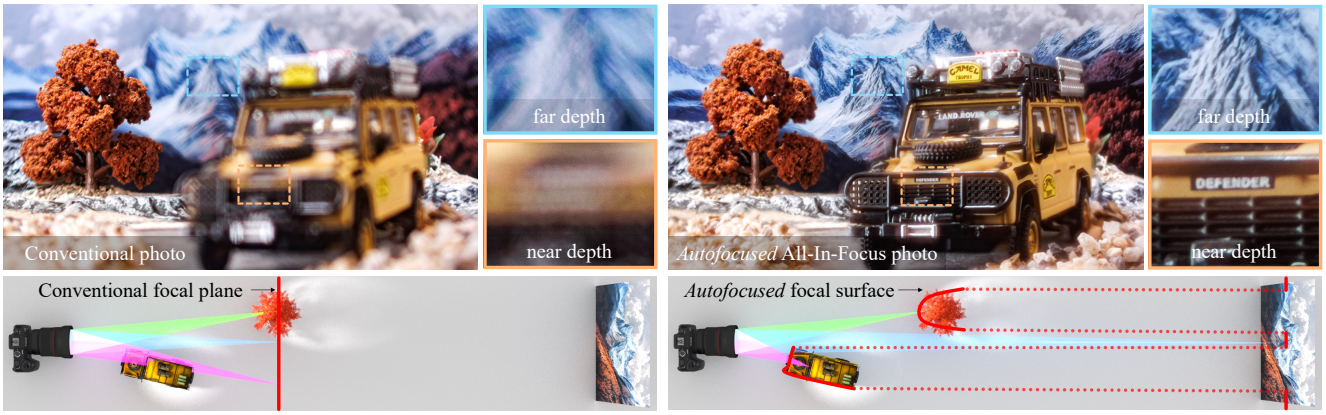
Spatially-Varying Autofocus

Yingsi Qin, Aswin C. Sankaranarayanan, Matthew O'Toole
Carnegie Mellon University



(a) Conventional photo and its confined focal plane

(b) All-In-Focus photo and its *spatially-varying autofocused* focal surface (ours)

Figure 1. Spatially-varying autofocus to produce an optical all-in-focus image. **Left:** A conventional photo with a regular lens, where objects at a single focal plane appear sharp. **Right:** An all-in-focus photo captured through spatially-varying autofocusing. To achieve this, we combine (i) a programmable lens with spatially-varying control over focus, and (ii) a spatially-varying autofocus algorithm to drive the focus of this lens. Note that this is an optically-captured image of a real scene with no post-capture processing used.

Abstract

A lens brings a single plane into focus on a planar sensor; hence, parts of the scene that are outside this planar focus plane are resolved under defocus. Can we break this precept by enabling a "lens" that can change its depth of field arbitrarily? This work investigates the design and implementation of such a computational lens with spatially-selective focusing. Our design uses an optical arrangement of a Lohmann lens and a phase-only spatial light modulator to allow each pixel to focus at a different depth. We extend classical autofocusing techniques to the spatially-varying scenario where the depth map is iteratively estimated using contrast and disparity cues, enabling the camera to progressively shape its depth-of-field to the scene's depth. By obtaining an all-in-focus image optically, our technique advances upon prior work in two key aspects: the ability to bring an entire scene in focus simultaneously, and the ability to maintain the highest possible spatial resolution.

1. Introduction

Core to any imaging system is the lens: an optical component designed to gather rays of light from a scene and form a focused image on a sensor. The focusing ability of a lens, however, only applies to a *single plane* in the scene. To form a focused image, the subject should be positioned at the *focal plane*, *i.e.*, at a fixed depth from the camera. Points that do not lie on this *focal plane* appear blurry, and the amount of defocus increases progressively as the points move further away from the plane.

Reducing the size of a lens' aperture decreases the amount of defocus and increases the depth of field (*i.e.*, the region near the focal plane where points appear to be in sharp focus). However, this comes at the cost of reduced light throughput. Furthermore, smaller apertures increase diffractive blur [8], making content within the depth of field less sharp. These restrictions on focus are attributed to the traditional design of a lens, which offers the ability to move the focal plane (*e.g.*, by adjusting a focus ring) but maintains its shape. Hence, we raise the question: in placeof a single *focal plane*, is it possible to optically program a *focal surface* that can adapt to any scene geometry?

This paper advances the design and implementation of a computational lens capable of *spatially-varying focus*—one that allows a scene *in its entirety* to be simultaneously in focus on an image sensor *even when the scene is highly non-planar*. Our approach relies on adapting the so-called Lohmann lens [21], which is a focus-tunable lens produced

	optical sharpness	# of images required	all-in-focus generation	outputs depth
small aperture	low	one	optical	no
cubic phase plate [9]	low	one	deconvolution	no
focal sweep [16, 22]	low	one	deconvolution	no
focal stack [23, 35]	high	many	contrast metric	yes
coded aperture [7, 11, 17, 18, 26, 27, 37, 41]	low	one	depth-dependent deconv.	yes
light field cameras [4, 10, 19, 24, 40]	low	one	contrast metric	yes
dual-pixel image deblurring [1, 2, 38, 42]	low	one	hard inverse problem	yes
spatially-varying autofocus (ours)	high	two	optical	yes

Table 1. Comparison of all-in-focus imaging techniques. **Optical sharpness:** Most methods either use a small effective aperture (increasing the amount of diffraction blur), or intentionally blur the photos (*e.g.*, to create depth-invariant blur). Our approach forms all-in-focus images by bringing each scene point into focus optically, while maintaining a large aperture. **# of images required:** Our method requires at least one image to approximate the scene geometry, and a second image to form the all-in-focus image. Moreover, our method is well suited for dynamic settings, where each frame determines the focus for the next frame. **All-in-focus generation:** Unlike most techniques, our approach forms images using an *all-optical* process; no additional computational post-processing is required. **Outputs depth:** A useful byproduct of several methods is the ability recover a scene's depth map.

by relative movement between two cubic lenses. Prior work by Qin et al. [28], in the context of near-eye virtual reality (VR) displays, has shown that a rearrangement of the Lohmann lens, along with the use of a phase-only spatial light modulator (SLM), can control the perceived depth of pixels on a display. Our work extends this concept to the imaging scenario to provide unprecedented control over a camera's focusing capabilities, and introduces a novel category to the solution space for all-in-focus imaging.

