Learned Display Radiance Fields with Lensless Cameras

Ziyang Chen
University College London
London, United Kingdom
ziyang.chen.22@ucl.ac.uk

Yuta Itoh
Institute of Science Tokyo
Tokyo, Japan
itoh@comp.isct.ac.jp

Kaan Akşit
University College London
London, United Kingdom
k.aksit@ucl.ac.uk

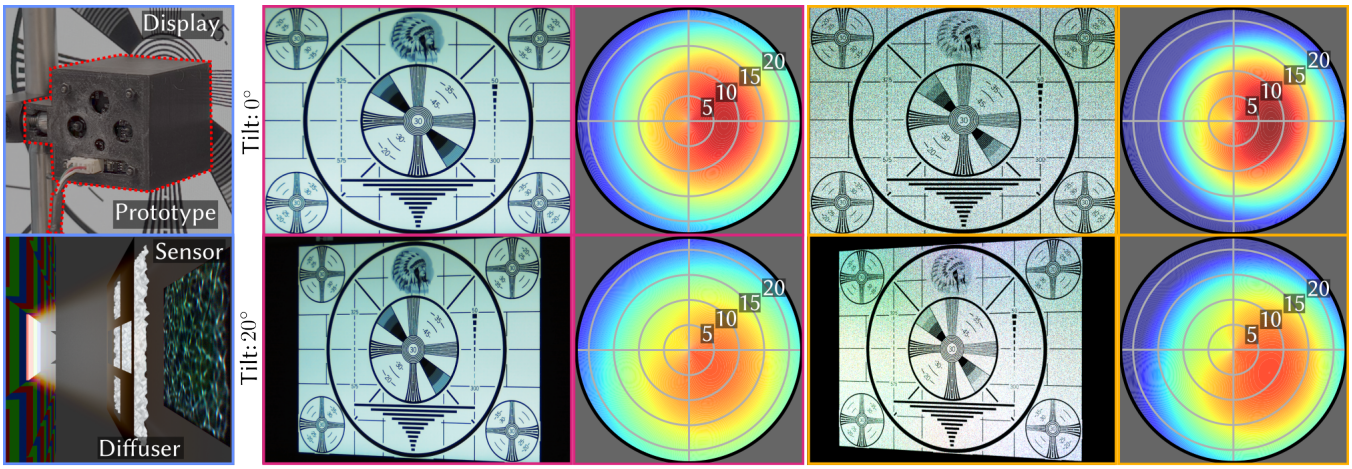

Figure 1: (a) Our lensless camera placed in front of the display captures the light field emitted from the display pixels. (b) A conventional camera captures a real-world photograph of the test pattern on the display. (c) Our learned model estimates a rendered view from the viewpoint of the real-world photograph displayed in the middle column. The right columns in (b) and (c) present the display's angular intensity distributions in spherical coordinates. The radius denotes the combined angular deviation from the optical axis, computed from the horizontal and vertical incidence angles. These plots illustrate how the relative intensity changes with viewing angle.

Abstract

Calibrating displays is a basic and regular task that content creators must perform to maintain optimal visual experience, yet it remains a troublesome issue. Measuring display characteristics from different viewpoints often requires bulky equipment and a dark room, making it inaccessible to most users. To avoid such hardware requirements in display calibrations, our work co-designs a lensless camera and an Implicit Neural Representation based algorithm for capturing display characteristics from various viewpoints. More specifically, our pipeline enables efficient reconstruction of light fields emitted from a display from a viewing cone of $46.6° \times 37.6°$. Our emerging pipeline paves the initial steps towards effortless display calibration and characterization.

CCS Concepts

• **Hardware** → **Emerging simulation**; **Hardware validation**; • **Computing methodologies** → **Computational photography**.

