

3D Imaging Using Extreme Dispersion in Optical Metasurfaces

Shiyu Tan, Frank Yang, Vivek Boominathan, Ashok Veeraraghavan, and Gururaj V. Naik*

Cite This: *ACS Photonics* 2021, 8, 1421–1429

Read Online

ACCESS |



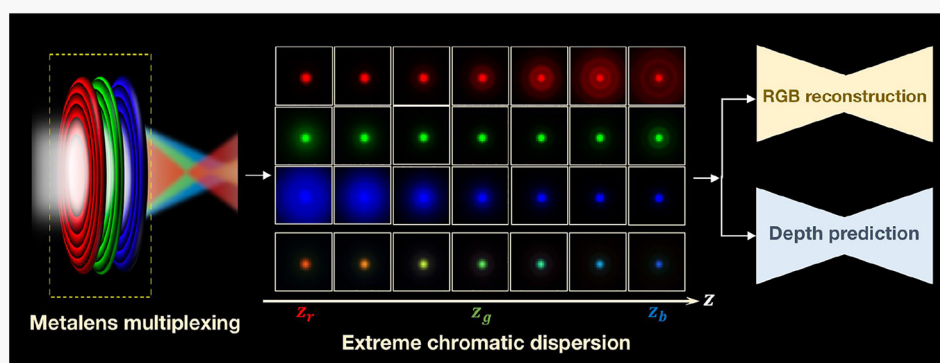
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: Metasurfaces have the potential to revolutionize imaging technologies due to their extreme control of phase, polarization, and amplitude of the incident light. They rely upon enhanced local interaction of light to achieve the desired phase profile. As a consequence of the enhanced local interaction of light, metasurfaces are highly dispersive. This strong dispersion has been recognized as a primary limitation as it relates to realizing conventional imaging with metasurfaces. Here, we argue that this strong dispersion is an added degree of design freedom for computational imaging, potentially opening up novel applications. In particular, we exploit this strongly dispersive property of metasurfaces to propose a compact, single-shot, and passive 3D imaging camera. Our device consists of a metalens engineered to focus different wavelengths at different depths and two deep networks to recover depth and RGB texture information from chromatic, defocused images acquired by the system. In contrast with other metasurface-based 3D sensors, our design can operate in the full visible range with a larger field-of-view (FOV) and can potentially generate dense depth maps of complicated 3D scenes. Our simulation results on a 1 mm diameter metalens demonstrate its ability to capture 3D depth and texture information ranging from 0.12 to 0.6 m.

KEYWORDS: metasurface, 3D reconstruction, depth from defocus, chromatic dispersion, convolutional neural network

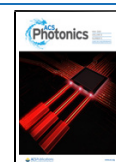
Metasurfaces have emerged as powerful substitutes to conventional diffractive optics.^{1–6} Similar to diffractive optical devices, metasurfaces are composed of two-dimensional arrays of optical scatterers (called meta-elements). However, their phase gradient does not arise from gradual phase changes via light propagation. Rather, phase discontinuities due to enhanced local interaction of light at each meta-element enable the desired phase gradient.^{3,7} This enhanced local light interaction modifies the phase of the incident light at a subwavelength scale, allowing for function multiplexing and increased design flexibility (as shown in Figure 1a). These unique aspects of metasurfaces make them compact and desirable for many functionalities, including 3D imaging.^{8–11}

Unfortunately, metasurfaces also lead to highly frequency-dependent phase discontinuities, resulting in much stronger chromatic dispersion than diffractive optics. For example, a metalens designed for 532 nm wavelength shows a dramatically different point spread function (PSF) at 612 and 452 nm wavelengths as shown in Figure 1b. This strong dispersion has been recognized as a primary limitation for imaging with

conventional techniques. Computational optics offers one pathway to overcome strong dispersion. In the recent first report of computational imaging with a metasurface, meta-optics and image postprocessing were combined to achieve direct imaging with white light in the full visible range.¹² A subsequent work extended these techniques to achieve simultaneous achromatic, varifocal imaging.¹³ Here, following these and other recent works,^{11,14–19} we argue that strong metasurface dispersion is an added degree of design freedom for computational imaging, potentially opening up novel applications. In particular, we exploit the strong dispersion and high design flexibility of metasurfaces by combining them with computational optics to demonstrate 3D imaging.