Contributions. This paper proposes a programmable, spatially-varying lens for optical all-in-focus imaging and flexible depth-of-field manipulation.

Optical all-in-focus (AIF) image. The centerpiece of our contribution is the ability to acquire an optical AIF image given knowledge of the scene depth, *i.e.*, our technique uses knowledge of the depth to resolve all scene points in sharp focus on the sensor. A hallmark of this result is that, unlike prior work in all-in-focus imaging, our imaging process does not require computational post-processing.

Spatially-varying autofocusing. To recover the depth map of the scene, we extend traditional ideas in contrast- and phase-based autofocus to their spatially-varying counterparts. We show that our system can progressively bring the entire scene into focus using as few as two images. This has the benefit of matching the quality and depth resolution of the focus stacking without requiring a large image set.

Code, datasets, and real-time video demonstrations of dynamic scenes are available on our project website: https://imaging.cs.cmu.edu/svaf [29].

Limitation. Our current optical prototype is light inefficient due to the use of polarization-based phase modulation and a beamsplitter; note that at most one-eighth of the incident light reaches the sensor, due to light passing through a polarizer once and through a 50/50 beamsplitter twice.

2. Related Work

We briefly discuss techniques used for extended depth-of-field imaging, with a focus on key differences from our proposed approach, as summarized in Tab. 1.

Coded aperture systems. One of the classic problems in computer vision is that of depth from defocus. For a traditional lens, this problem is ill-conditioned since the circular shape of the defocus kernel is not sufficiently discriminative. To resolve this, coded-aperture systems reshape the aperture of the lens using amplitude or phase masks. A seminal work in this space is that of Dowski and Cathey [9], who show that a cubic phase mask has a depth-invariant blur kernel that can be used to deblur and obtain the AIF image. More recent work have concentrated on enabling better discriminability of depth, either by using amplitude apertures [17, 37] or phase-based ones [7, 11, 18, 26, 27, 41].

Also loosely falling under the broad umbrella of coded aperture system are dual-pixel (DP) sensors, where each pixel has two sub-pixels under a single microlenslet. The resulting system acquires two images simultaneously, each from different halves of the lens aperture, thereby emulating a small baseline stereo setup. DP sensors have found extensive use in autofocusing systems [13]; however, their ability to provide stereo pairs has also enabled their use in AIF image and depth estimation [1, 2, 38, 42].

The proposed work differs from this class of techniques in two distinct ways. First, we do not need any computational post-processing since we optically form an all-in-focus image on the sensor. Second, since we optically focus on scene points, our images are sharper, modulo non-idealities of the optics.

Focus stacking. One of the ways to regularize the depth-from-defocus problem is to capture a dense focal stack—that is, a collection of images obtained by sweeping the focus plane through the scene. Using local contrast as a cue,

we can estimate depth and construct an all-in-focus image of the scene [23, 35]. However, this technique is slow, due to the need to capture a collection of images, and does not handle scene dynamics well.

Focal sweep and flexible depth-of-field imaging. Nagahara et al. [22] explore the idea of capturing a single image, where the focal plane changes throughout its exposure. One of their results is the construction of a near-invariant defocus kernel by sweeping focus, linearly in diopter space, within the exposure; similar to Dowski and Cathey's work, this creates a depth-invariant blur that can be computationally removed. A different result enables a flexible and non-planar depth of field by synchronizing the focus plane to a rolling shutter [16]. Our work goes beyond these results by avoiding post-process computation, as well as enabling a freeform shape for the depth of field. Such control of the depth of field was previously only possible via post-processing, for example, a focal stack [12, 31, 44].

Light fields. Light field cameras [4, 10, 19] sample the incident 4D space of light rays, and enables refocusing and AIF image synthesis to be done as a post-processing operation. Light fields can be measured by placing a microlens array in front of a sensor [24], or using a camera array [40], both of which sacrifice the spatial resolution of the sensors to increase angular resolution. A high angular resolution is critical for better depth selectivity, but this comes at a commensurate loss in spatial resolution for the reconstructed images.

Autofocusing systems. Our spatially-selective focusing technique revisits classical ideas in passive autofocusing to estimate the depth map of the scene, in order to obtain an all-in-focus image. Specifically, we develop spatially-varying counterparts to the traditional contrast and phase detection autofocusing techniques.

Contrast detection autofocus (CDAF) is one of the primary methods used by digital cameras to focus a lens. The approach involves adjusting the focus settings of the lens until the camera detects the highest contrast (usually at a few select locations). CDAF techniques can be sped up using various hill-climbing techniques [15, 33, 36] to estimate the peak contrast using a sparser set of photographs.

Phase detection autofocus (PDAF) is an alternate technique that is commonly available in cameras with a DP sensor. When a scene point is in focus, the two images captured by the corresponding sub-pixels match. Otherwise, disparity is introduced between the two views. The signed disparity determines the lens focus for the scene point.

In this work, we extend both techniques to our spatially-selective framework, where we apply them to bring all regions of the scene into focus simultaneously.

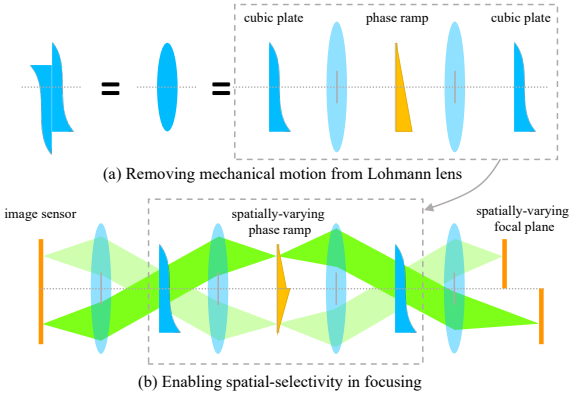


Figure 2. Split-Lohmann computational lens expands upon the Lohmann lens in two key steps. **(a)** First, it achieves a programmable focus-tunable lens by collocating the two cubic phase plates of the Lohmann lens using a 4f relay and placing a phase ramp at the Fourier plane whose slope controls the effective focal length. **(b)** We can optically collocate an image sensor with the SLM. Since each pixel on the sensor is resolved on the SLM, displaying a spatially-varying phase ramp allows local focus control.

