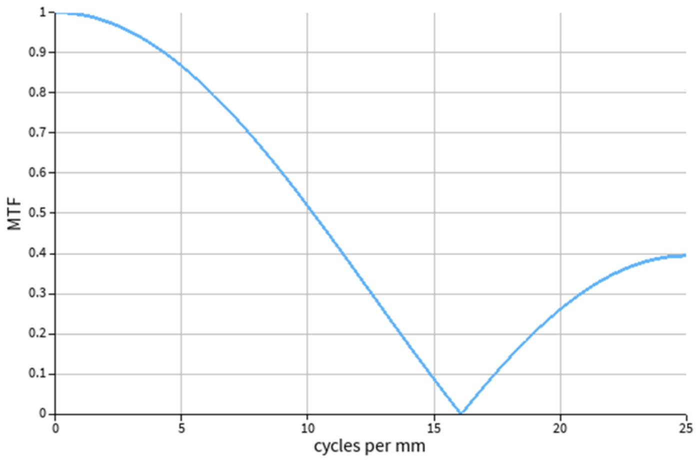

Figure 7. Calculated modulation transfer function of the MLA system. (Focal length 100 μm, unit diameter 20 μm).

2.2. Example of Light Field Camera Design Optimization

This section applies the key findings from the preceding discussion to provide a practical design example of a seed phenotyping 3D reconstruction system, demonstrating the light field system workflow illustrated in Figure 8. Similar to conventional optical design approaches, the process begins by establishing initial parameters, followed by iterative optimization to meet system specifications. The flowchart outlining the complete light field system design and optimization procedure is presented below.

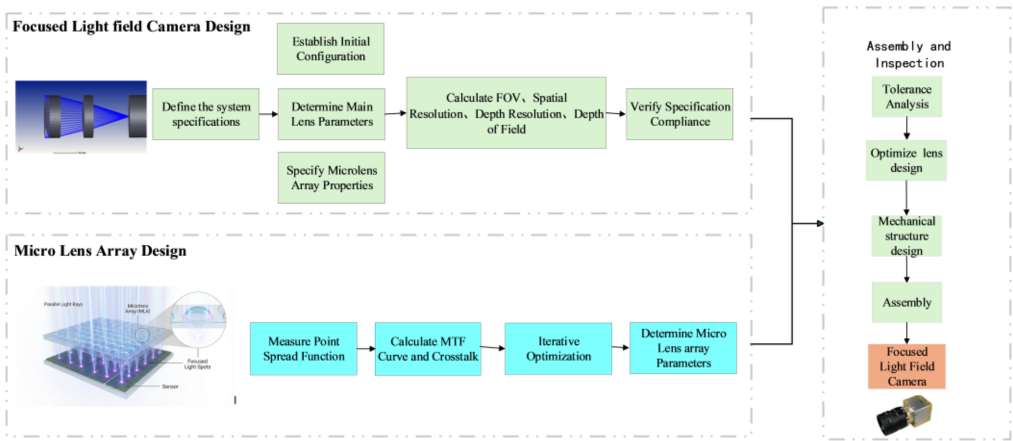

Figure 8. Design flowchart of the focused light field camera.

First, the key specifications of the target scene, such as object distance, FOV and DOF, are defined. Next, MLA parameters are determined, including its arrangement (hexagonal), focal length, diameter, and virtual depth. Subsequently, the focal length and diameter of the main lens are selected. Finally, all design parameters are calculated and verified. If any requirement is not satisfied, the design undergoes further optimization.

In 3D reconstruction applications for seeds, depth estimation often fails due to poor image quality across different areas in the sub-aperture images. To address these issues, we follow the FOV priority procedure mentioned at the beginning, focusing on pushing the lateral resolution to the limits while keeping the desired FOV obtained from Equation (3). In the system diagram shown in Figure 9, we define the first lens as the main lens, which is used to improve optical etendue with a large aperture, delivers collimated light and gives extra tunability for the tunable lens. The tunable lens is used to adjust the effective focal length for different requirements of FOV and resolution. The effective focal length could be obtained by the equation:

$$\frac{1}{f_{eff}} = \frac{f_1 \times f_2}{f_1 + f_3 - d} \tag{24}$$

where f_1 and f_3 are the focal lengths of the main lens and the tunable lens and d is the distance between them, which could be tuned for different f_{eff}. Therefore, by adjusting the lens separation d, the effective focal length f_{eff} can be tuned to achieve the desired FOV. Resolution optimization is subsequently addressed within the constraints governed by Equation (13). Furthermore, the MLA configuration critically influences resolution and DOF performance, requiring careful implementation, as analyzed in Section 2.1.3.

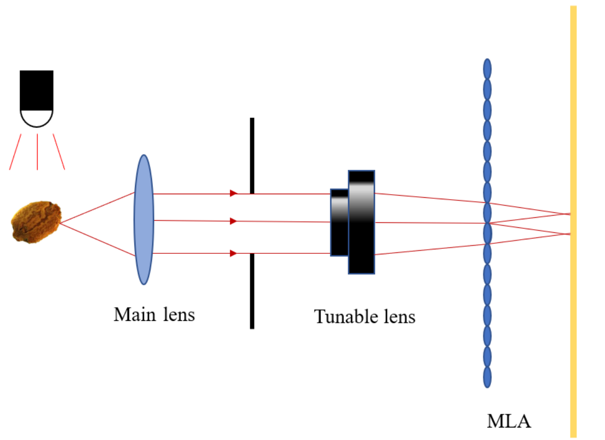

Figure 9. Schematic diagram of the tunable light field system.