Received: January 20, 2021

Published: April 16, 2021



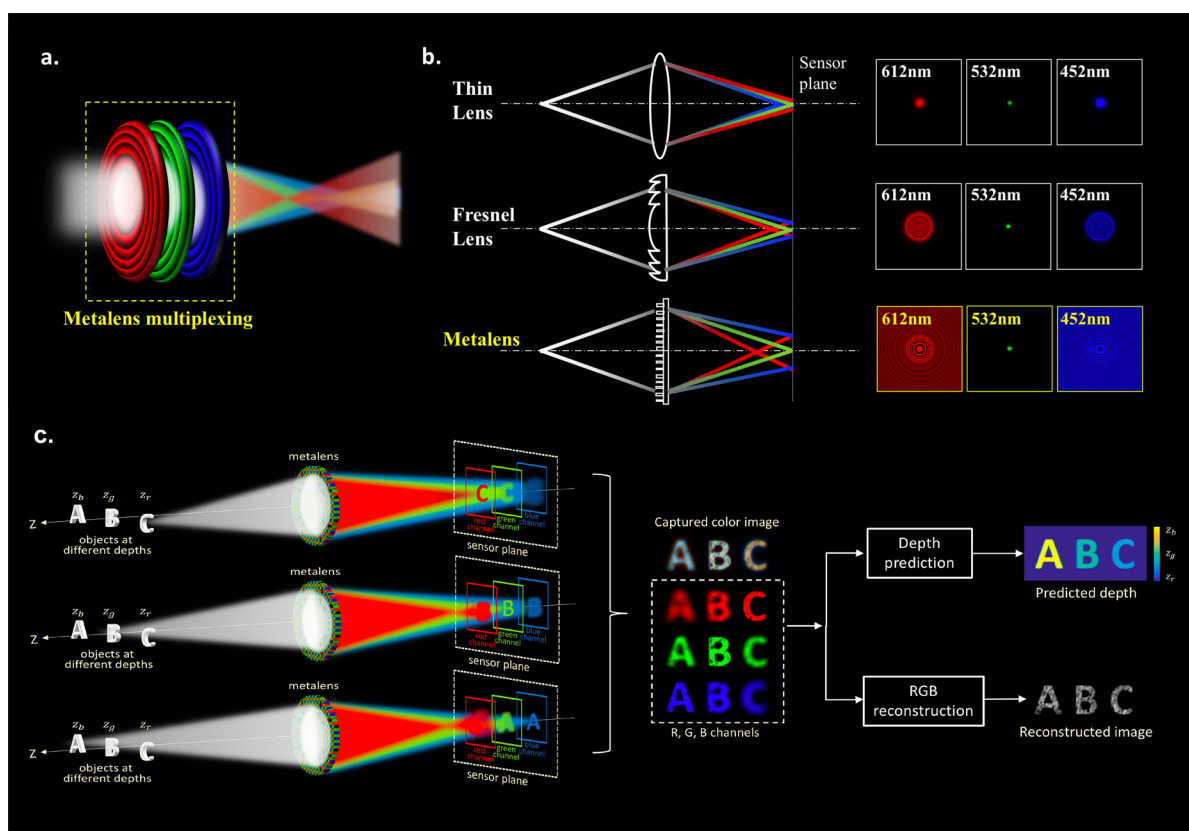


Figure 1. Overview: (a) The enhanced local light interaction of metalenses results in strong chromatic dispersion, a new design axis for computational imaging. Further, the localized light interaction in meta-elements enables multiplexing of functionalities, increasing design flexibility. (b) Focusing schematic and point spread functions (PSF) for a thin lens, Fresnel lens, and metalens (all designed for 532 nm) at different wavelengths. The metalens PSF shows a dramatic spread (when deviated from the design wavelength) due to stronger chromatic dispersion relative to thin lenses and Fresnel lenses. (c) We exploit the strong dispersion of a metalens to show a compact, single-shot, and passive 3D imaging system. Different wavelengths focus at different depths, encoding depth information over R (red), G (green), and B (blue) channels of captured color images. Two deep networks are used to recover depth and RGB texture information from the chromatic, defocused image.

Compared to typical 3D imaging techniques,^{20–26} metasurface-based 3D cameras can be compact, single-shot, and passive, without requiring active light sources or multiple viewpoints/exposures. They enable the miniaturization of 3D cameras and can be applied to mobile or wearable platforms with low cost and complexity. Prior metasurface-based 3D cameras are generally based on depth from defocus (DFD) techniques.^{8,9,11,27} In DFD, images are captured at different levels of defocus from which the depth information is recovered.^{28–33} In such cases, a single metasurface could consist of interleaved metalenses with different focal lengths.⁸ Alternatively, two adjacent metasurfaces, one with depth-dependent and the other with a complementary depth-independent response, could enable the recovery of depth information.^{9,11} Both these systems inherit the limitation of field-of-view (FOV) and are designed for single wavelength operation. Thus, their light efficiency is limited, which is disadvantageous for low-power or energy-constrained applications. Recently, metasurface lenslet based imaging techniques^{34,35} have been proposed for full-color imaging and depth estimation. However, such implementations require an extra main lens (together with the metalens array), making them more complicated in design and less compact than DFD-based 3D cameras.^{8,9,11}

In this article, we propose a metasurface-based 3D imaging system that, for the first time, operates in the full visible

spectrum (380–700 nm). An overview of our system is shown in Figure 1c. Inspired by previous work,³⁶ we exploit the fact that the reflectance spectra of most real-world objects are sufficiently broadband³⁷ to simultaneously acquire two-dimensional images of texture and depth. Our system consists of an optical metasurface that focuses red, green, and blue (RGB) light with different central wavelengths (460, 530, and 620 nm) at different depths. A sensor located at a fixed distance from the metasurface captures the RGB image in one shot. Therefore, the three channels of the captured image are defocused differently and are sensitive to changes in object depth. We leverage chromatic differences in the captured image using two U-Net^{30,38,40} based convolutional neural networks (CNN) to create depth maps and sharp RGB image reconstructions. The deep networks implicitly learn 3D information from spatial blur and prior statistics and can produce accurate dense reconstructions.

METASURFACE DESIGN

Our metasurface is made of three multiplexed metalenses having different focal lengths f_c for each target wavelength, as shown in Figure 2a. $c \in \{r, g, b\}$ denotes the target channel with the central wavelengths $\lambda_c = \{620 \text{ nm}, 530 \text{ nm}, 460 \text{ nm}\}$, respectively.

The metalens phase profile for each color channel is given below:

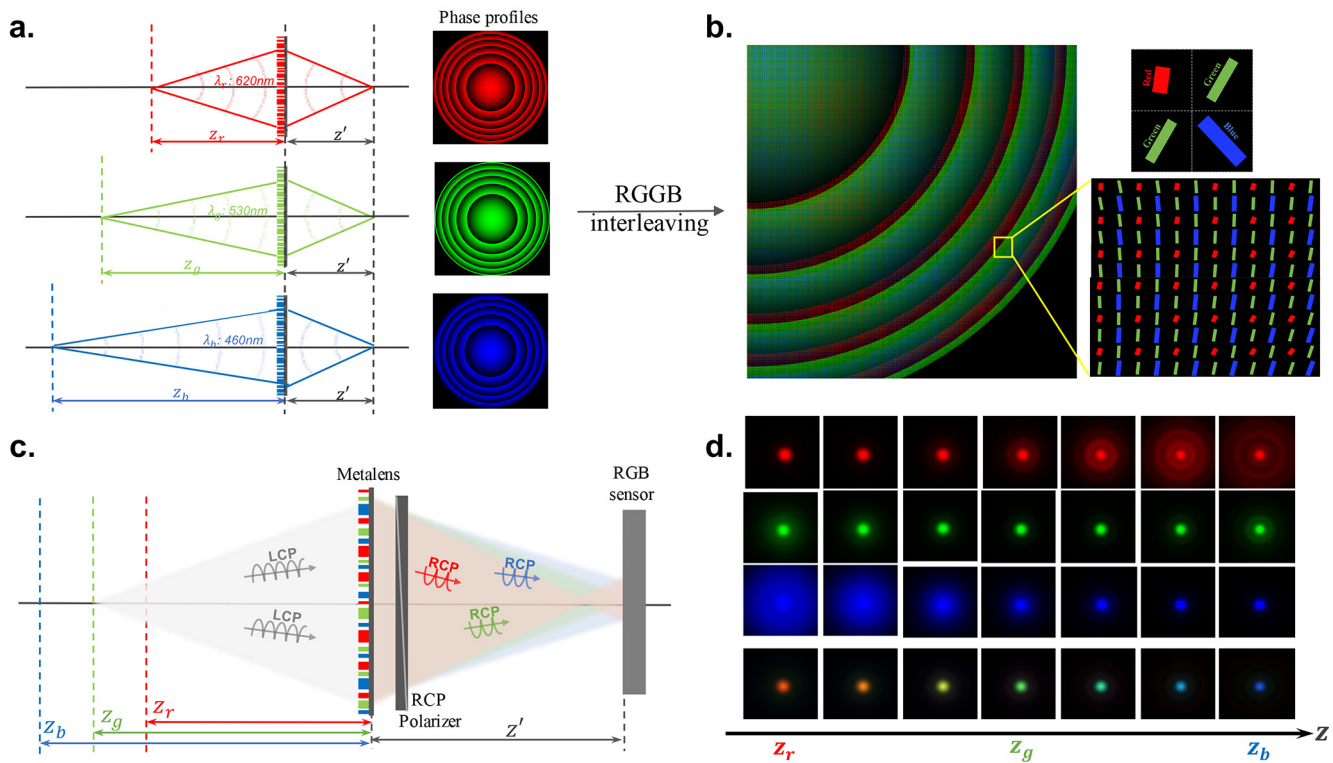


Figure 2. Metalens design: (a) Design of RGB phase profiles with different focal lengths. (b) RGGB spatial interleaving. Nanofins corresponding to the three RGB phase profiles are colored red, green, and blue. (c) Multiplexed metalens focuses different colors at different depths on the sensor. (d) Simulated point spread functions (PSF) for different channels at different depths (PSFs are normalized for visualization). We use the chromatic dispersion shown in the PSFs to estimate accurate depth maps of a 3D scene.