3. The Split-Lohmann Computational Lens

The ideas of this paper are inspired by a recent result on the design of a VR display [28], which proposes the Split-Lohmann lens, a computational lens that can spatially vary the focal length. For completeness, we briefly discuss this prior work before delineating our specific contributions.

The Split-Lohmann design relies on a specific kind of focus-tunable lens called a Lohmann or Alvarez lens [5, 21] consisting of two translating cubic phase plates. Suppose that the optical profiles of the two cubic plates are given by $h_1(x) = \kappa x^3$ and $h_2(x) = -\kappa x^3$, where κ is a curvature-related parameter. When stacked together with a lateral offset Δ, the resulting phase modulation is given as

$$h_1(x+\Delta) + h_2(x-\Delta) = \kappa\left(6\Delta x^2 + 2\Delta^3\right) \qquad (1)$$

Ignoring the constant term that is independent of x, we get a phase modulation that is quadratic in x, *i.e.*, a lens whose focal length is inversely proportional to Δ. Hence, lenses of different focal lengths can be obtained by changing the amount of translation between the cubic phase plates (see Fig. 2). The Split-Lohmann display advances the Lohmann lens by first removing the mechanical translation required, and second, achieving independent local control of focal length for different regions on an OLED display.

We propose *inverting the function* of this optical system, by replacing the OLED display with a camera sensor and adding a camera lens. The result is a Split-Lohmann computational lens that now offers a camera the ability to spatially-vary its focus.

4. Spatially-Varying Autofocus

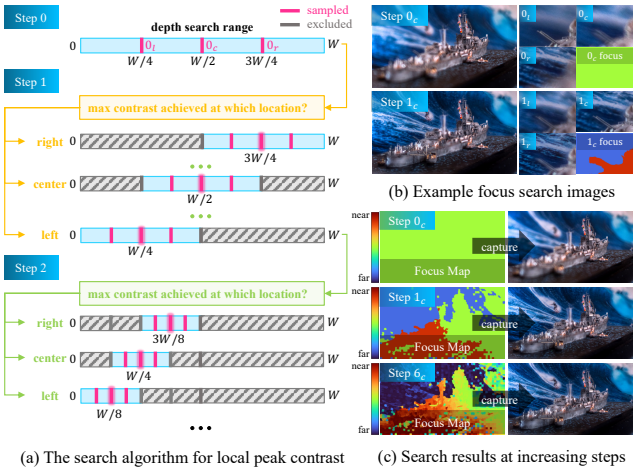
Programming a Split-Lohmann computational lens to form all-in-focus images requires solving a spatially-varying autofocusing problem. While we can use a second device devoted to depth estimation—either in the form of a passive stereo camera, a structured light 3D scanner, or a time-of-flight device—we consider a self-contained autofocus loop where our device progressively estimates depth and revises its focus setting. We take inspiration from existing AF solutions used in conventional cameras to drive lens focus: *Contrast Detection Autofocus (CDAF)* and *Phase Detection Autofocus (PDAF)*. Our goal here is to design the spatially-varying counterparts to these techniques, where we recover a depth image that is used to compute the necessary SLM pattern to perform simultaneous local autofocus and bring the entire field of view in focus simultaneously.

4.1. Contrast Detection Autofocus (CDAF)

Contrast is a popular focus metric for driving focus, as it can be readily implemented on most imaging systems. The basic premise of this approach relies on the observation that contrast of a local region (say, a patch) is maximized when it is in focus. CDAF techniques, hence, search for the focus setting that maximizes the contrast of a pre-determined patch. We extend CDAF to its spatially-varying counterpart by identifying an independent focus parameter for every region that maximizes its image contrast.

Our approach relies on the insight that contrast of a patch as a function of depth is often smooth and, more importantly, unimodal. This property allows us to design an efficient search strategy by progressively narrowing down the focus/depth range where the mode can lie. Suppose that the total working range of focus, as measured in diopters (the reciprocal of depth), is from 0 to W diopters. We obtain three images that correspond to the focus at $\frac{W}{4}$, $\frac{W}{2}$, and $\frac{3W}{4}$ diopters. We can estimate the contrast at each local patch across these three images, and the focus which has the maximum contrast (across the three images) allows us to reduce the range centering the contrast maxima. For example, if the maximum value occurs at $\frac{W}{4}$, then the true maxima must be between the 0 and $\frac{W}{2}$ diopters. In each iteration, the search range reduces by half, and we can repeat the technique over the refined focus range. Note that, this procedure happens in parallel and independently at each patch or region in the image, enabled by the spatially-selective focusing capability of our device. Since each iteration comprises of taking three images[1] spanning a reduced search range, we obtain the same performance of linear search, but with a number of images that is logarithmic in comparison.

[1]In reality, two images suffice as the center focus in an iteration is one of the three focus settings used in a previous iteration; however, the specific patching strategy that we adopted complicates the reuse of previously captured images, thereby requiring a recapture of the center focus.



(a) The search algorithm for local peak contrast

(b) Example focus search images

(c) Search results at increasing steps

Figure 3. Contrast-based autofocusing pipeline. **(a)** The objective is to identify a per-superpixel focus (or depth) setting that maximizes contrast. At step K, we capture three measurements and identify the focus setting that maximizes contrast. We then refine the search around the next best candidate focus. **(b)** Example focus searches from the first two iterations. **(c)** Focus (or depth) maps and corresponding images captured after K iterations.

Patching strategy. A key challenge in the technique described above is the definition of the local region or patch where the search technique maximizes contrast. In particular, we would like each patch to have minimal depth variations and especially avoid depth discontinuities, so that a single focus setting suffices for all pixels in the patch. We achieve this by performing superpixel segmentation; since depth edges invariably align with texture edges, this strategy of patching based on superpixels allows us to avoid having large depth variations or discontinuities inside a superpixel. We compute the superpixels using the k-means clustering-based SLIC method [3]. We update the superpixels after each autofocus iteration to benefit from the improved sharpness in the captured photographs, which in turn allows for finer segmentation of depth boundaries. When the updated superpixel contains pixels previously assigned to different depths, we assign the new depth to the most common (contrast-maximizing) depth in the superpixel.