In particular, we set the object distance a_1 = 76 mm. As shown in Figure 9, the lens parameters used in the optical system design are listed below:

- Main lens: Focal length $f_1 = 25.8$ mm, $F/\#_1 = 2$.
- Tunable lens: Focal length $f_3 = 76.2$ mm, $F/\#_2 = 2$, distance to main lens $d = 20$ mm.
- MLA: Focal length $f_2 = 100$ µm, diameter $D_2 = 20$ µm, pitch $p = 20$ µm, pixel size $d_{pixel} = 1.5$ µm. Distance from main lens to MLA: 35 mm, distance from MLA to sensor: 80 µm.
- Wavelength: $\lambda = 550$ nm.
- Sensor size: 10.82 mm × 7.52 mm.

We compare the performance of a single-lens system with a focal length of 25.8 mm to that of our tunable-lens design, where the focal length used in the tunable optical system is 23.98 mm. As shown in Table 1, better resolution and FOV performance can be achieved after the optimization. However, DOF is obviously reduced due to the trade-off between resolution and DOF. In MLA design, Lumerical FDTD simulations were conducted to investigate the microlens's focusing efficiency.

As shown in Table 1, the analysis of the optical system indicates that integrating a tunable lens with a microlens array improves overall performance. Reducing the main lens focal length to 23.98 mm successfully expands FOV from 23.68° to 25.42°, meeting the need for wider imaging coverage. This adjustment also enhances the theoretical spatial resolution, which increases from 1.11 lp/mm to 9.41 lp/mm. Moreover, the tunable lens enables flexible DOF adjustment to accommodate varying imaging depths. Further supported by FDTD simulations, a microlens array with a radius of 10 μm and a focal length of 100 μm achieves an effective balance among resolution, spatial confinement, and DOF performance. Quantitatively, the simulated system attains a spatial frequency of 10.2 lp/mm at the MTF50 threshold. The result is in line with the calculated value of 9.41 lp/mm, validating both the accuracy of the theoretical model and the resolving capability of the optimized design.

Table 1. Initial design vs. optimized parameters with improved design.

Parameter	Single-Lens Design	Tunable-Lens Design
Focal length	25.8 mm	23.98 mm
FOV	H: 23.68°, V: 16.58°	H: 25.42°, V: 17.82°
Lateral resolution	1.11 lp/mm	9.41 lp/mm
Depth of field (DOF)	4.54 mm	0.5142 mm

The DOF reduction in the optimized design is a direct consequence of prioritizing lateral resolution, and this priority can be justified by examining the factors that limit depth estimation precision. Using the system parameters listed above, the geometric blur spot of the microlens S_2 is approximately 4.39 μm, the diffraction-limited spot size S_0 is 3.36 μm, and the pixel pitch d_{pixel} is 1.5 μm. Since S_2 exceeds d_{pixel} by a factor of approximately three, the bottleneck in EPI-based depth estimation is not the sensor sampling but the optical blur of the sub-aperture images. Under this condition, improving the lateral resolution by reducing S_2 toward d_{pixel} directly improves the disparity estimation precision in the EPI and brings the actual depth resolution closer to the theoretical limit given by Equation (21). The cGAN-based deblurring pipeline described in Section 2.3.2 further reduces the effective blur spot size in the sub-aperture images, thereby enhancing both lateral and depth resolution simultaneously. For samples whose surface relief exceeds the 0.51 mm DOF, the computational refocusing capability of the light field system can generate a focal stack from a single acquisition, from which an all-in-focus image is synthesized to recover surface details across the full depth range without additional mechanical scanning.

2.3. Image Processing Pipeline Development

In this work, we utilize a light field imaging system wherein each macro-pixel comprises a 15 × 15-pixel array. By resampling the raw light field data, a sequence of 15 × 15 sub-aperture images is generated. To address aberrations in the sub-images, we first correct most of the aberrations using Zhang's method [23] based on the raw light field image line feature detection (see the Supplementary Materials for details). To address the remaining blurring effects induced by spherical aberration and field curvature, we propose a dedicated correction method. Furthermore, the depth estimation framework is developed, and a 3D point cloud optimization strategy is applied. These approaches effectively mitigate geometric distortions, especially in low-texture and high-curvature regions on the seed surface, thereby achieving the micrometer-level precision required by the system design.

2.3.1. Disparity-to-Depth Fitting

The calibration procedure consists of camera-parameter calibration and disparity-to-depth fitting. The disparity, which is computed from the sub-aperture images of the light field camera, exhibits a linear relationship with depth z. In lenslet-based light field systems, the disparity between corresponding points in sub-aperture views is inversely

https://doi.org/10.3390/photonics13040385

proportional to the object depth due to the angular displacement of light rays captured by different microlenses [24]. This yields the depth z as:

$$\frac{1}{z}=\frac{1}{a_1}+\frac{B_2D_2}{d_{pixel}b_1}\cdot(-\Delta x) \tag{25}$$

To quantify measurement precision, we analyze residuals in disparity space, rather than depth space. Because the disparity–depth mapping is nonlinear, a constant disparity error corresponds to depth errors that vary across the working range; using depth residuals would therefore obscure the true underlying uncertainty. After fitting the model $z = f(d,\theta)$ using the Levenberg–Marquardt algorithm, we compute the disparity residuals at each measurement point:

$$r_i = d_i^{measured} - d_i^{predicted} \tag{26}$$

The standard deviation of the disparity residuals is computed as:

$$\mathrm{RSD}_d = \sqrt{\frac{\sum_{i=1}^{n} r_i^2}{n-p}} \tag{27}$$

where n is the number of measurement points and p is the number of fitted parameters. The value $2 \times \mathrm{RSD}_d$ is adopted as the disparity-space uncertainty threshold, corresponding to the 95% confidence interval.

To evaluate depth resolution across the operating range, we select three representative depth positions: the near point z_1, the far point z_n, and the midpoint $z_{mid} = (z_1 + z_n)/2$. At each position, Δz_{res} is computed as:

$$\Delta z_{res} = |f(d - 2\cdot \mathrm{RSD}_d, \theta) - f(d,\theta)| \tag{28}$$

where d is the disparity corresponding to the given depth. This calculation reveals how depth resolution evolves with working distance, reflecting the inherent nonlinearity of the disparity–depth relationship.