$$\phi_c(x, y, \lambda_c) = \frac{2\pi}{\lambda_c}(\sqrt{x^2 + y^2 + f_c^2} - f_c) \quad (1)$$

where (x, y) is the spatial coordinate on the metalens. We spatially interleave the meta-elements of the three individual metalenses corresponding to red, green, and blue channels. Spatially interleaving the three profiles $\phi_r(x, y, \lambda_r)$, $\phi_g(x, y, \lambda_g)$, and $\phi_b(x, y, \lambda_b)$ results in the overall phase profile $\Phi(x, y)$, shown in Figure 2b.

The phase profile above may be implemented by many nanophotonic designs.⁴¹ For simplicity, we design a metasurface that uses geometric phases to achieve the desired phase profile.¹⁷ Our metasurface takes input left-hand circularly polarized light and outputs right-hand circularly polarized light, as shown in Figure 2c. The wavelength and depth-dependent point spread functions of this metasurface are shown in Figure 2d.

To design the individual meta-elements, we employ full-wave finite-difference time-domain (FDTD) simulations. The basic metasurface unit cell contains a titanium dioxide nanofin on a silica substrate, as shown in Figure 3a. By varying the geometric parameters of the nanofins, we engineer the phase jump for individual unit cells. Optical constants for silica⁴² and titanium dioxide⁴³ are taken from literature, and the background index is 1. We control the phase discontinuity of each meta-element using its geometric phase. In this configuration, each nanofin functions as a half-wave plate, transforming the helicity of incident circularly polarized light. By rotating the nanofins by angle θ , as shown in Figure 3a, incident circularly polarized light is transformed to the opposite helicity and imparted with a phase of $\varphi = 2\theta$. Using this method, we can

achieve the full 2π phase range needed for metasurface operation.

Because we seek to engineer the focal length of the metasurface lens for red, green, and blue light, nanofins are designed to maximize efficiency at target wavelengths while suppressing transmission at other wavelengths. We engineer the nanofins to maximize polarization conversion efficiency, which is defined as the transmitted power of light of the output helicity divided by the total incident power, given below (assuming LCP incident light).

$$\text{PCE} = \frac{\text{transmitted power in RCP}}{\text{total incident power (LCP)}} \times 100 \quad (2)$$

Optimized nanofin parameters are $L = 160\text{ nm}$, $W = 90\text{ nm}$ for the blue channel; $L = 290\text{ nm}$, $W = 60\text{ nm}$ for the green channel; and $L = 380\text{ nm}$, $W = 90\text{ nm}$ for the red channel. In all cases, $H = 600\text{ nm}$ and $P = 400\text{ nm}$. This geometry results in maximum PCE at target wavelengths for each channel, as shown in Figure 3b. Phase and PCE are plotted against nanofin rotation angle for each target wavelength in Figure 3c. These parameters can be used to implement designed phase profiles and simulate system PSFs.

3D RECONSTRUCTION

Our system focuses the red, green, and blue light at different depths, enabling the depth prediction and texture reconstruction of the captured RGB image of a 3D scene. We make use of two separate deep networks for the 3D reconstruction. The pipeline of the networks' training procedures is shown in Figure 4. Conceptually there are three steps involved: (a) PSF simulation, the system PSF is simulated at 21 discrete samples