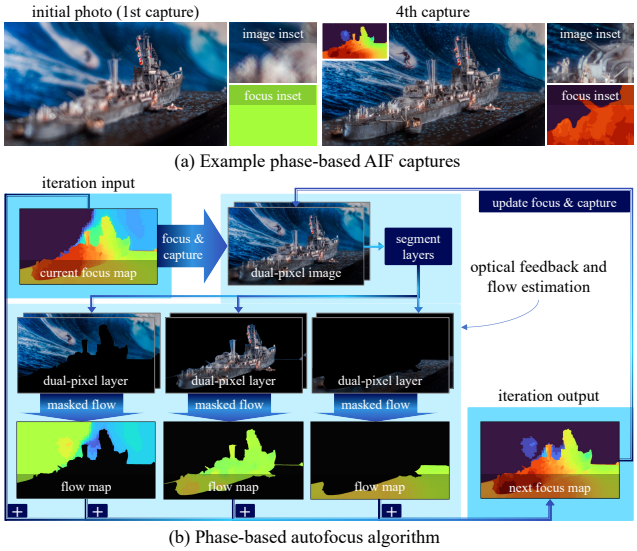
We illustrate our spatially-varying CDAF algorithm in Fig. 3. We also provide pseudocode in the supplemental pdf. Contrast-based autofocusing often requires multiple captures to identify the focus plane with the highest contrast. After establishing focus, one can then refine the existing focal surface over time to accommodate dynamic scenes, which is far less expensive than recapturing a full focal stack. In contrast, the technique we present next, PDAF, offers a single-shot approach to determining focus, at the cost of requiring a specialized sensor that uses dual pixel autofocus technology.

4.2. Phase Detection Autofocus (PDAF)

Recent advancements in DP sensors enabled Phase Detection Autofocus (PDAF) in consumer cameras and smartphones. A DP sensor consists of an array of microlenses, with two (or more) photodiodes under each lens. These photodiodes capture the light traveling through different parts of the aperture, producing a stereo image pair. The disparity observed across the two images determines the (signed) distance of the scene point from the focal plane (see Fig. 14 in supplemental pdf). When a scene point is sharply focused on the sensor, there is no disparity. When a scene point is out of focus, *i.e.*, focused onto a plane in front of or behind the sensor, the disparity appears directionally proportional to the direction and depth of the focused plane. To solve for the signed disparity, we adapt the conjugate gradient-based optical flow solver by Liu [20]. From the magnitude and direction of optical flow, we can calculate the focus correction needed to bring each region into focus, thereby enabling a single-shot approach to autofocus the entire field of view.

Challenges at depth boundaries. Since disparity in a stereo configuration is always horizontal, this suggests that much of our disparity estimation is driven by pixels with strong vertical gradients. However, in the context of focusing, we would also need to associate this disparity to the two regions on either side of the vertical gradient. As an example, suppose that we observe an edge that exhibits a disparity of a few pixels. When this edge corresponds to a depth boundary (which often have strong texture gradients), all we can observe is the relative shifts between the patches on either side of the boundary, and we cannot—just with local context—identify the exact shifts for each. To resolve this, we resort to segmenting the scene into semantically meaningful regions, and computing disparity for each region in isolation. In particular, we segment the total DP image using SegmentAnything [14] into masked labeled regions (layers), independently compute the masked optical flow for each layer, and sum all layers into a result flow map. The result flow map is used to drive the next autofocus. We illustrate this layered optical flow algorithm in Fig. 4 and detail the pseudocode in the supplemental pdf.

PDAF has many desirable properties. Unlike CDAF, which requires a search procedure to identify the sharpest focus setting, PDAF requires a single image to identify the spatially-varying focus map; this allows it to adapt to scene dynamics and is less likely to get stuck in local minima. When the working range is large, we can expect significant defocus in regions with larger disparity, which degrades the quality of the disparity estimate; even in this case, as long as the sign of the disparity is accurate, the technique converges within a few steps. All of these benefits come with the requirement of a specialized sensor that has dual pixels.



(a) Example phase-based AIF captures

(b) Phase-based autofocus algorithm

Figure 4. Phase-based autofocusing pipeline. **(a)** Example 1st and 4th captures from phase-based autofocus. **(b)** The algorithm starts by capturing an initial DP image, and segments the result into semantically meaningful layers. For each layer, we use optical flow to compute a smooth flow map, that indicates how to adjust the focus settings. After updating focus, we capture another DP image and repeat the process.

5. Results

Prototype. We show our prototype in Fig. 5. The Point Spread Function (PSF) of the prototype, shown in Fig. 6, is captured for a dot placed at varying distances spanning the entire tilting range we use of the SLM. To obtain DP images, we use the Canon R10 camera sensor with a 6000×4000 resolution and a $3.72\,\mu\text{m}$ pixel pitch. The sensor is a dual-pixel sensor which we use for our spatially-varying PDAF algorithm. We use gphoto2 commands to programmatically acquire raw images and the LibRaw library to extract the DP views from the raw image. To perform spatially-varying autofocusing, we use the Holoeye GAEA-2 Phase-Only SLM for phase modulation, which has a pixel pitch of $\delta_{\text{SLM}} = 3.74\,\mu\text{m}$ and a resolution of 4160×2464. The cubic phase plate is custom fabricated using subtractive laser etching. The relay lenses consist of three Samyang $85\,\text{mm}$ f/1.4 AS lenses. Our objective lens is a AF-S DX Micro NIKKOR 40mm f/2.8G.

Freeform depth-of-field photography. The ability to spatially-vary focus opens up a lot of composition capabilities beyond all-in-focus photography. After obtaining a depth map, our system can also intentionally add freeform defocus to the scene. For example, we can perform tilt-shift photography without requiring a Scheimpflug adapter to tilt the plane of focus, as seen in Fig. 7. We also show in Fig. 7 the ability to selectively focus on isolated regions at different depths, while the rest of the scene remains defocused.

Figure 5. Our prototype camera (see supplemental pdf for details).