Calibration experiments were conducted under two optical architectures: a single-lens configuration and a tunable-lens configuration. Disparity measurements were extracted along four orientations (horizontal, vertical, and two diagonal directions), and linear regression was performed to establish the depth calibration model.

2.3.2. Image Deblurring

The nature of the angular sensing in light field technology enhances information dimension, but a wide FOV can also introduce off-axis aberrations, such as spherical aberration, coma, and distortion [19,25]. Although the optical aberration has been primarily reduced in optical design, there are still spherical and coma aberrations from the main lens, combined with the off-axis vignetting effects of the microlens, resulting in laterally varying hybrid blur kernels in raw light field data. For example, a symmetric Gaussian-like blur kernel exists in the central area of the FOV due to dominant spherical aberration. The aberrations existing in light field cameras can be classified into two primary categories (see Supplementary Materials for details).

To address this challenge, we propose a cGAN framework for hybrid aberration correction. This approach leverages laterally adaptive deblurring to handle non-uniform blur patterns across the FOV while preserving light field angular–lateral consistency.

$$I_B = k(M) * I_S + N \tag{29}$$

In the equation above, I_B represents the blurred image, $k(M)$ denotes the unknown system blur kernel, which convolves with the true image I_S, and N is the noise introduced during imaging. The deblurring problem can be categorized into blind deblurring (unknown blur kernel) and non-blind deblurring (known blur kernel). In light field imaging systems, accurately obtaining the system's PSF (point spread function) is not experimentally feasible. To address this, we introduce a conditional adversarial generative network that leverages an attention mechanism to separate target textures from aberration noise in the frequency domain. By designing a generator to synthesize blurred images containing comprehensive system aberrations, the adversarial training with a discriminator evaluates and refines the deblurred output, enhancing resolution performance for subsequent depth image computation and 3D reconstruction rendering.

The conditional generative adversarial network does not need to explicitly estimate the blur kernel for the above complex aberrations and can directly learn the mapping from blurry images to clear images. The structure is shown in Figure 10, and the generator uses the publicly available pretrained model from DeblurGAN [26]. Here we combine Wasserstein GAN (WGAN-GP) with perceptual loss to balance global structure recovery and detail generation. A lightweight generator architecture based on residual blocks (ResBlock) and skip connections (skip connection) is adopted with few system parameters and fast inference speed. Among them, the generator consists of 2× stride convolution downsampling modules, 9× ResBlock residual blocks and 2× transposed convolution upsampling modules. The skip connection is:

$$I_S = I_B - ResOut \tag{30}$$

The generator is not a purely data-driven black box. The global skip connection in Equation (30) formulates the task as residual learning, where the network estimates only the degradation component caused by optical aberrations and subtracts it from the blurred input. This is consistent with the classical restoration model in Equation (29), in which the goal is to remove the convolution effect of the blur kernel and noise. The network thus learns a general degradation prior that approximates the inverse mapping of the blur process without requiring explicit PSF estimation. In the WGAN-GP framework, the discriminator functions as a critic that outputs an unbounded scalar score rather than a classification probability. The discriminator in this work adopts the PatchGAN architecture with five convolutional layers using 4 × 4 kernels. The first four layers apply a stride of 2 and increase the feature map channels from 64 to 512 progressively. The final layer outputs a single-channel map, and all layers except the last are followed by InstanceNorm and LeakyReLU with a slope of 0.2. Each spatial position in the output corresponds to a 70 × 70 receptive field on the input image. The averaged output serves as the score used in the WGAN-GP loss defined in Equation (32). We design the loss function as:

$$L_{\text{Total}} = \lambda_{\text{GAN}} L_{\text{GAN}} + \lambda_{\text{Perceptual}} L_{\text{Perceptual}} \tag{31}$$

Here, deblurring is perfromed by the trained CNN G_{θ_G}, the critic network is D_{θ_G}, and E_{I_B} denotes the expectation over the distribution of blurred input images I_B. Adversarial loss (WGAN-GP) is given by:

$$L_{\text{GAN}} = E_{I_B}\left[D_{\theta_G}\left(G_{\theta_G}(I_B)\right)\right] \tag{32}$$

Perceptual loss:

$$L_{\text{Perceptual}} = \sum_{i,j} \left\|\phi_{i,j}(I_S) - \phi_{i,j}\left(G(I_B)\right)\right\|_2^2 \tag{33}$$

$\phi_{i,j}$ is the feature map of the j-th convolutional layer of the i-th layer of the pretrained network. The discriminator optimization objective is designed to maximize the Wasserstein distance between the real clear image and the generated image.

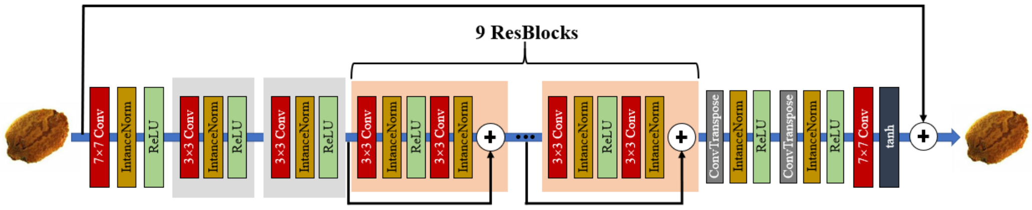

Figure 10. The architecture of the ResNet-based generator for image deblurring. The "+" denotes element-wise addition, implementing the global skip connection for residual learning.