SA Technical Communications '25, December 15–18, 2025, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2136-6/2025/12
https://doi.org/10.1145/3757376.3771381

ACM Reference Format:
Ziyang Chen, Yuta Itoh, and Kaan Akşit. 2025. Learned Display Radiance Fields with Lensless Cameras. In *SIGGRAPH Asia 2025 Technical Communications (SA Technical Communications '25), December 15–18, 2025, Hong Kong, Hong Kong*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3757376.3771381

1 Introduction

Professional content creators use off-the-shelf colorimeters [1] regularly to maintain a color consistency in the workflow. Similarly, display engineers conduct calibrations by measuring various aspects of screens including but not limited to luminance, chromaticity, and contrast ratios [2]. This process requires specialized equipment such as spectroradiometers or goniophotometers [3].

Display chromaticity and luminance vary with viewing angle, affecting how users perceive content from different positions in front of a screen. But most calibration tools assume a fixed viewing position leading to invalid assessments for other viewing positions. This is not only problematic for display manufacturers but also for

[1] https://www.adobe.com/creativecloud/video/discover/how-to-calibrate-monitor.html
[2] https://www.admesy.kr/articles/display-calibration-workflows-rd-to-mass-production/
[3] https://www.lisungroup.com/news/technology-news/application-of-goniophotometers-in-display-and-projection-technologies.html

professionals that need consistent color edits for various geometric configurations. The ISO standards [ISO 2008] define procedures for assessing the uniformity distribution of Flat Panel Display (FPD) at various viewing directions by rotating either the display or measurement instrument in a dark room. However, these requirements with numerous captures hinder the end users from performing regular calibration, which is critical for virtual production display walls [4], professional graphic design, and video editing [5]. *An accessible method with a reasonable amount of captures is needed to measure the image quality of the displays from arbitrary viewpoints.*

Our work introduces a calibration tool that characterizes the display pixel emission patterns at different viewing angles with simpler setup and fewer captures. Specifically, we design a paired hardware and software: (1) a lensless camera [Chen et al. 2024; Kingshott et al. 2022] capturing the incoming light of a display pixel from a range of directions and (2) an Implicit Neural Representation (INR) that encodes the display angular responses. Our work makes the following contributions:

- **Lensless Camera Prototype**. We design a lensless camera using a phase mask and aperture array to capture display characteristics across a practical angular range. The phase mask is positioned 30 *mm* from the display pixel, with the imaging sensor placed 10 *mm* behind it to ensure spatial invariance of the diffuser. This configuration enables horizontal and vertical incident angle coverage of approximately **46.6** and **37.6** degrees, respectively.
- **INR for Display Pixel Characterization**. To efficiently represent the light fields from the lensless captures and generate novel views of unseen display pixels, we train an Multi-Layer Perceptron (MLP) that performs end-to-end reconstruction based on different display positions and incident angles. This MLP model can be trained with only **nine** positions on the display, enabling the reconstruction of the light field for each pixel on the display.

We conduct experiments on a Liquid Crystal Display (LCD) and demonstrate that our method can reconstruct the display with different viewpoints. Our codebase is available at GitHub:complight.

2 Methods

Our method reconstructs the light field representation of a display from lensless captures of individual pixels. We detail our pipeline consisting of hardware and software components in the subsequent sections. Our notations are summarized in Tbl. 1. Upper-case bold letters denote matrices, while lower-case bold letters represent vectors. Scalars and spatial indices are denoted by non-boldface letters. The calligraphic letters indicate the functions.

2.1 Lensless 2D Image Formation

Capturing the incoming light from pixels across multiple directions is crucial in our application. We choose the lensless cameras for their ease of modification and flexibility.

[4]https://partnerhelp.netflixstudios.com/hc/en-us/articles/1500002086641-Common-Virtual-Production-Challenges-Potential-Solutions
[5]https://displaycal.net/

$\mathbf{H}$	Coefficient matrix for lensless forward model
$\mathbf{Y}$	Lensless camera 2D measurements
$\mathbf{X}$	Lensless camera unknown scene intensity
$\mathbf{y}$	Flattened lensless camera measurements
$\mathbf{x}$	Flattened unknown scene intensity
h_I, w_I	Height and width of the lensless image
h_S, w_S	Height and width of the scene
$C(\cdot)$	Cropping function
$\mathcal{Z}(\cdot)$	Zero padding function
$\mathbf{h}$	Pre-captured PSF
x, y	Display pixel coordinates
u, v	Light field angular coordinates
s, t	Image coordinates
$\mathbf{L}$	Light field
$\mathbf{I}_{\text{pred}}$	Predicted lensless capture
$\mathbf{I}_{\text{gt}}$	Ground truth lensless capture
$\mathbf{F}_\theta$	Multi-layer perceptron
$\gamma(\cdot)$	Positional encoding
n	Aperture size
m	Distance between display pixel and aperture
α	Incident angle of the light rays

Table 1: Notation used throughout this section.