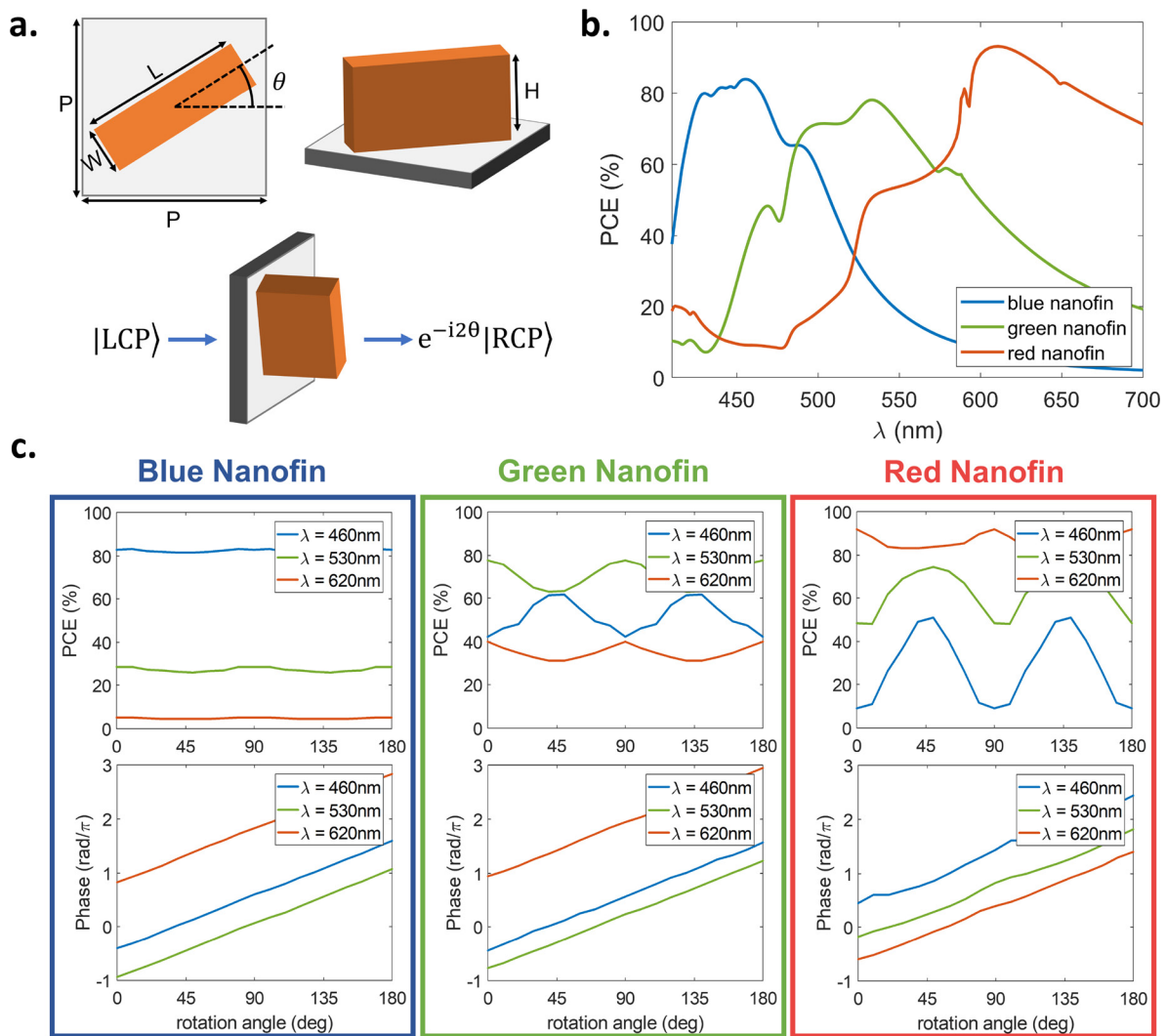


Figure 3. Nanofin design: (a) Simulation schematic of metasurface unit cell containing a titanium dioxide nanofin on a silica substrate. Scatterers convert left circularly polarized (LCP) light to right circularly polarized (RCP) light with an additional geometric phase of 2θ . (b) Polarization conversion efficiency of nanofin designs optimized for target wavelengths (blue, 460 nm; green, 530 nm; red, 620 nm). Optimized nanofins for the blue, green, and red channels are $160 \text{ nm} \times 90 \text{ nm}$, $290 \text{ nm} \times 60 \text{ nm}$, and $380 \text{ nm} \times 90 \text{ nm}$, respectively. Nanofin height and periodicity are 600 and 400 nm, respectively. (c) PCE and phase vs rotation angle θ for optimal nanofin geometries at target wavelengths of 460, 530, and 620 nm.

in the depth ranging from z_r to z_b for RGB channels (using parameters obtained in full-wave simulations); (b) Image rendering, the image captured by the sensor is rendered as the convolution of the all-in-focus image and the depth and color dependent PSF; (c) Depth and RGB reconstruction, the depth map and RGB image are reconstructed from the captured image using U-Net based deep networks.

PSF Simulation. We first simulate the PSFs at different depths based on Fourier optics.⁴⁴ The metasurface nanofins spatially modulate the wavefront of the incident light at each wavelength. This spatial modulation is described by $T_\lambda(x, y)$, defined below:

$$T_\lambda(x, y) = A(x, y) \sqrt{\text{PCE}_\lambda(x, y)} \exp(j\Phi_\lambda(x, y)) \quad (3)$$

Here, $A(x, y)$ is a circular aperture function, which is 1 within the diameter of the metasurface and 0 elsewhere. $\sqrt{\text{PCE}_\lambda(x, y)}$ and $\Phi_\lambda(x, y)$ are the changes to amplitude (square root of polarization conversion efficiency) and phase induced by the meta-element located at (x, y) . They are obtained from full-

wave simulation. The transmission efficiency and phase shift differ for different wavelengths due to metasurface dispersion.

The system PSF $P_{\lambda,z}$ is the response to a point source at wavelength λ and distance z from the metasurface plane. We derive our PSF using light field propagation methods under the Fresnel approximation, assuming $\lambda \ll z$.

$$P_{\lambda,z}(x', y') \propto \left| \mathcal{F} \left\{ A(x, y) \sqrt{\text{PCE}_\lambda(x, y)} \exp \left\{ j \left[\frac{\pi}{\lambda} \left(\frac{1}{z} + \frac{1}{z'} \right) (x^2 + y^2) + \Phi_\lambda(x, y) \right] \right\} \right\} \right|^2 \quad (4)$$

where z' is the distance between metalens and the sensor, and (x', y') is the coordinate on the sensor plane. Next, the spectral response of a color sensor $\kappa_{c,\lambda}$ is adopted to convert computed spectral PSFs to color PSFs.

$$P_{c,z}(x', y') = \sum_{\lambda} \kappa_{c,\lambda} P_{\lambda,z}(x', y'), \quad c \in \{r, g, b\} \quad (5)$$

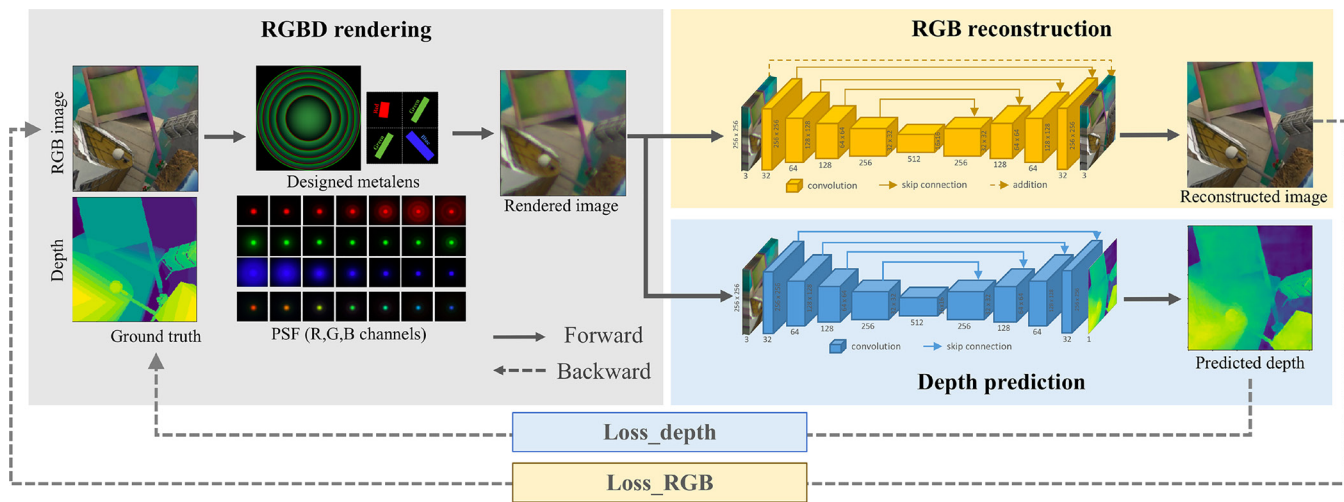


Figure 4. 3D reconstruction pipeline. We use separate U-Net based reconstruction networks to recover depth and texture information on a 3D scene from the captured image. The reconstruction networks are trained beforehand with simulated captured images. During training, the depth-dependent PSF is first simulated for different wavelengths, given the phase and PCE profiles of the designed nanofins. The captured color image is then rendered using all-in-focus texture and depth map as inputs (by convolving the PSFs with the layered textures to account for the depth-dependent defocus effect). The two networks take the rendered image as input and estimate the depth map and RGB image. The loss of reconstructed depth and texture are backpropagated to update the network parameters.

A simulation of color PSFs at different depths is shown in Figure 4.