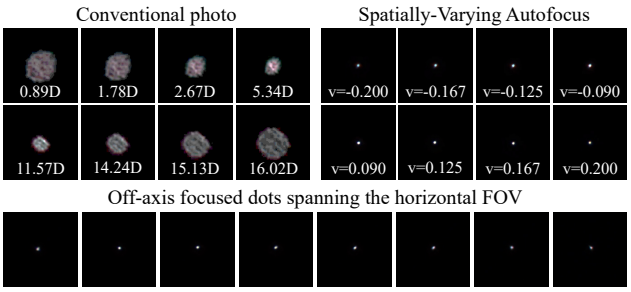


Figure 6. Point-spread functions of our prototype camera. **Left:** Images of a point placed at different distances from the camera, created by puncturing black paper with a needle and back-illuminating it with full-spectrum white light. The dot is 10-by-10 pixels, or 37.2 µm-by-37.2 µm on the sensor. We label the distance of each dot in diopters. **Right:** Focused images of the dot, captured with an SLM spatial frequency v. **Bottom:** Focused images of multiple dots at $v = 0.09$ spanning the camera's FOV (see supplement for zoom-ins).

Another application of our freeform depth-of-field selection can be seen in Fig. 8, where we suppress the presence of thin structures from a photo by optically defocusing only those particular pixels. We achieve this by selecting a focus that is far from its depth.

All-in-focus (AIF) imaging. We capture AIF results for six scenes: *Adventure*, *Planes*, *Flowers*, *Bunny*, *Ship*, and *Rainbow*. The full results gallery is in the supplemental pdf.

For *Adventure* and *Rainbow*, we provide a quantitative analysis of AIF methods in Fig. 11 using PSNR and SSIM [39] as metrics. To obtain the ground truth target, we capture a dense focal stack with 69 depth planes, with each focus setting having 2 repeated captures averaged to reduce noise. We then combine these images to form a computational AIF image from this focal stack to serve as our ground truth. Given the recovered depth map, we also program our lens to capture an AIF image optically, shown in our qualitative comparisons. When evaluating all-in-focus results with respect to number of photos captured, we observe phase-based autofocusing having the best performance, followed by both contrast-based autofocus and capturing a focal stack. For focal sweep, we search for the regularization parameter that produces the best result.

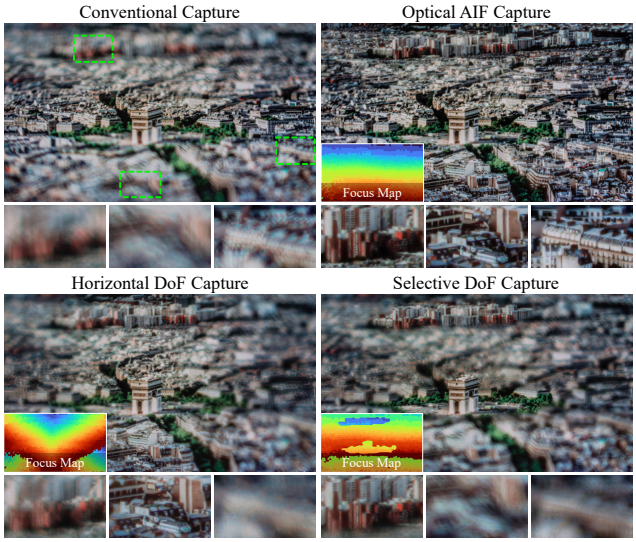


Figure 7. Example freeform depth-of-field captures of Arc de Triomphe displayed on a vertically tilted OLED. We show the capability of capturing an optical AIF image (top right), Scheimpflug-focusing to scale the defocus horizontally (bottom left), and selective focusing where a user specifies select regions to be in focus while all other regions are defocused (bottom right).

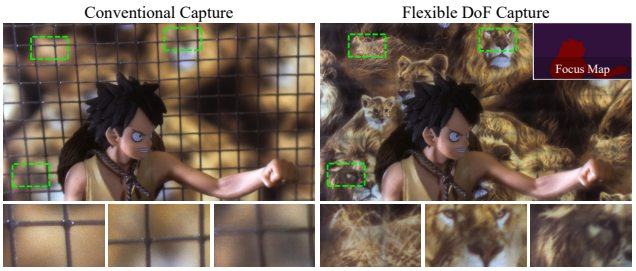


Figure 8. A thin structure removal example using freeform depth-of-field. The character stands in front of a thin wire mesh, far away from the lions print background. **Left:** A conventional capture shows the character in focus but with a visible wire mesh over the background. **Right:** Our proposed prototype optically removes the wire mesh by focusing on the background. The large defocus blur of the wire mesh, therefore, makes it hardly visible.

For all scenes, we show qualitative results for contrast-based after 3 steps (10 photos) and phase-based after 3 steps (4 photos). We also perform focal sweep (1 photo) and reconstruct AIF images via focus stacking both computationally and optically (20 photos). For the *Planes* scene, we also capture a small aperture photo with a separate camera and a long exposure ($f/36$ with a 55 mm lens on Nikon Z5).

Spatial resolution. We assess the spatial resolution of our imaging prototype, and the difference in performance across autofocus methods. Fig. 12 compares the modulation transfer function (MTF) of our spatially-varying autofocus methods to focal stack techniques and focal sweep. Both of our phase- and contrast-based autofocus methods show