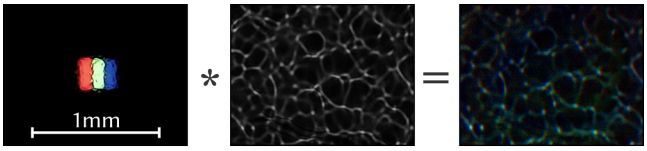

Figure 2: The display pixel captured with a microscope (left). The Point Spread Function (PSF) captured with the proposed lensless camera (middle). The captured lensless image from the display pixel (right).

Our lensless camera contains a diffuser that alters the light propagation paths, resulting in distorted images on the sensor as shown in Fig. 2. Let h_I and w_I denote the height and width of the sensor, and h_S and w_S denote the height and width of the scene. The forward imaging process can be expressed as

$$\mathbf{y}_I = \mathbf{H}\mathbf{x}_S, \tag{1}$$

where $\mathbf{y}_I \in \mathbb{R}^{(h_I \cdot w_I)}$ denotes the flattened measurement vector from the sensor, $\mathbf{H} \in \mathbb{R}^{(h_I \cdot w_I)\times(h_S \cdot w_S)}$ is a coefficient matrix modeling the input light's interaction with the diffuser, and $\mathbf{x}_S \in \mathbb{R}^{(h_S \cdot w_S)}$ represents the flattened intensity vector of the unknown scene. However, reconstructing the full $\mathbf{H}$ matrix is computationally intractable. To address this, we position the diffuser close to the sensor to ensure spatial invariance in the forward model, allowing it to be approximated as a linear convolution:

$$\begin{aligned}\mathbf{Y} &= C\big(\mathcal{Z}(\mathbf{X}) * \mathcal{Z}(\mathbf{h})\big) \\ &= C\left(\mathcal{F}^{-1}\left(\mathcal{F}\big(\mathcal{Z}(\mathbf{X})\big) \cdot \mathcal{F}\big(\mathcal{Z}(\mathbf{h})\big)\right)\right),\end{aligned} \tag{2}$$

where $\mathbf{Y} \in \mathbb{R}^{h_I \times w_I}$ denotes the 2D measurements, $\mathbf{X} \in \mathbb{R}^{h_S \times w_S}$ denotes the unknown scene intensity, $*$ denotes the 2D linear convolution operation, $C(\cdot)$ is the cropping operation, $\cdot$ is the Hadamard product, $\mathcal{Z}(\cdot)$ is the zero padding operation to ensures dimension consistency and avoids circular convolution artefact, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Discrete Fourier Transform (DFT) and Inverse Discrete Fourier Transform (IDFT), respectively, and $\mathbf{h} \in \mathbb{R}^{h_S \times w_S}$ is the pre-captured PSF. To reconstruct for the unknown scene $\mathbf{X}$, we need to solve an inverse problem of the Eq. (2) by minimizing the loss:

$$\hat{\mathbf{X}} \leftarrow \underset{\mathbf{X}}{\operatorname{argmin}}\ \mathcal{L}\big(\mathbf{Y}_{\text{gt}}, C(\mathcal{Z}(\mathbf{X}) * \mathcal{Z}(\mathbf{h}))\big), \tag{3}$$

where $\mathbf{Y}_{\text{gt}}$ is the ground truth lensless image.