RGBD Rendering. The captured images are then rendered given the simulated PSFs, all-in-focus RGB images, and depth maps. The all-in-focus images refer to the image of 3D scenes captured by a pinhole camera and are considered as the RGB image ground truth. Similar to PSF-based image formation in DOE,^{30,40,45} we treat the 3D scene as a set of segmented layers as different depths and convolve each layer $I_{c,z}$ with the corresponding PSF $P_{c,z}$. The rendered image I'_c for each channel is computed as

$$I'_c = \sum_z I_{c,z} * P_{c,z} + \text{noise} \quad (6)$$

where $*$ denotes a convolution operator. To account for the effect of noise, we apply additive Gaussian noise with standard deviation based on the aperture size, light level, and exposure duration.

Reconstruction Networks. We use two separate networks to reconstruct the depth and RGB texture images. The depth prediction network, as shown in Figure 4, has an encoder-decoder architecture that can predict pixel-size depth maps.³⁰ It takes as input the three-channel captured RGB image and outputs the one-channel depth map with the same resolution. The texture reconstruction network is based on a modified residual U-Net^{38,39} in which the differences between the coded image and the ground truth image (i.e., residual image, which encourage high-frequency information recovery) are learned. Such residual learning techniques have been shown to have good performance in all-in-focus image reconstruction (details of the network architectures are provided in the Supporting Information).

During training, the loss of depth prediction and RGB texture reconstruction are back-propagated to update the network parameters. We enforce the root mean squared error for both estimated RGB images $\hat{I}$ and predicted depth $\hat{z}$. A gradient-based regularization loss is used for depth estimation to encourage sharper boundaries.³⁰

$$\text{loss_depth} = \frac{1}{\sqrt{N}} (\|z - \hat{z}\|_2 + \alpha \|\nabla \hat{z}\|_2) \quad (7)$$

$$\text{loss_RGB} = \frac{1}{\sqrt{M}} \|I - \hat{I}\|_2 \quad (8)$$

Here, ∇ denotes the spatial gradient operator, α is the weight of gradient loss, and M and N are the number of pixels in the RGB image and disparity map.

SIMULATION AND RESULTS

To demonstrate our method, we design a metalens-based 3D camera for the 0.12–0.6 m depth range and simulate its 3D reconstruction performance. In our design, the metalens focuses at $z_r = 0.12$ m, $z_g = 0.2$ m, and $z_b = 0.6$ m for the R (620 nm), G (540 nm), and B (460 nm) channels, respectively. The sensor is located 57.2 mm away from the metalens, and the focal lengths are $f_r = 38.8$ mm, $f_g = 44.5$ mm, and $f_b = 52.2$ mm. The overall phase profile is formed by multiplexing the three RGB phase profiles in the RGGGB configuration, with a 1 mm² circular aperture. The simulated sensor pixel size is 9.6 μm .

Implementation Details. During training, the PSFs are simulated over 21 discrete depths linearly sampled in the disparity (i.e., inverse depth) space and 17 wavelengths within the visible range (380 nm–700 nm in 20 nm intervals). The spectral response of the color Sony IMX183 CMOS sensor is used to convert spectral PSFs to color PSFs, as given in eq 5. In image modulation, we apply Gaussian noise with a standard deviation of $\sigma = 0.01$. The reconstruction networks are trained end-to-end on the SceneFlow data set,⁴⁶ a synthetic data set consisting of dense ground truth disparity maps (enabling our RGBD rendering) for 35, 454 training, and 4370 testing images. We use the “cleanpass” subset with only all-in-focus images (randomly cropped into a size of 256 $\times$ 256) to simulate the captured images for our camera settings. The network parameters are optimized using Adam optimizer⁴⁷ ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a batch size of 21 for 50 epochs, on

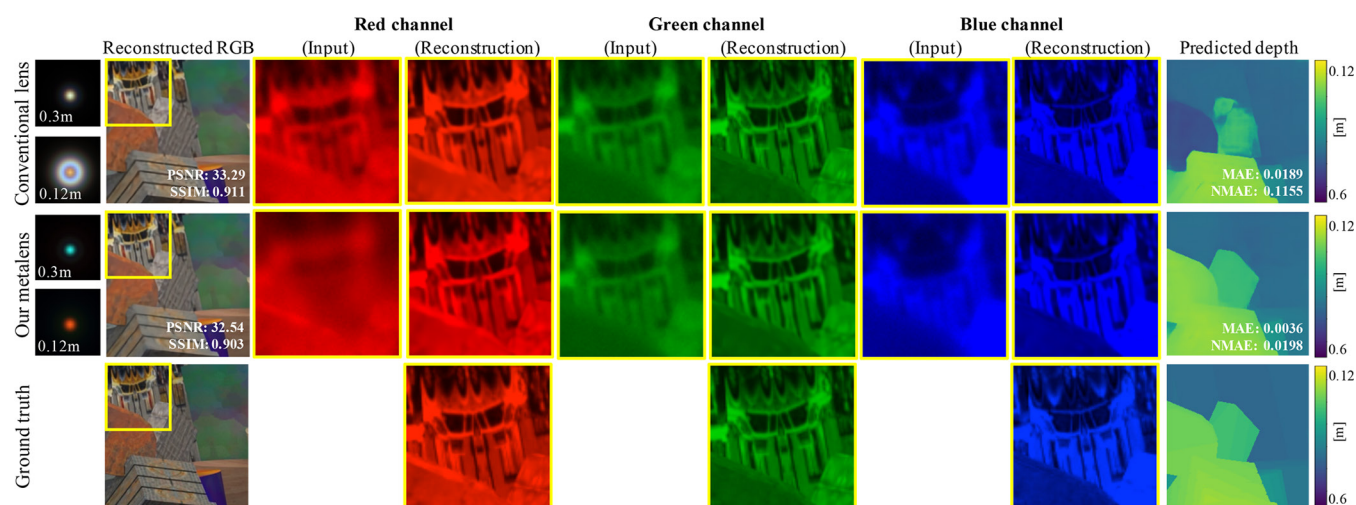


Figure 5. Comparison with a conventional lens: The conventional lens is designed to have the same focal length (44.5 mm) and focus at the same depth (0.2 m) as our metalens (for green light centered at 540 nm). In particular, it is designed to have a quadratic height profile $h(x, y)(n_g - 1) = \left(\sqrt{x^2 + y^2 + f_g^2} - f_g \right)$ (without phase wrapping), where $f_g = 44.5$ mm is the focal length and $n_g = 1.52$ is the refractive index at 540 nm. The red and blue light is slightly dispersed due to the change of central wavelengths and refractive indices, but will be overall focused around 0.2 m. As a comparison, our designed metalens (with strong chromatic dispersion and subwavelength scale) can be multiplexed to focus at very different depths for R, G, and B channels. Shown above are 3D reconstruction comparisons between our metalens and the conventional lens. The depth maps and RGB textures are reconstructed by the same network architectures. Our metalens outperforms the conventional lens in depth prediction with higher accuracy and maintains a similar performance in the RGB texture reconstruction.