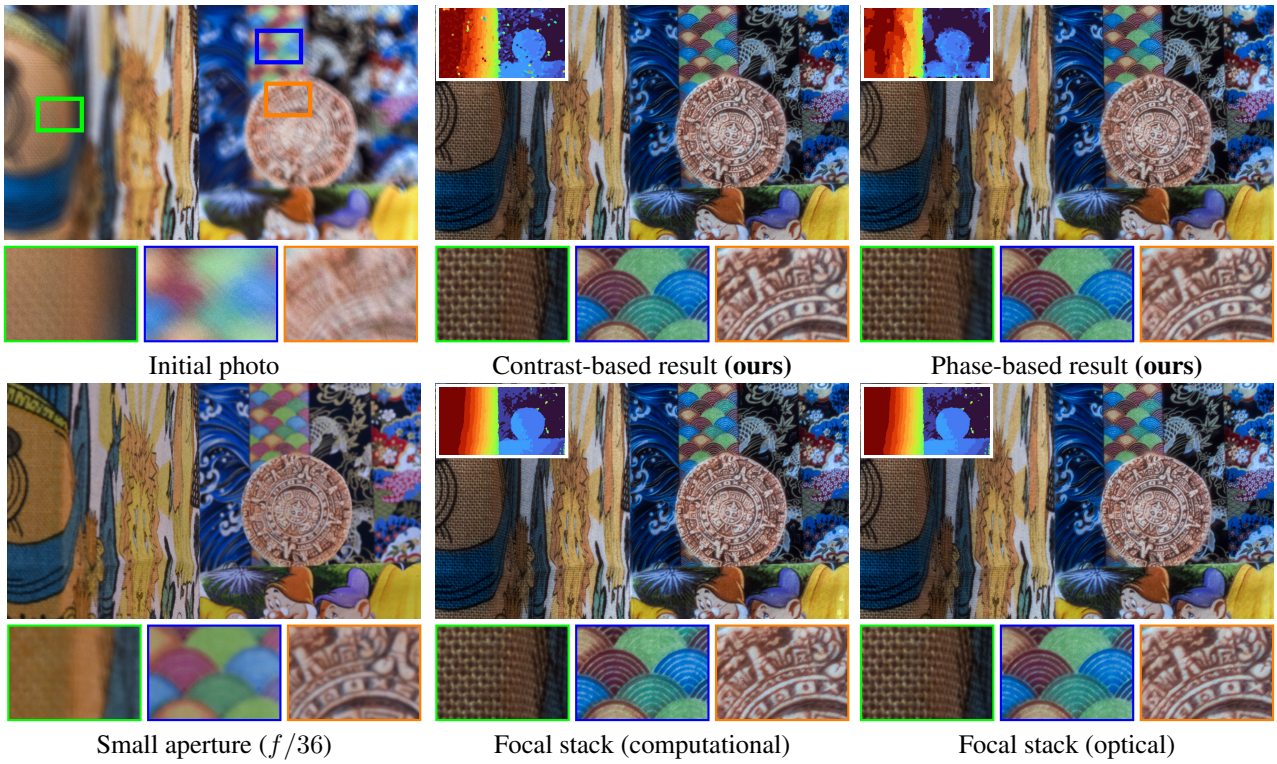


Figure 9. Qualitative comparison for the *Planes* scene. Each recovered depth map is shown in the top-left corner. All insets are blurrier for small aperture ($f/36$) when compared to our results due to diffraction (orange inset) and extreme depth range (green and blue insets).

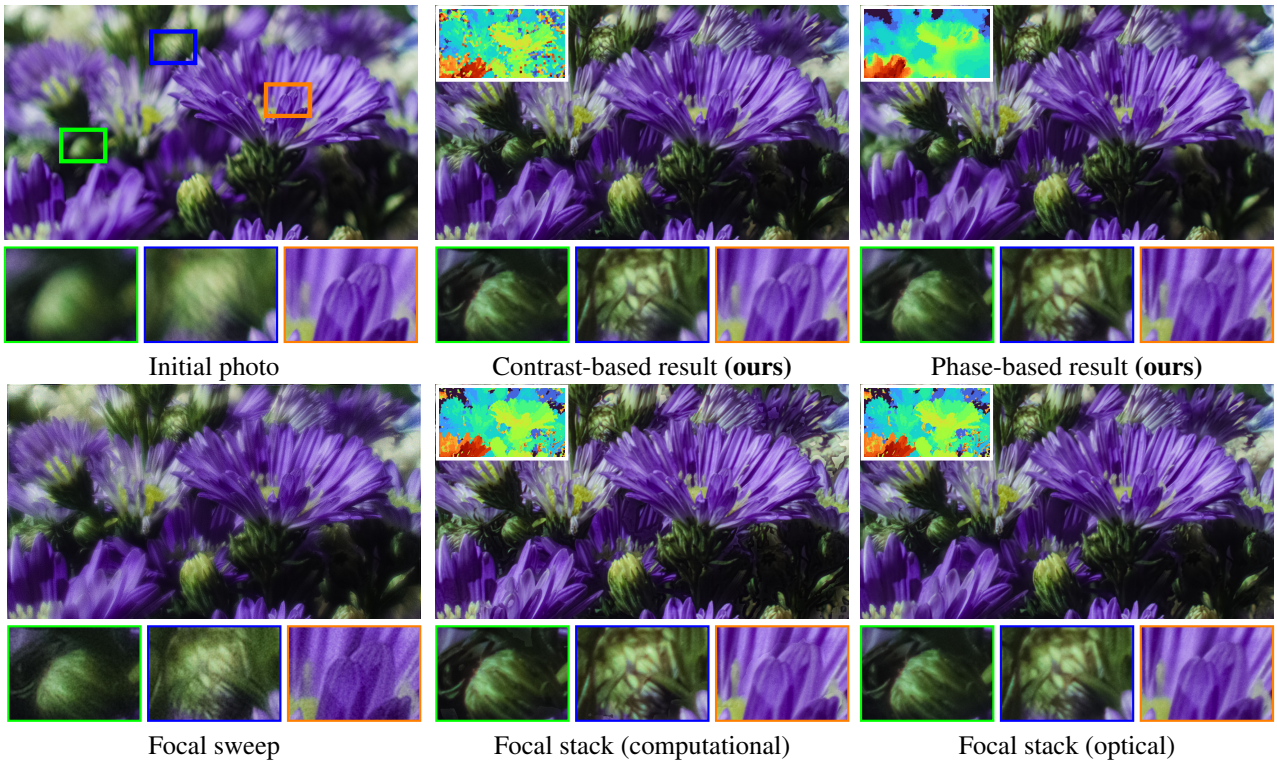


Figure 10. Qualitative comparison for the *Flowers* scene. We include a focal sweep photo (does not output depth) for additional comparison.