2.2 Lensless Light Field Formation

We leverage the lensless camera system to capture the unknown light fields for pixels at different locations on the display $\mathbf{L} = (u, v, s, t)$, where the (u, v) and (s, t) are the angular and spatial coordinates. An image formed by the directional light rays from a single or multiple light sources satisfies:

$$\mathbf{I}(s,t) = \int_{\Omega_x} \int_{\Omega_y} \mathbf{L}(u,v,s,t)\, du\, dv, \tag{4}$$

where $\{\Omega_x, \Omega_y\}$ is a subset of all the angles which the light rays reach for the imaging sensor. To simplify the problem, we model the phase mask as an array of pinholes that scatter incident light rays arriving at angles (u', v') onto the imaging sensor, forming distinct patterns that we refer as sub-aperture images. The formation of each sub-aperture image can be modeled as the 2D convolution of the scene information with its corresponding PSF patch, $\mathfrak{h}(u', v')$. To compute this convolution efficiently across all angular dimensions, we employ a 4D DFT on the spatial coordinates:

$$\mathbf{I}_{\text{gt}} = C\left(\mathcal{F}^{-1}\left(\mathcal{F}(\mathcal{Z}(\mathbf{L})) \cdot \mathcal{F}(\mathcal{Z}(\mathfrak{h}))\right)\right). \tag{5}$$

Finally, we pose the recovery of the light field as an inverse problem:

$$\hat{\mathbf{L}} \leftarrow \underset{\mathbf{L}}{\operatorname{argmin}}\ \mathcal{L}\Big(\mathbf{I}_{\text{gt}}, C\big(\mathcal{Z}(\mathbf{L}) * \mathcal{Z}(\mathfrak{h})\big)\Big), \tag{6}$$

where $\mathbf{I}_{\text{gt}}$ denotes the ground truth lensless capture.

2.3 Implicit Neural Representation

We utilize an MLP to efficiently represent the continuous pixel light field samples from the lensless captures to avoid extensive sampling on the target screen, which satisfies:

$$\hat{\mathbf{L}}(\mathbf{x}, \mathbf{y}, u, v, s, t) = \mathbf{F}_\theta(\mathbf{x}, \mathbf{y}, u, v, s, t), \tag{7}$$

where $\mathbf{F}_\theta$ is an MLP with parameters θ. This network takes the spatial and angular coordinates as input and outputs the corresponding pixel color value. The variables $\mathbf{x}$ and $\mathbf{y}$ denote the illuminated pixel coordinates on the display. Before the input coordinates p are fed into the MLP, we apply positional encoding [Mildenhall et al. 2021] to preserve high-frequency details:

$$\gamma(p) = \Big(\sin(2^0\pi p), \cos(2^0\pi p), \cdots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p)\Big), \tag{8}$$

We apply the positional encoding with different frequency levels in the input coordinate groups: $\mathbf{e} = \big(\gamma_0(\mathbf{x}, \mathbf{y}), \gamma_1(u, v), \gamma_1(s, t)\big)$. Finally,

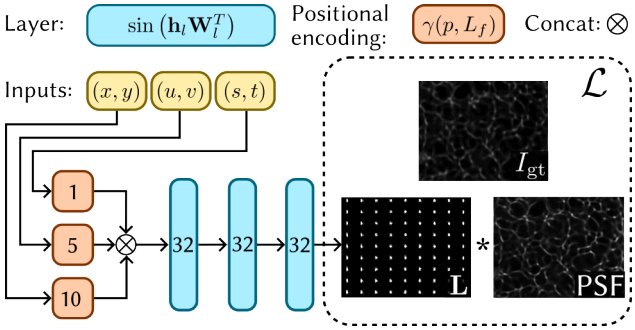

Figure 3: Each MLP layer consists of 32 neurons and a sinusoidal activation function. We apply positional encoding with varying frequency levels (L_f) to each input coordinate group. After concatenating these encoded features, the model processes them to reconstruct the light field. We then perform linear convolution between this reconstruction and the pre-captured PSF to generate the predicted lensless image ($\mathbf{I}_{\text{pred}}$). Finally, we compute the loss with $\mathbf{I}_{\text{pred}}$ and $\mathbf{I}_{\text{gt}}$.

the minimization problem in Eq. (6) is reformulated as:

$$\hat{\theta} \leftarrow \underset{\theta}{\operatorname{argmin}}\ \mathcal{L}\Big(\mathbf{I}_{\text{gt}}, C\big(\mathcal{Z}(\mathbf{F}_\theta(\mathbf{e})) * \mathcal{Z}(\mathfrak{h})\big)\Big), \tag{9}$$

where $\hat{\theta}$ denotes the optimized parameters of the MLP.