2.3.3. Depth Image and 3D Rendering

The light field images capture rich lateral and angular information, which can be resampled to obtain a 15 × 15 sub-image sequence, known as the EPI. Depth estimation is performed by analyzing the pixel displacement characteristics across different viewpoints to reconstruct the scene, as shown in Figure 11 below. The method for constructing the EPI is as follows: along a fixed spatial coordinate (horizontal or vertical), the pixels from the same row or column of the sub-aperture image sequence are stacked in order of viewpoints, forming a slice image that contains the point trajectory of the scene. In the EPI, the light rays from the same point in the spatial-angular joint domain appear as straight lines with specific slopes. When there are depth differences in the scene, the edges of the object in the EPI will show a linear tilting feature. Depth information can be calculated by combining the baseline and slope parameters with the depth calculation formula. The relationship between the slope of the EPI texture line, the difference between images, and the depth estimation can be treated as equivalent. As shown in the simplified EPI diagram, the light rays emitted from a point in space will project to different positions in adjacent sub-images. The baseline length of the sub-image array, represented by h, is essential for depth estimation. It is clear that the slope of the line in the EPI is related to the differences between points.

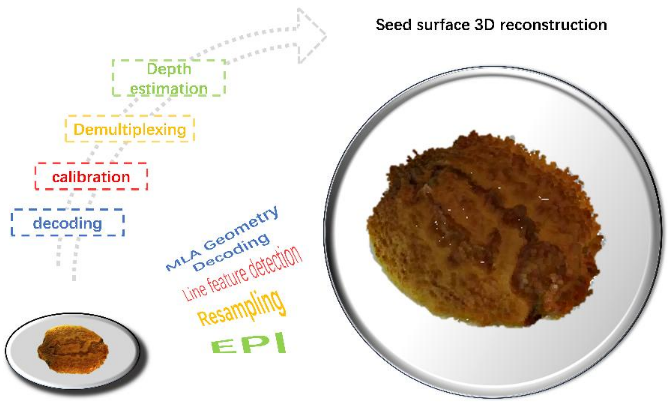

Figure 11. Pipeline of 3D reconstruction of seed surface with EPI and full color 3D rendering.

When integrating the depth map obtained from the EPI estimation with the two-dimensional RGB image, a pixel-level geometric-color alignment model must be established. First, the 2D pixel coordinates (u, v) and their corresponding depth Z_{EPI} are mapped to the 3D space through the pinhole camera projection relationship.

$$\begin{cases} X = \dfrac{(u - c_x) \cdot Z}{f} \\ Y = \dfrac{(v - c_y) \cdot Z}{f} \\ Z = Z_{\mathrm{EPI}} \end{cases} \tag{34}$$

Here f is the camera's focal length and (c_x, c_y) are the coordinates of the principal point. This step transforms each RGB pixel into a 3D point (X, Y, Z), preserving the color attributes (R, G, B), thus forming a colored point cloud. The algorithmic processing pipeline from raw light field image through deblurring, color correction, resampling, depth estimation, and 3D reconstruction is illustrated in Figure 11.

3. Experiments

We constructed a light field camera verification platform for seed surface imaging, as shown in Figure 12. The light field imaging system consists of a main lens with focal length f = 25.8 mm, F-number, a tunable lens with focal length $f = 9{\sim}76.8$ mm with $F/\# = 2$, a microlens array (focal length: 100 μm, diameter: 20 μm, baseline: 20 μm), and a CMOS sensor (sensor size: 10.82 mm × 7.52 mm, pixel size: 1.5 μm). The standard USAF 1951 resolution test board was used to estimate the resolution, and the captured results are shown in Figure 13. Figure 13a shows that the image resolution and MTF results are significantly improved to 3.25 lp/mm at MTF50 by the enhanced optical hardware. Figure 13b shows that the images in (a) and the MTF results are improved to 6.2 lp/mm by the deblurring algorithm. Discrepancies between theoretical predictions and experimental results arise from multiple sources. On the fabrication side, MLA surface profile errors and assembly tolerances between the MLA and the sensor introduce deviations from the ideal optical geometry. On the modeling side, the theoretical analysis relies on thin-lens approximations and monochromatic illumination, which do not fully account for higher-order aberrations, chromatic effects, and the actual wavefront distortion of the main lens and tunable lens assembly. In addition, sensor-level factors such as pixel crosstalk and electronic noise further degrade the measured resolution relative to the theoretical prediction. A comprehensive error budget quantifying individual contributions is planned for future work.

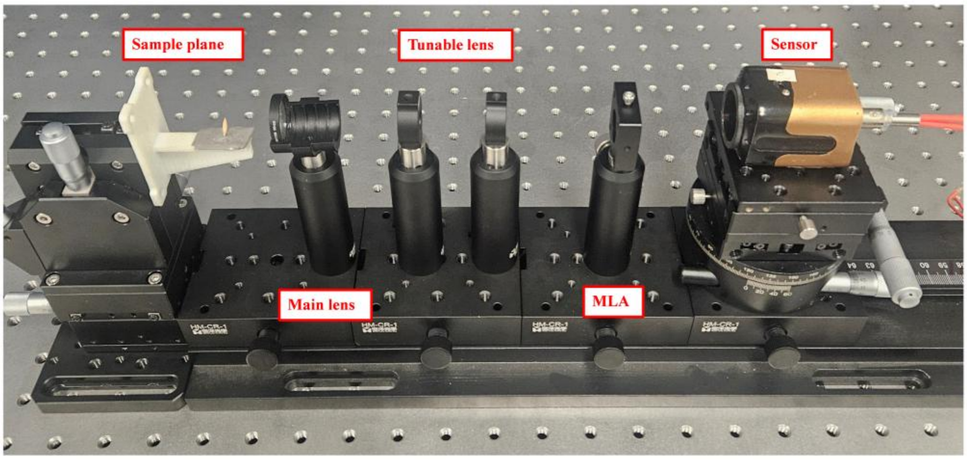

Figure 12. Picture of the experimental setup. The system consists of a sample holder at the sample plane, a main lens, a tunable lens, an MLA and a CMOS sensor.