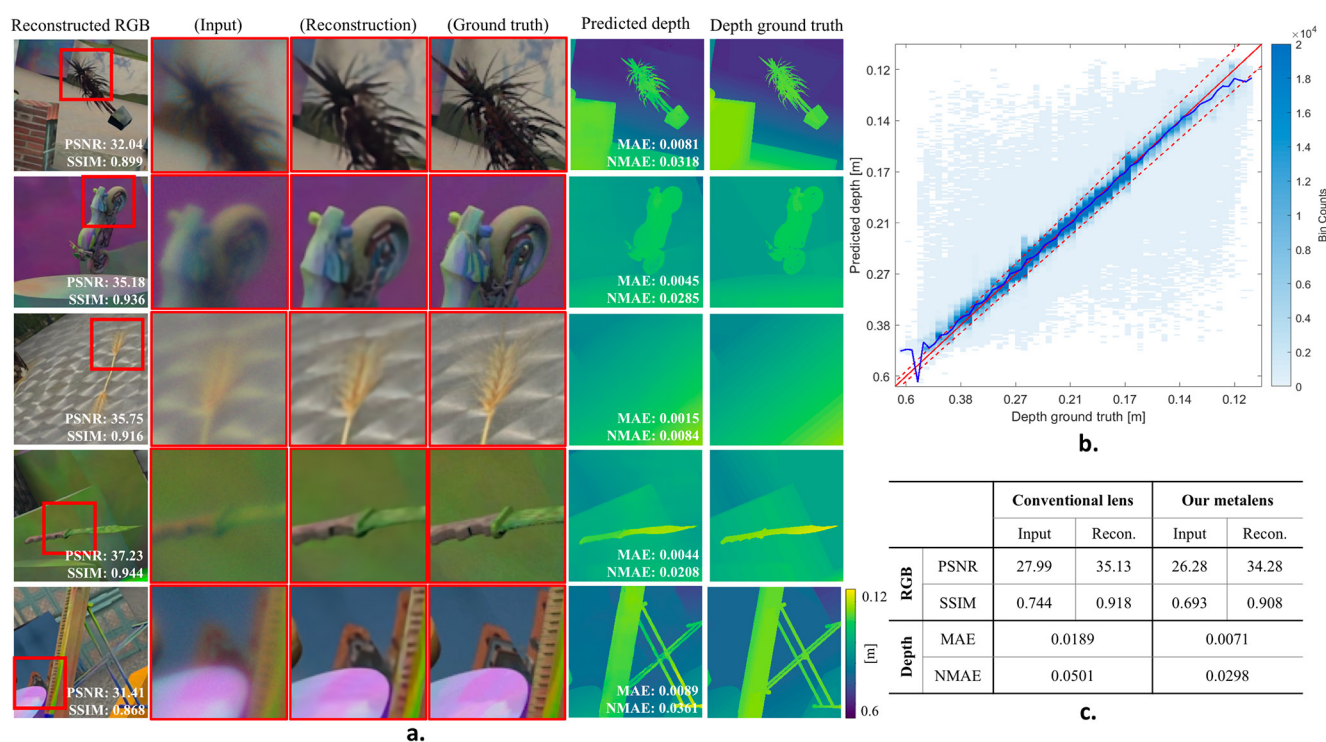


Figure 6. Performance analysis: (a) Reconstruction results for our simulated metalens. (b) Binned scatter plots for predicted depth as a function of ground truth. The plot partitions the depth space into rectangular bins and displays the count of depth points in each bin using different shades of blue (darker shades for more counts). The solid blue curve is the mean of the predicted depth over ground truths. The diagonal solid red line represents the ideal predictions, and the two dashed red lines represent the $\pm 5\%$ boundaries. For visualization, the x and y axes are linearly sampled in disparity (inverse depth) space. Most of the predictions fall within the $\pm 5\%$ range of the true depths, over the target depth range of 0.12–0.6 m. (c) Comparison with a conventional diffractive lens in average reconstruction accuracy. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) are used on texture reconstruction (the higher the better), and the mean absolute error (MAE) and the normalized MAE (NMAE, absolute error divided by its ground truth) are used on depth prediction (the lower the better).

GeForce RTX 2080 Ti GPUs. The regularization weight of the gradient loss is $\alpha = 1$.

Results. Simulation results of depth and RGB texture reconstruction are shown in Figures 5 and 6. We also compare our metalens with a conventional lens, a case in which the wavelength dependence is quite small. The conventional lens is designed to have the same focal length (44.5 mm) and focus at the same depth (0.2 m) as our metalens (for green light centered at 540 nm). In particular, it is designed to have a quadratic height profile

$$h(x, y)(n_g - 1) = \left(\sqrt{x^2 + y^2 + f_g^2} - f_g \right) \quad (\text{without phase wrapping}),$$

where $f_g = 44.5$ mm is the focal length and $n_g = 1.52$ is the refractive index at 540 nm. The feature sizes of the conventional lens are all much bigger than wavelength (unlike designs based on metasurfaces). Therefore, dispersion in such conventional lenses arises only from material dispersion. Material dispersion is a small effect for most transparent materials. On the other hand, metasurfaces can have much larger dispersion. Metasurfaces rely on the resonant interaction of light with subwavelength scale nanostructures. The resonance is a narrowband feature and hence makes the metasurface highly dispersive. Additionally, the narrowband operation of metasurface allows interleaving and wavelength-specific design. In our case, we interleave meta-elements to multiplex three different metalenses. Each metalens focuses a different design wavelength at a different depth with maximized efficiency (polarization conversion efficiency in our case) while suppressing transmission of other wavelengths (as shown in Figure 3b). This degree of freedom is not accessible with conventional lenses. Thus, the use of a metasurface provides us a way to predict depth information on 3D scenes with much higher accuracy than with a conventional lens. A comparison of the reconstruction results of the conventional lens and our method is shown in Figure 5. We adopt the same PSF-based RGBD rendering procedure to simulate its captured images and the same U-Net based networks to estimate the depth and RGB images. For quantitative evaluation, the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) are used on texture reconstruction (the higher the better),⁴⁸ and the mean absolute error (MAE) and the normalized MAE (NMAE, absolute error divided by its ground truth) are adopted on depth prediction (the lower the better). Our designed metalens outperforms the conventional lens in depth prediction accuracy and maintains similar performance in RGB texture reconstruction. A summary of average reconstruction performance (over the testing data set) of the conventional lens and metalens is shown in Figure 6c. To further analyze the depth prediction accuracy of our metalens, we show the binned scatter plots for the predicted depth as a function of ground-truth depth in Figure 6b. The plot indicates that most of the predicted depths (dark blue) hover within $\pm 5\%$ of the true depths. A small number of depth predictions for textureless or color-unbalanced objects (light blue) fall beyond the $\pm 5\%$ range, as the chromatic-defocus cues are not available for objects without texture or color information. More simulation results for various scenes are shown in Figure 6a.