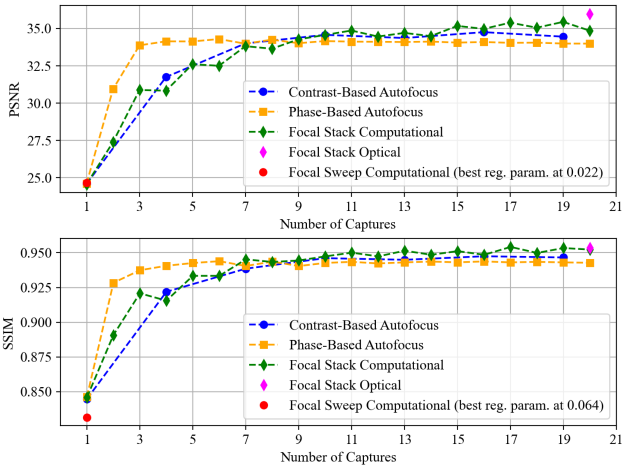


Figure 11. Quantitative analysis of all-in-focus imaging methods, evaluated as a function of number of photos. These plots represent the average performance over the *Adventure* and *Rainbow* scenes.

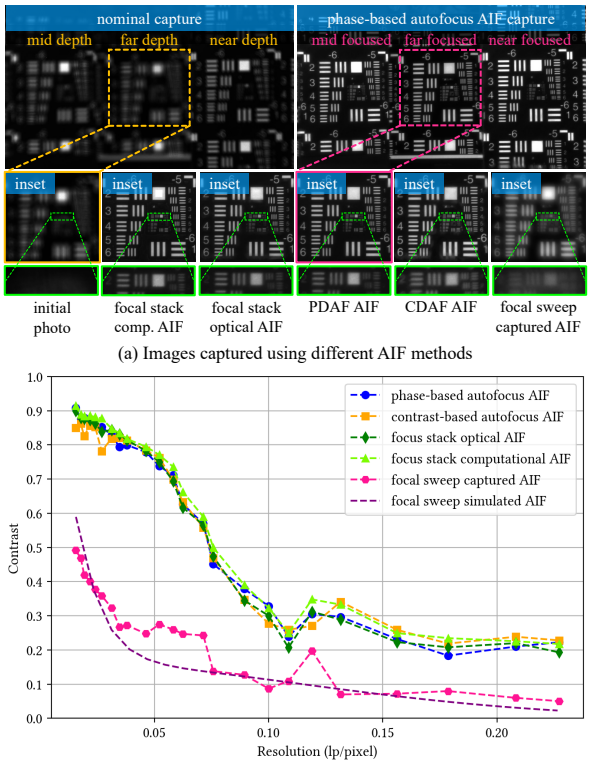
comparable performance with focal stack techniques, while focal sweep has consistently lower performance.

6. Discussion

Aperture. We estimate that our prototype imaging system has a rest state f number of $f/6.8$ (*i.e.*, the slope of the phase ramp is 0). When a phase ramp is applied to the SLM, the cubic phase plates are optically translated relative to one another, resulting in a smaller effective aperture. In other words, as our system moves the focus away from the nominal plane, this changes the shape of the effective aperture (impacting the bokeh) and reduces its area (limiting the total amount of light transmitted through the system). At the most extreme focus settings, we calculate that light throughput is 76% when compared to the system in the rest state. See the supplemental materials for additional details.

It is important to note that the f-number depends on both the diameter and curvature of the cubic phase plate, among other optics. In principle, we can increase the light throughput by using a cubic phase plate with a large diameter.

Aberration correction. As with the continual evolution of optical systems from singlets to achromatic doublets, apochromatic designs, and modern computer-optimized multi-element systems, the optical design of our proposed system can be improved to correct for aberrations. This includes the following: (i) redesign the lenses used in our system in an end-to-end fashion, and (ii) optimize the cubic phase plate profile. For the former, Sitzmann et al. [32] and Sun et al. [34] have shown that end-to-end optical system design allows aberrations to be fixed with a higher degree of freedom. Such methods suggest that lenses can be jointly optimized with parametric SLM patterns to improve all axes of performance of the imaging system: chromatic focal shift, artifacts at depth discontinuities, and f-number. For the latter, one can potentially optimize for a hybrid refractive-diffractive phase plate using the fully differentiable hybrid ray-tracing and wave-propagation (ray-wave) model recently proposed by Yang et al. [43], which significantly improved aberration correction.



(a) Images captured using different AIF methods

(b) MTF curves comparing different AIF methods

Figure 12. MTF of our prototype. **(a)** We arranged printed USAF targets at three different depths and illuminated them with full-spectrum white light, and computed AIF images for different methods. **(b)** Each MTF is computed and averaged from both horizontal and vertical line pairs across the three depths. We observe that our methods, including phase-based, contrast-based, and focus stack optical, perform consistently comparable to the focus stack computational AIF method, while focal sweep performs consistently worse than all of them.

7. Conclusion

This paper introduces a first-of-its-kind imaging technique, one that provides spatially-varying autofocusing capabilities. The proposed imaging system and the autofocusing algorithms can be interpreted as an optical optimizer of image contrast; we show that standard autofocusing techniques based on contrast and phase can be readily adopted to the spatially-varying setting. In general, we believe that this novel approach to imaging has widespread applications where focus is of paramount importance, such as long-range surveillance, machine vision, and microscopy.

Acknowledgements. This work was supported by the Air Force Office of Scientific Research (FA 95502410244), a NSF CAREER award (IIS 2238485), and a James Sprague Presidential Fellowship.

References

[1] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *ECCV*, 2020. 2

[2] Abdullah Abuolaim, Mauricio Delbracio, Damien Kelly, Michael S Brown, and Peyman Milanfar. Learning to reduce defocus blur by realistically modeling dual-pixel data. In *IEEE/CVF International Conference on Computer Vision*, 2021. 2

[3] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012. 4

[4] Edward H Adelson, James R Bergen, et al. The plenoptic function and the elements of early vision. *Computational models of visual processing*, 1(2):3–20, 1991. 2, 3

[5] Luis W Alvarez and William E Humphrey. Variable-power lens and system, 1970. US Patent 3,507,565. 3

[6] Terry A Bartlett, William C McDonald, and James N Hall. Adapting texas instruments dlp technology to demonstrate a phase spatial light modulator. In *Emerging Digital Micromirror Device Based Systems and Applications XI*, pages 161–173. SPIE, 2019. 14

[7] Julie Chang and Gordon Wetzstein. Deep optics for monocular depth estimation and 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10193–10202, 2019. 2

[8] Julie Chang, Isaac Kauvar, Xuemei Hu, and Gordon Wetzstein. Variable aperture light field photography: overcoming the diffraction-limited spatio-angular resolution tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3737–3745, 2016. 1