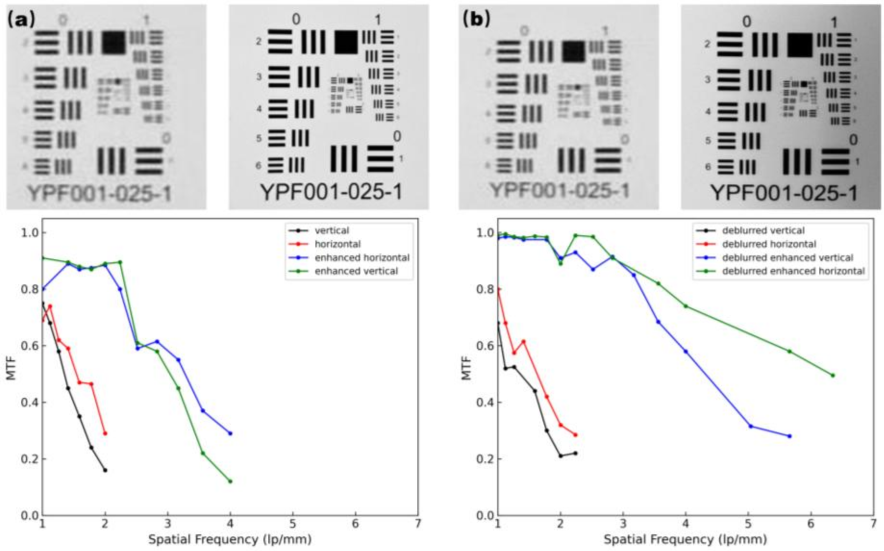

Figure 13. USAF 1951 sub-images and MTF curves: (**a**) the image captured by the original optical system and the image captured by the modified system. (**b**) The deblurred image with the original system and after the modified system.

To evaluate the disparity-depth relationship, calibration experiments were conducted under two optical configurations: the single-lens design and the tunable-lens design. Linear fitting was performed for disparity measurements along four directional axes (horizontal, vertical, and two diagonal directions) to establish the depth calibration model. The fitting results are shown in Figure 14:

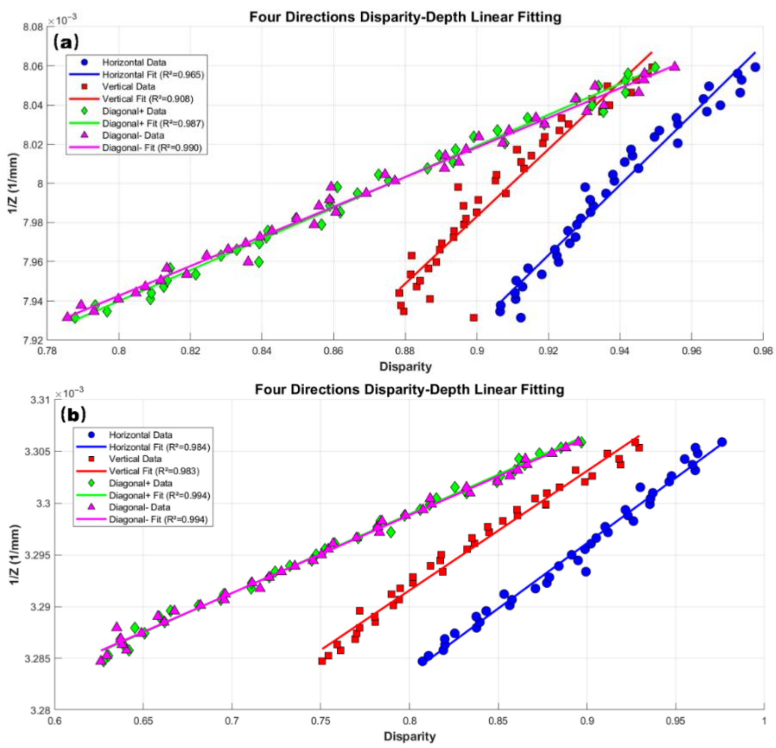

Figure 14. Four-directional disparity and depth linear fitting. (**a**) Single-lens design. (**b**) Tunable-lens design.

The calibration results demonstrate that the tunable-lens design markedly outperforms the single-lens configuration. Compared with the single-lens system, the tunable-lens setup exhibits superior linearity, with R^2 values consistently exceeding 0.98, and greatly enhanced precision, as the RSD_d is reduced by approximately fivefold. According to the Rayleigh criterion, where depth resolution is defined as twice the residual standard deviation, the tunable-lens system achieves an average depth resolution of 65.96 μm over the 2 mm scanning range, representing a 4.99-fold improvement over the single-lens configuration (328.86 μm). Moreover, the tunable-lens design maintains excellent spatial stability; depth resolution varies by less than 0.4 μm throughout the scanning range, whereas the single-lens design shows a progressive degradation. These improvements indicate that the tunable-lens architecture provides significantly better disparity–depth calibration consistency and accuracy, which is essential for high-precision depth estimation in light field imaging applications that require sub-100 μm depth discrimination.

The residual aberrations in sub-aperture images of bitter gourd seeds are corrected by the cGAN algorithm. Figure 15 shows that the deblurred images achieved a PSNR of 34.63 dB and an SSIM of 0.9570 after deblurring, validating the feasibility of the cGAN approach to improve the sub-image quality.

Figure 15. Original bitter melon seed image and deblurred result and residual map of the conditional generative adversarial network.

The cGAN-based deblurring algorithm shows a 90.76% (MTF50) improvement in the high-frequency signal region. With the combination of the enhanced optical system and the computational algorithm, clear sub-aperture images of bitter gourd and rice seeds are successfully obtained, exhibiting more prominent surface textures and structural features. As shown in Figure 16, the 3D reconstruction results reveal that a dense point cloud distribution of up to about 15,000 points per square centimeter is feasible, and the software in this work can offer a reliable and information-rich dataset for seed phenotyping research.