In contrast to the two previous designs,^{8,9} our metalens spatially multiplexes three lens phase profiles with distinct focal lengths designed at different wavelengths. Thereby, our design enables single-shot operation with high transmission efficiency

over the entire visible spectrum. The broadband operation increases the light efficiency and signal-to-noise ratio (SNR) compared to a monochromatic system and enables us to reconstruct RGB color images of scenes. Further, our design has an advantage over the prior depth-sensor systems. Earlier designs were based on two adjacent non-overlapping images,^{8,9} requiring a large sensor and/or limiting the field of view (FOV). In our system, only one image is captured by the sensor for simultaneous depth estimation and RGB reconstruction. Further, the reflectance spectra of most real-world objects are sufficiently broadband and hence, our technique is general and applicable to real-world situations.³⁶

SUMMARY AND CONCLUSION

We have proposed a metalens 3D sensor that encodes depth information in color space, that is, different wavelengths focus at different depths. We take advantage of strong chromatic dispersion of metalenses to achieve the 3D imaging functionality over the entire visible spectrum. Compared to other metalens-based 3D imaging,^{8,9} this chromatic encoding eliminates the limitations of spectral bandwidth and field-of-view, enhances energy efficiency, and is compatible with various aperture designs.

Our design builds on a previous experimental demonstration exploiting chromatic dispersion to extend the depth of field of imaging devices.³⁶ Similar to this past work, we make use of the fact that the reflectance spectra of most real-world objects are broadband. Therefore, typical captured scenes contain sufficient information for simultaneous depth and texture reconstruction.

Furthermore, we use reconstruction deep networks to estimate the depth and texture information on a 3D scene simultaneously. Deep networks provide more reliable dense 3D reconstructions of complicated scenes compared to traditional deconvolution methods. Also, deep networks can handle cases such as transparent objects or textureless regions, which are challenging for deconvolution techniques.³⁰

Despite the advantages of our approach, some challenges remain. These include the reconstruction of large textureless areas and objects with sharp features in their reflectance spectra. The prediction accuracy tends to be lower for large depth values because the network estimates depth from defocus blur, and the defocus phase changes inversely with depth. Moreover, our metalens is designed for circularly polarized light, limiting the energy efficiency compared to the scalar diffractive lenses (polarization insensitive). However, it can reconstruct 3D information with greater accuracy, making it useful for low-power applications. Nevertheless, optimizing the metasurface and system design can significantly mitigate these limitations. Possible approaches include exploring polarization-insensitive meta-elements, tuning the focal lengths of the red, green, and blue metalenses, and employing end-to-end learning techniques^{30,40} to jointly optimize the optical and computational components.

Overall, we have presented, for the first time, a strategy to implement 3D imaging over the entire visible spectrum in a single shot. We exploited the strong chromatic dispersion of metasurfaces and combined it with computational optics to achieve 3D imaging. Our method can be adapted to various depth ranges. Here, we demonstrated the design for the depth range 0.12–0.6 m using a 1 mm aperture. Simulation results suggest significant improvements in depth prediction accuracy over conventional lens-based techniques. This work demon-

strates how computational optics can turn a drawback of metasurfaces, their strong chromatic dispersion, into a novel functionality suitable for mobile or wearable platforms.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsphotonics.1c00110>.

Reconstruction network architectures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Gururaj V. Naik – Electrical and Computer Engineering, Rice University, Houston, Texas 77005, United States;

orcid.org/0000-0002-3800-3392; Phone: +1 713 348 2528; Email: guru@rice.edu

Authors

Shiyu Tan – Electrical and Computer Engineering, Rice University, Houston, Texas 77005, United States;

orcid.org/0000-0002-0779-5956

Frank Yang – Electrical and Computer Engineering, Rice University, Houston, Texas 77005, United States;

orcid.org/0000-0001-5591-4035

Vivek Boominathan – Electrical and Computer Engineering, Rice University, Houston, Texas 77005, United States

Ashok Veeraraghavan – Electrical and Computer Engineering, Rice University, Houston, Texas 77005, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsphotonics.1c00110>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the NSF CAREER Grant (A Signal Processing Framework for Computational Imaging: From Theory to Applications) IIS-1652633 and the NSF SaTC CORE Medium Collaborative Grant (Presentation-attack-robust biometrics systems via computational imaging of physiology and materials) CNS-1801372.

■ REFERENCES

- (1) Kildishev, A. V.; Boltasseva, A.; Shalae, V. M. Planar photonics with metasurfaces. *Science* **2013**, *339*, 1232009.
- (2) Yu, N.; Capasso, F. Flat optics with designer metasurfaces. *Nat. Mater.* **2014**, *13*, 139–150.
- (3) Lin, D.; Fan, P.; Hasman, E.; Brongersma, M. L. Dielectric gradient metasurface optical elements. *Science* **2014**, *345*, 298–302.
- (4) Genevet, P.; Capasso, F.; Aieta, F.; Khorasaninejad, M.; Devlin, R. Recent advances in planar optics: from plasmonic to dielectric metasurfaces. *Optica* **2017**, *4*, 139–152.
- (5) Kamali, S. M.; Arbabi, E.; Arbabi, A.; Faraon, A. A review of dielectric optical metasurfaces for wavefront control. *Nanophotonics* **2018**, *7*, 1041–1068.
- (6) Ding, F.; Pors, A.; Bozhevolnyi, S. I. Gradient metasurfaces: a review of fundamentals and applications. *Rep. Prog. Phys.* **2018**, *81*, 026401.
- (7) Yu, N.; Genevet, P.; Kats, M. A.; Aieta, F.; Tetienne, J.-P.; Capasso, F.; Gaburro, Z. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *Science* **2011**, *334*, 333–337.
- (8) Guo, Q.; Shi, Z.; Huang, Y.-W.; Alexander, E.; Qiu, C.-W.; Capasso, F.; Zickler, T. Compact single-shot metalens depth sensors

inspired by eyes of jumping spiders. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 22959–22965.

(9) Colburn, S.; Majumdar, A. Single-shot three-dimensional imaging with a metasurface depth camera. *arXiv:1910.12111 [physics.optics]* **2019**, na.

(10) Jin, C.; Afsharnia, M.; Berlich, R.; Fasold, S.; Zou, C.; Arslan, D.; Staude, I.; Pertsch, T.; Setzpfandt, F. Dielectric metasurfaces for distance measurements and three-dimensional imaging. *Adv. Photonics* **2019**, *1*, 036001.

(11) Colburn, S.; Majumdar, A. Metasurface Generation of Paired Accelerating and Rotating Optical Beams for Passive Ranging and Scene Reconstruction. *ACS Photonics* **2020**, *7*, 1529–1536.

(12) Colburn, S.; Zhan, A.; Majumdar, A. Metasurface optics for full-color computational imaging. *Sci. Adv.* **2018**, *4*, No. eaar2114.

(13) Colburn, S.; Majumdar, A. Simultaneous Achromatic and Varifocal Imaging with Quartic Metasurfaces in the Visible. *ACS Photonics* **2020**, *7*, 120–127.

(14) Huang, L.; Whitehead, J.; Colburn, S.; Majumdar, A. Design and analysis of extended depth of focus metalenses for achromatic computational imaging. *Photonics Res.* **2020**, *8*, 1613–1623.