Loss function. We employ the L_1 norm to quantify the discrepancy between $\mathbf{I}_{\text{pred}}$ and $\mathbf{I}_{\text{gt}}$, and use $\mathcal{R}$ to penalize predicted pixel values exceeding the range $[0, 1]$. We define:

$$\mathcal{R}(\mathbf{I}_{\text{pred}}) = \sum \max(\mathbf{I}_{\text{pred}} - 1, 0) + \sum \max(-\mathbf{I}_{\text{pred}}, 0). \tag{10}$$

The total training loss is then:

$$\mathcal{L} = \lambda_0 \|\mathbf{I}_{\text{gt}} - \mathbf{I}_{\text{pred}}\|_1 + \lambda_1 \mathcal{R}(\mathbf{I}_{\text{pred}}), \tag{11}$$

where λ_0, and λ_1 represent weights ($\lambda_0 = 1$, $\lambda_1 = 10^{-7}$).

3 Evaluation and Discussion

Hardware. Our hardware uses a 0.5° engineered diffuser (Edmund 35-860) with five 2.5 $mm \times 3$ mm apertures positioned 10 mm from the imaging sensor as depicted in Fig. 4. We experimentally determine this spacing to ensure spatial invariance and sharp caustic patterns, satisfying Eq. (2). Moreover, we place the apertures such that we maximize the angle of rays arriving from a pixel landing on an imaging sensor. This aperture placement maximizes ray angles from pixels to the sensor, though sensor dimensions limit the maximum supported angle. In our implementation, the limited imaging sensor dimensions dictate the locations of these apertures, posing a constrain on the maximum angle we can support with our aperture arrangement. We 3D print a housing maintains 30 mm between diffuser and screen, as demonstrated in left image Fig. 5. We capture a stack of PSF for each aperture using a white LED with a 100 μm pinhole (Thorlabs P100HKb), then mount the prototype in front of the display and activate pixels sequentially. To account

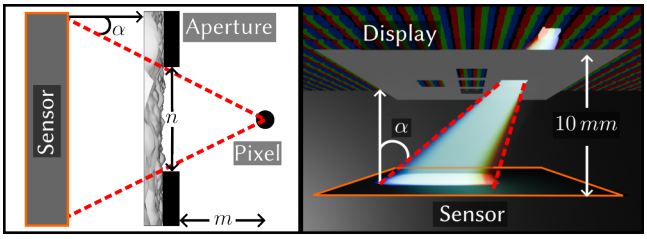

Figure 4: The incident angle (α) of the light rays is determined by the distance between the aperture and the display pixel (m), as well as the size of the aperture (n) (left). Our proposed aperture array expands the incident angle range by turning on each pixel one by one (right).

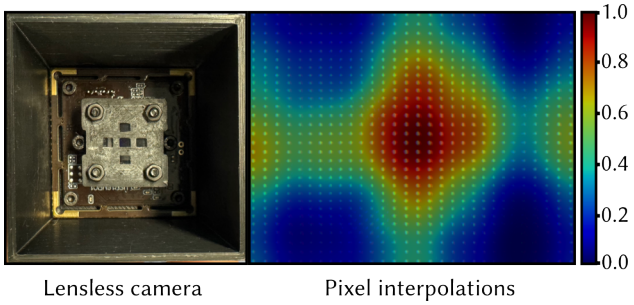

Figure 5: The top view of our lensless camera (left). The simulated display pixels with unseen incident angles and the corresponding overlay intensity heatmap (right).

for LCD backlight bleeding [6], we sample nine screen positions to capture the full spatial distribution (Supplementary Sec. 1).