Figure 16. Bitter melon seed and rice seeds 3D rendering.

Table 2 provides a comprehensive comparison of six depth-sensing technologies across multiple performance dimensions. The comparison evaluates key metrics, including fundamental principles, lateral resolution, depth accuracy, measurement range, acquisition speed, surface dependency, and primary advantages and limitations for phenotyping applications. As shown in Table 2, the light field system achieves a lateral resolution of approximately 80 μm and a depth accuracy of 66 μm, while maintaining robustness to challenging surface conditions such as specularity and low texture—properties that are common in seed and leaf phenotyping. Unlike laser scanning and structured light, it is less sensitive to surface optical properties and can provide significantly better resolution. Relative to other stereo vision techniques, it does not require strong texture cues. Although slower in acquisition (86 s per sample), the single-shot capability and upcoming optimization to 1–2 s position it as a promising, motion-robust solution for high-resolution 3D phenotyping, despite being computationally intensive. It should also be noted that no publicly available light field dataset for seed phenotyping currently exists, which precludes direct benchmarking against end-to-end deep learning-based 3D reconstruction methods. The primary goal of this work is to perform 3D reconstruction based on geometric optics and physical modeling of the light field system rather than data-driven depth prediction, which we consider more suitable for application scenarios where domain-specific training data are unavailable. As summarized in Table 2, our light field system offers a unique combination of high lateral resolution and depth accuracy that is robust to surface texture, outperforming traditional methods in key phenotyping scenarios.

Table 2. Comparison of depth-sensing technologies for phenotyping applications.

Feature/Technology	**Laser Scanning**	**Structured Light**	**Flash LiDAR**	**Stereo Vision**	**This Work (Light Field)**
Principle	Point-by-point/time-of-flight	Pattern deformation/phase shifting	Area illumination/time-of-flight	Epipolar geometry/disparity	Multi-view synthesis/EPI analysis
Lateral Resolution	Very High (μm)	High	Low	Medium (texture-dependent)	Approximately 80 μm
Depth Accuracy	Very High (μm level)	High (10–100 μm)	Medium (mm–cm level)	Medium-High (50–500 μm)	High (66 μm), Measured
Depth Range	Meters	0.1–10 m	0.1–100+ m	0.1–10 m	Approximately 2 mm
Measurement Speed (per sample)	Slow (min–hour)	Fast (ms–s)	Very Fast (μs–ms)	Fast (ms–s)	Current: 86 s, Target: 1–2 s
Surface Dependency (Leaves/Seed Coat)	Fails on wet/specular leaves	Fails on wet/specular leaves	Works on most surfaces	Fails on textureless leaves	Robust to specularity & texture
Key Advantage for Phenotyping	Ultimate accuracy for static samples	Good speed-accuracy balance	scan in motion	Low cost, intuitive	Single-shot, high-res, and motion-robust
Main Limitation for Phenotyping	Slow, sensitive to motion & specularity	Sensitive to ambient light and surface	Low resolution limits detail	Easy to fail on smooth surfaces	Computationally intensive (addressable via our optimization roadmap)
Representative Algorithms	ICP registration [27], Hierarchical point feature learning [28]	Phase-shifting profilometry [29], CNN-based fringe-to-phase recovery [30]	ToF depth correction [31], Point set feature network [32]	Semi-global matching [33], Iterative geometry encoding volume [34]	EPI-based depth estimation [35], EPINET [36]

4. Conclusions

This work has presented a hardware–software co-optimized light field imaging framework designed to address the multi-scale imaging challenges inherent in 3D seed phenotyping. By integrating a dynamically tunable optical front end with a deep learning-enhanced reconstruction pipeline, the proposed system offers an effective solution to the fundamental trade-off between field of view and lateral resolution that has long constrained conventional plenoptic cameras.

The key methodological contribution of this research lies in the systematic integration of optical reconfiguration and computational imaging. At the hardware level, the tunable-focus architecture, combined with a custom-designed microlens array, enables flexible optimization of optical sampling efficiency across different imaging scales. The two-stage correction strategy, which integrates geometric calibration with cGAN-based deblurring, has been experimentally validated to improve the effective resolution (MTF50) from 1.52 lp/mm to 6.2 lp/mm, a 307% enhancement. These improvements facilitate high-fidelity EPI depth estimation and the generation of dense point clouds, even under the optical constraints imposed by a wide-field configuration. Quantitative depth calibration further demonstrates a depth resolution of approximately 66 μm over a 2 mm range, representing a fivefold improvement over the single-lens baseline.

Beyond these technical metrics, this study establishes a scalable method for the digitization of germplasm resources. By reconstructing subpixel-level surface details and true-color textures across seed varieties of substantially different sizes, from rice grains to bitter melon seeds, the system supports the extraction of precise morphological traits essential for functional genomics and intelligent breeding. This capability accelerates the transition from manual observation to high-throughput digital phenotyping, providing reliable data infrastructure for modern agricultural analysis.

Looking forward, several directions merit further investigation. First, system miniaturization through the integration of silicon-based optical devices [37,38] or emerging metalens arrays could significantly reduce footprint while maintaining performance, enabling portable and field-deployable phenotyping platforms. Second, the current processing time of 86 s per sample is primarily limited by CPU-based computation. The two most time-consuming stages, EPI-based depth estimation and point cloud generation, are both amenable to GPU acceleration. The EPI slope computation is performed independently at each spatial position across multiple angular directions, making it naturally parallelizable and well suited for GPU-based batch processing. The point cloud generation and registration stages involve per-pixel back-projection and nearest-neighbor operations, which can be accelerated through CUDA-based parallel computation. With these optimizations, a target processing time of 1 to 2 s per sample is feasible, meeting the throughput demands of large-scale breeding programs. Finally, although primarily validated for agricultural applications, the reconfigurable computational imaging architecture demonstrated here holds promise for broader domains requiring adaptive depth sensing, including robotic vision [39], biomedical endoscopy, and industrial precision inspection [40].