(15) Kwon, H.; Arbabi, E.; Kamali, S. M.; Faraji-Dana, M.; Faraon, A. Computational complex optical field imaging using a designed metasurface diffuser. *Optica* **2018**, *5*, 924–931.

(16) Chen, W. T.; Zhu, A. Y.; Capasso, F. Flat optics with dispersion-engineered metasurfaces. *Nat. Rev. Mater.* **2020**, *5*, 604–620.

(17) Khorasaninejad, M.; Chen, W. T.; Devlin, R. C.; Oh, J.; Zhu, A. Y.; Capasso, F. Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging. *Science* **2016**, *352*, 1190–1194.

(18) Chen, C.; Song, W.; Chen, J.-W.; Wang, J.-H.; Chen, Y. H.; Xu, B.; Chen, M.-K.; Li, H.; Fang, B.; Chen, J.; Kuo, H. Y.; Wang, S.; Tsai, D. P.; Zhu, S.; Li, T.; et al. Spectral tomographic imaging with aplanatic metalens. *Light: Sci. Appl.* **2019**, *8*, 1–8.

(19) Khorasaninejad, M.; Chen, W.-T.; Oh, J.; Capasso, F. Super-dispersive off-axis meta-lenses for compact high resolution spectroscopy. *Nano Lett.* **2016**, *16*, 3732–3737.

(20) Maddern, W.; Newman, P. Real-time probabilistic fusion of sparse 3d lidar and dense stereo. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE, 2016; pp 2181–2188.

(21) Shrestha, S.; Heide, F.; Heidrich, W.; Wetzstein, G. Computational imaging with multi-camera time-of-flight systems. *ACM T. Graphics* **2016**, *35*, 1–11.

(22) Hansard, M.; Lee, S.; Choi, O.; Horaud, R. P. *Time-of-flight cameras: principles, methods and applications*; Springer Science & Business Media, 2012; doi: DOI: 10.1007/978-1-4471-4658-2.

(23) Gupta, M.; Nayar, S. K.; Hullin, M. B.; Martin, J. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Trans. Graphics* **2015**, *34*, 1–18.

(24) Geng, J. Structured-light 3D surface imaging: a tutorial. *Adv. Opt. Photonics* **2011**, *3*, 128–160.

(25) Lazaros, N.; Sirakoulis, G. C.; Gasteratos, A. Review of stereo vision algorithms: from software to hardware. *Int. J. Optomechatronics* **2008**, *2*, 435–462.

(26) Geiger, A.; Ziegler, J.; Stiller, C. Stereoscan: Dense 3d reconstruction in real-time. *IEEE intelligent vehicles symposium IV*; IEEE, 2011; pp 963–968.

(27) Jin, C.; Zhang, J.; Guo, C. Metasurface integrated with double-helix point spread function and metalens for three-dimensional imaging. *Nanophotonics* **2019**, *8*, 451–458.

(28) Haim, H.; Elmaleh, S.; Giryes, R.; Bronstein, A. M.; Marom, E. Depth estimation from a single image using deep learned phase coded mask. *IEEE Trans. Comput. Imag.* **2018**, *4*, 298–310.

(29) Chang, J.; Wetzstein, G. Deep optics for monocular depth estimation and 3d object detection. *Proceedings of the IEEE International Conference on Computer Vision*; IEEE, 2019; pp 10193–10202.

(30) Wu, Y.; Boominathan, V.; Chen, H.; Sankaranarayanan, A.; Veeraraghavan, A. Phasecam3d-learning phase masks for passive

single view depth estimation. *IEEE International Conference on Computational Photography (ICCP)*; IEEE, 2019; pp 1–12.

(31) Chaudhuri, S.; Rajagopalan, A. N. *Depth from defocus: a real aperture imaging approach*; Springer Science & Business Media, 2012; DOI: 10.1007/978-1-4612-1490-8.

(32) Darrell, T.; Wohn, K. Pyramid based depth from focus. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; IEEE, 1988; pp 504–505.

(33) Xiong, Y.; Shafer, S. A. Depth from focusing and defocusing. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* 1993, 68–73.

(34) Lin, R. J.; et al. Achromatic metalens array for full-colour light-field imaging. *Nat. Nanotechnol.* 2019, 14, 227–231.

(35) Fan, Z.-B.; Qiu, H.-Y.; Zhang, H.-L.; Pang, X.-N.; Zhou, L.-D.; Liu, L.; Ren, H.; Wang, Q.-H.; Dong, J.-W. A broadband achromatic metalens array for integral imaging in the visible. *Light: Sci. Appl.* 2019, 8, 1–10.

(36) Cossairt, O.; Nayar, S. Spectral focal sweep: Extended depth of field from chromatic aberrations. *IEEE International Conference on Computational Photography (ICCP)*; IEEE, 2010; pp 1–8.

(37) Parkkinen, J. P.; Hallikainen, J.; Jaaskelainen, T. Characteristic spectra of Munsell colors. *J. Opt. Soc. Am. A* 1989, 6, 318–322.

(38) Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*; MICCAI Society, 2015; pp 234–241.

(39) Tan, S.; Wu, Y.; Yu, S.-I.; Veeraraghavan, A. CodedStereo: Learned Phase Masks for Large Depth-of-field Stereo, 2021, <https://arxiv.org/abs/2104.04641>.

(40) Sitzmann, V.; Diamond, S.; Peng, Y.; Dun, X.; Boyd, S.; Heidrich, W.; Heide, F.; Wetzstein, G. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)* 2018, 37, 1–13.

(41) Campbell, S. D.; Sell, D.; Jenkins, R. P.; Whiting, E. B.; Fan, J. A.; Werner, D. H. Review of numerical optimization techniques for meta-device design [Invited]. *Opt. Mater. Express* 2019, 9, 1842–1863.

(42) Palik, E. D. *Handbook of optical constants of solids*; Academic Press, 1998; Vol. 3; DOI: 10.1016/C2009-0-20920-2.

(43) Devlin, R. C.; Khorasaninejad, M.; Chen, W. T.; Oh, J.; Capasso, F. Broadband high-efficiency dielectric metasurfaces for the visible spectrum. *Proc. Natl. Acad. Sci. U. S. A.* 2016, 113, 10473–10478.

(44) Goodman, J. W. *Introduction to Fourier Optics*; Roberts and Company Publishers, 2005; DOI: 10.1063/1.3035549.

(45) Scofield, C. *Graphics Gems III (IBM Version)*; Elsevier, 1992; pp 36–38.

(46) Mayer, N.; Ilg, E.; Hausser, P.; Fischer, P.; Cremers, D.; Dosovitskiy, A.; Brox, T. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition* 2016, 4040–4048.

(47) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv:1412.6980 [cs.LG]* 2014, na.

(48) Hore, A.; Ziou, D. Image quality metrics: PSNR vs SSIM. 2010 20th International Conference on Pattern Recognition; IEEE, 2010; pp 2366–2369.