Implicit Neural Representation. We divide lensless captures into 9×9 sub-aperture images using a 54×70 pixel sliding window. Later, we feed them into a MLP that consists of 3 fully connected layers with sinusoidal activation functions and 32 channels each. To improve generalization, we inject random noise into the coordinates from Sec. 2.3: display (std: 5×10^{-3}), angular (std: 1×10^{-2}), and subview (std: 1×10^{-3}). We then apply positional encoding Eq. (2.3) with 1, 5, and 10 frequency levels for display, angular, and spatial coordinates, respectively (see Fig. 3 and Supplementary Sec. 2). We optimize using Adam optimizer with an initial learning rate of 0.001 that decays linearly over 800 epochs and clip gradient norms to 1.0. We train the model on independent color channels to mitigate crosstalk, which takes approximately one hour on an NVIDIA RTX 2070 GPU, while inference takes 0.01 seconds per 1080p frame.

Quantitative Evaluations. To assess pixel consistency, we measure intensity variations in reconstructed light fields across continuous horizontal and vertical incident angles. We sample nine display positions with five measurements at different apertures per position, train the MLP on this data, and estimate pixel values at unseen angles. Fig. 5 shows smooth predicted intensity changes,

[6] https://pixiogaming.com/blogs/latest/understanding-backlight-bleed-in-ips-panels-causes-and-solutions

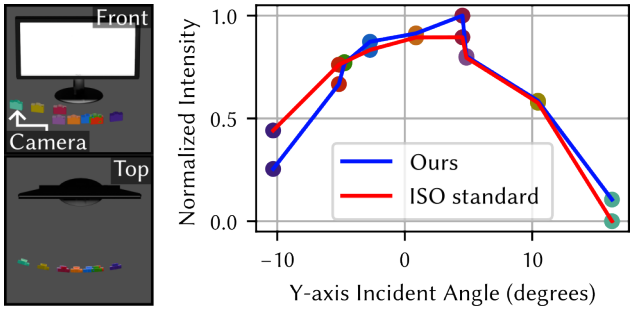

Figure 6: The illustration of the camera poses and the display (left). The average intensity that normalized to $[0, 1]$ from the ISO standard and ours, with data samples color-coded by the corresponding camera poses (right).

demonstrating the model's angle-dependent reconstruction capability. Our architecture outperforms a vanilla MLP baseline with fully connected ReLU layers across all metrics (Tbl. 2). To benchmark our method, we follow the ISO standard [ISO 2008]: align a camera with the display in a dark room, display full-screen white stimuli, and capture images from $-10°$ to $16°$ vertical incident angles. Fig. 6 shows our method reproduces the ISO intensity trend, validating its physical plausibility. We capture light field data from nine display positions without camera rotation or controlled lighting, reducing measurement time and simplifying view-dependent calibration while maintaining comparable accuracy.

Table 2: Model Comparison

Methods	PSNR ↑ (dB)	SSIM ↑	MSSIM ↑	Train (h)	Inference (s)
Ours	**19.54**	**0.9165**	**0.9549**	1	0.01
Vanilla	9.93	0.5327	0.8049	0.8	0.003

Limitations and Future Works. Our $46.6°$ angular coverage remains well below the $240°$ required in professional display calibration. We could extend this range by co-optimizing the apertures and the diffuser design. Beyond angular coverage, pixel-wise training hinders scalability for full-panel characterization. To overcome these constraints, we outline several promising research directions. First, developing an end-to-end pipeline that eliminates PSF convolutions, cropping, and padding. Second, hash encoding [Müller et al. 2022] or Gaussian splatting could model light fields more efficiently.

References

Ziyang Chen, Mustafa Dogan, Josef Spjut, and Kaan Akşit. 2024. SpecTrack: Learned Multi-Rotation Tracking via Speckle Imaging. In *SIGGRAPH Asia 2024 Posters*. 1–2.

ISO. 2008. ISO 9241-305:2008(E) – Ergonomics of human-system interaction – Part 305: Optical laboratory test methods for electronic visual displays. https://www.iso.org/standard/39091.html. Accessed: 2025-06-16.

Oliver Kingshott, Nick Antipa, Emrah Bostan, and Kaan Akşit. 2022. Unrolled primal-dual networks for lensless cameras. *Optics Express* 30, 26 (2022), 46324–46335.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)* 41, 4 (2022), 1–15.