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/photonics13040385/s1, Figure S1: Calibration results of line feature regions in the original light field image; Table S1: Main lens aberrations.

Author Contributions: J.Y.: Methodology, formal analysis, software, writing—original draft, data curation, and visualization. Q.Z.: Data curation, supervision, and writing—review and editing. S.L.: Data curation and validation. M.X.: Formal analysis. J.G.: Investigation. Y.Y.: Data curation and resources. C.L.: Methodology and software. X.T.: Resources and project administration. S.W.:

Resources and validation. Q.H.: Formal analysis. F.G.: Conceptualization and methodology. Q.L.: Conceptualization and funding acquisition. M.Z.: Writing—review and editing, project administration, and funding acquisition. Q.S.: Conceptualization, methodology, supervision, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: Yuelushan Laboratory Breeding Program, YLS-2025-ZY03022; Changsha Municipal Key Special Projects (kq2404013). Hunan Agricultural Science and Technology Innovation Fund Project (No. 2025CX35).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: Author Chao Li was employed by the company Qualcomm Inc. Author Qi Song was employed by the Soleilware Photonics LLC. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Fiorani, F.; Schurr, U. Future Scenarios for Plant Phenotyping. *Annu. Rev. Plant Biol.* **2013**, *64*, 267–291. https://doi.org/10.1146/annurev-arplant-050312-120137.
2. Li, L.; Zhang, Q.; Huang, D. A Review of Imaging Techniques for Plant Phenotyping. *Sensors* **2014**, *14*, 20078–20111. https://doi.org/10.3390/s141120078.
3. Liu, Y.; Pu, H.; Sun, D.-W. Hyperspectral Imaging Technique for Evaluating Food Quality and Safety during Various Processes: A Review of Recent Applications. *Trends Food Sci. Technol.* **2017**, *69*, 25–35.
4. Yang, J.; Zhao, Q.; Liu, S.; Guo, J.; Guan, F.; Wang, S.; Hu, Q.; Liu, Q.; Song, Q.; Zhu, M.; et al. Towards Next-Generation Smart Seed Phenomics: A Review and Roadmap for Metasurface-Based Hyperspectral Imaging and a Light-Field Platform for 3D Reconstruction. *Photonics* **2026**, *13*, 61. https://doi.org/10.3390/photonics13010061.
5. Tanabata, T.; Shibaya, T.; Hori, K.; Ebana, K.; Yano, M. SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis. *Plant Physiol.* **2012**, *160*, 1871–1880. https://doi.org/10.1104/pp.112.205120.
6. Roussel, J.; Geiger, F.; Fischbach, A.; Jahnke, S.; Scharr, H. 3D Surface Reconstruction of Plant Seeds by Volume Carving: Performance and Accuracies. *Front. Plant Sci.* **2016**, *7, 745*. https://doi.org/10.3389/fpls.2016.00745.
7. Toda, Y.; Okura, F.; Ito, J.; Okada, S.; Kinoshita, T.; Tsuji, H.; Saisho, D. Training Instance Segmentation Neural Network with Synthetic Datasets for Crop Seed Phenotyping. *Commun. Biol.* **2020**, *3*, 173. https://doi.org/10.1038/s42003-020-0905-5.
8. Preibisch, S.; Saalfeld, S.; Tomancak, P. Globally Optimal Stitching of Tiled 3D Microscopic Image Acquisitions. *Bioinformatics* **2009**, *25*, 1463–1465. https://doi.org/10.1093/bioinformatics/btp184.
9. Wu, G.; Masia, B.; Jarabo, A.; Zhang, Y.; Wang, L.; Dai, Q.; Chai, T.; Liu, Y. Light Field Image Processing: An Overview. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 926–954. https://doi.org/10.1109/JSTSP.2017.2747126.
10. Bolles, R.C.; Baker, H.H.; Marimont, D.H. Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion. *Int. J. Comput. Vis.* **1987**, *1*, 7–55. https://doi.org/10.1007/BF00128525.
11. Ihrke, I.; Restrepo, J.; Mignard-Debise, L. Principles of Light Field Imaging: Briefly Revisiting 25 Years of Research. *IEEE Signal Process. Mag.* **2016**, *33*, 59–69. https://doi.org/10.1109/MSP.2016.2582220.
12. Georgiev, T.; Lumsdaine, A. Focused Plenoptic Camera and Rendering. *J. Electron. Imaging* **2010**, *19*, 021106. https://doi.org/10.1117/1.3442712.
13. Perwaß, C.; Wietzke, L. Single Lens 3D-Camera with Extended Depth-of-Field. In Proceedings of the Human Vision and Electronic Imaging XVII, Burlingame, CA, USA, 1 March 2012; Volume 8291, p. 829108.
14. Bishop, T.E.; Favaro, P. The Light Field Camera: Extended Depth of Field, Aliasing, and Superresolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 972–986. https://doi.org/10.1109/TPAMI.2011.168.
15. Hahne, C.; Aggoun, A.; Velisavljevic, V.; Fiebig, S.; Pesch, M. Refocusing Distance of a Standard Plenoptic Camera. *Opt. Express OE* **2016**, *24*, 21521–21540. https://doi.org/10.1364/OE.24.021521.
16. Hu, X.; Li, Z.; Miao, L.; Fang, F.; Jiang, Z.; Zhang, X. Measurement Technologies of Light Field Camera: An Overview. *Sensors* **2023**, *23*, 6812. https://doi.org/10.3390/s23156812.