[9] Edward R Dowski and W Thomas Cathey. Extended depth of field through wave-front coding. *Applied optics*, 34(11): 1859–1866, 1995. 2

[10] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen. The lumigraph. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 453–464. 2023. 2, 3

[11] Adam Greengard, Yoav Y Schechner, and Rafael Piestun. Depth from diffracted rotation. *Optics letters*, 31(2):181–183, 2006. 2

[12] David E Jacobs, Jongmin Baek, and Marc Levoy. Focal stack compositing for depth of field control. *Stanford Computer Graphics Laboratory Technical Report*, 1(1):2012, 2012. 3

[13] Jinbeum Jang, Yoonjong Yoo, Jongheon Kim, and Joonki Paik. Sensor-based auto-focusing system using multi-scale feature extraction and phase correlation matching. *Sensors*, 15(3):5747–5762, 2015. 2

[14] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv:2304.02643*, 2023. 5, 14

[15] Eric P Krotkov. *Active computer vision by cooperative focus and stereo*. Springer Science & Business Media, 2012. 3

[16] Sujit Kuthirummal, Hajime Nagahara, Changyin Zhou, and Shree K Nayar. Flexible depth of field photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):58–71, 2010. 2, 3

[17] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman. Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26 (3):70–es, 2007. 2

[18] Anat Levin, Samuel W Hasinoff, Paul Green, Frédo Durand, and William T Freeman. 4d frequency analysis of computational cameras for depth of field extension. *ACM Transactions on Graphics (TOG)*, 28(3):1–14, 2009. 2

[19] Marc Levoy. Light fields and computational imaging. *Computer*, 39(8):46–55, 2006. 2, 3

[20] Ce Liu. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, USA, 2009. AAI0822221. 5, 14, 15

[21] Adolf W Lohmann. A new class of varifocal lenses. *Applied Optics*, 9(7):1669–1671, 1970. 1, 3

[22] Hajime Nagahara, Sujit Kuthirummal, Changyin Zhou, and Shree K Nayar. Flexible depth of field photography. In *European Conference on Computer Vision*, pages 60–73. Springer, 2008. 2, 3

[23] Shree K Nayar and Yasuo Nakagawa. Shape from focus. *IEEE Transactions on Pattern analysis and machine intelligence*, 16(8):824–831, 1994. 2, 3

[24] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. *Light field photography with a hand-held plenoptic camera*. PhD thesis, Stanford university, 2005. 2, 3

[25] Deepak Pathak, Ross Girshick, Piotr Dollár, Trevor Darrell, and Bharath Hariharan. Learning features by watching objects move. In *CVPR*, 2017. 14

[26] Sri Rama Prasanna Pavani and Rafael Piestun. Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system. *Optics express*, 16 (26):22048–22057, 2008. 2

[27] Sri Rama Prasanna Pavani, Michael A Thompson, Julie S Biteen, Samuel J Lord, Na Liu, Robert J Twieg, Rafael Piestun, and William E Moerner. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences*, 106(9):2995–2999, 2009. 2

[28] Yingsi Qin, Wei-Yu Chen, Matthew O'Toole, and Aswin C. Sankaranarayanan. Split-lohmann multifocal displays. *ACM Trans. Graph.*, 42(4), 2023. 2, 3, 11, 14

[29] Yingsi Qin, Aswin C. Sankaranarayanan, and Matthew O'Toole. Project page: Spatially-varying autofocus. `https://imaging.cs.cmu.edu/svaf`, 2025. 2

[30] Daniel J. Reiley. Nikon AF Nikkor 85mm f/1.4D IF Lens prescription design. `https://www.lens-designs.com/photographic-primes`, 2017. Licensed under the MIT License. 14

[31] Parikshit Sakurikar and P. J. Narayanan. Focal stack representation and focus manipulation. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 250–255, 2017. 3

[32] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph.*, 37(4), 2018. 8, 14

[33] Muralidhara Subbarao and J-K Tyan. Selecting the optimal focus measure for autofocusing and depth-from-focus. *IEEE transactions on pattern analysis and machine intelligence*, 20(8):864–870, 1998. 3

[34] Qilin Sun, Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. End-to-end complex lens design with differentiate ray tracing. *ACM Trans. Graph.*, 40(4), 2021. 8, 14

[35] Supasorn Suwajanakorn, Carlos Hernandez, and Steven M Seitz. Depth from focus with your mobile phone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3497–3506, 2015. 2, 3

[36] Dong-Chen Tsai and Homer H. Chen. Reciprocal focus profile. *IEEE Transactions on Image Processing*, 21(2):459–468, 2012. 3

[37] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3):69, 2007. 2

[38] Neal Wadhwa, Rahul Garg, David E Jacobs, Bryan E Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T Barron, Yael Pritch, and Marc Levoy. Synthetic depth-of-field with a single-camera mobile phone. *ACM Transactions on Graphics (ToG)*, 37(4):1–13, 2018. 2

[39] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 6

[40] Bennett Wilburn, Neel Joshi, Vaibhav Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, Andrew Adams, Mark Horowitz, and Marc Levoy. High performance imaging using large camera arrays. In *ACM SIGGRAPH 2005 Papers*, pages 765–776. 2005. 2, 3

[41] Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin C. Sankaranarayanan, and Ashok Veeraraghavan. Phasecam3d — Learning phase masks for passive single view depth estimation. In *IEEE Intl. Conf. Computational Photography (ICCP)*, 2019. 2

[42] Shumian Xin, Neal Wadhwa, Tianfan Xue, Jonathan T Barron, Pratul P Srinivasan, Jiawen Chen, Ioannis Gkioulekas, and Rahul Garg. Defocus map estimation and deblurring from a single dual-pixel image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2228–2238, 2021. 2

[43] Xinge Yang, Matheus Souza, Kunyi Wang, Praneeth Chakravarthula, Qiang Fu, and Wolfgang Heidrich. End-to-end hybrid refractive-diffractive lens design with differentiable ray-wave model. In *SIGGRAPH Asia 2024 Conference Papers*, New York, NY, USA, 2024. Association for Computing Machinery. 8, 14

[44] Changyin Zhou, Daniel Miau, and Shree K Nayar. Focal sweep camera for space-time refocusing. 2012. 3