17. Wanner, S.; Goldluecke, B. Variational Light Field Analysis for Disparity Estimation and Super-Resolution. *IEEE Transac-Tions Pattern Anal. Mach. Intell.* **2013**, *36*, 606–619.
18. Hahne, C.; Aggoun, A.; Velisavljevic, V.; Fiebig, S.; Pesch, M. Baseline and Triangulation Geometry in a Standard Plenoptic Camera. *Int. J. Comput. Vis.* **2018**, *126*, 21–35. https://doi.org/10.1007/s11263-017-1036-4.
19. Cui, Q.; Zhu, S.; Gao, L. Developing an Optical Design Pipeline for Correcting Lens Aberrations and Vignetting in Light Field Cameras. *Opt Express* **2020**, *28*, 33632–33643. https://doi.org/10.1364/OE.399735.
20. Kim, H.M.; Yoo, Y.J.; Lee, J.M.; Song, Y.M. A Wide Field-of-View Light-Field Camera with Adjustable Multiplicity for Practical Applications. *Sensors* **2022**, *22*, 3455. https://doi.org/10.3390/s22093455.
21. Adelson, E.H.; Bergen, J.R. The Plenoptic Function and the Elements of Early Vision. In *Computational Models of Visual Processing*; Landy, M., Movshon, J.A., Eds.; The MIT Press: Cambridge, MA, USA, 1991; ISBN 978-0-262-29089-0.
22. Ng, R.; Levoy, M.; Brédif, M.; Duval, G.; Horowitz, M.; Hanrahan, P. Light Field Photography with a Hand-Held Plenoptic Camera. Doctoral Dissertation, Stanford University, Stanford, CA, USA, 2005.
23. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. https://doi.org/10.1109/34.888718.
24. Jin, D.; Zhang, S.; Huo, X.; Zhang, W.; Yang, F. A Two-Step Calibration Method for Unfocused Light Field Camera Based on Projection Model Analysis. *arXiv* **2021**, arXiv:2001.03734.
25. Dansereau, D.G.; Pizarro, O.; Williams, S.B. Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Portland, OR, USA, 2013; pp. 1027–1034.
26. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Salt Lake City, UT, USA, 2018; pp. 8183–8192.
27. Chen, L.; Feng, C.; Ma, Y.; Zhao, Y.; Wang, C. A review of rigid point cloud registration based on deep learning. *Front. Neurorobot.* **2024**, *17*, 1281332. https://doi.org/10.3389/fnbot.2023.1281332.
28. Qin, Z.; Yu, H.; Wang, C.; Guo, Y.; Peng, Y.; Ilic, S.; Hu, D.; Xu, K. GeoTransformer: Fast and Robust Point Cloud Registration With Geometric Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 9806–9821. https://doi.org/10.1109/TPAMI.2023.3259038.
29. Feng, S.; Zuo, C.; Yin, W.; Gu, G.; Chen, Q. Micro deep learning profilometry for high-speed 3D surface imaging. *Opt. Lasers Eng.* **2019**,*121*,416-427. https://doi.org/10.1016/j.optlaseng.2019.04.020.
30. Wang, Y. Deep learning-based high dynamic range 3D reconstruction. *Sci. Rep.* **2026**, *16*, 563. https://doi.org/10.1038/s41598-025-30204-0.
31. Li, L.; Xiang, S.; Yang, Y.; Yu, L. Multi-camera interference cancellation of time-of-flight (TOF) cameras. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, USA, 27–30 September 2015; pp. 556–560.
32. Wu, X.; Jiang, L.; Wang, P.S.; Liu, Z.; Liu, X.; Qiao, Y.; Ouyang, W.; He, T.; Zhao, H. Point Transformer V3: Simpler, Faster, Stronger. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024; pp. 4840–4851.
33. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, 30, 328–341. https://doi.org/10.1109/TPAMI.2007.1166.
34. Xu, G.; Wang, X.; Ding, X.; Yang, X. Iterative Geometry Encoding Volume for Stereo Matching. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–22 June 2023; pp. 21919–21928.
35. Tosi, F.; Bartolomei, L.; Poggi, M. A Survey on Deep Stereo Matching in the Twenties. *Int. J. Comput. Vis.* **2025**, *133*, 4245–4276. https://doi.org/10.1007/s11263-024-02331-0.
36. Shin, C.; Jeon, H.G.; Yoon, Y.; Kweon, I.S.; Kim, S.J. Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4748–4757.
37. Song, Q.; Campione, S.; Boyraz, O.; Capolino, F. Silicon-Based Optical Leaky Wave Antenna with Narrow Beam Radiation. *Opt. Express OE* **2011**, *19*, 8735–8749. https://doi.org/10.1364/OE.19.008735.
38. Campione, S.; Guclu, C.; Song, Q.; Boyraz, O.; Capolino, F. An Optical Leaky Wave Antenna with Si Perturbations inside a Resonator for Enhanced Optical Control of the Radiation. *Opt. Express OE* **2012**, *20*, 21305–21317. https://doi.org/10.1364/OE.20.021305.

39. Wang, S.; Chen, X.; Ding, G.; Li, Y.; Xu, W.; Zhao, Q.; Gong, Y.; Song, Q. A Lightweight Localization Strategy for LiDAR-Guided Autonomous Robots with Artificial Landmarks. *Sensors* **2021**, *21*, 4479. https://doi.org/10.3390/s21134479.
40. Antensteiner, D.; Štolc, S.; Valentín, K.; Blaschitz, B.; Huber-Mörk, R.; Pock, T.; Štolc, S.; Valentín, K.; Blaschitz, B.; Huber-Mörk, R.; et al. High-Precision 3D Sensing with Hybrid Light Field & Photometric Stereo Approach in Multi-Line Scan Framework. *Electron. Imaging* **2017**, *29*, 52–60. https://doi.org/10.2352/ISSN.2470-1173.2017.9.IRIACV-268.