OPTICA

Foveated thermal computational imaging prototype using all-silicon meta-optics

Vishwanath Saragadam,[1] Zheyi Han,[2] Vivek Boominathan,[1] Luocheng Huang,[2] Shiyu Tan,[1] Johannes E. Fröch,[2,3] Karl F. Böhringer,[2,4] Richard G. Baraniuk,[1] Arka Majumdar,[2,3] and Ashok Veeraraghavan[1,*]

[1]Department of ECE, Rice University, Houston, Texas 77005, USA
[2]Department of ECE, University of Washington, Seattle, Washington 98195, USA
[3]Department of Physics, University of Washington, Seattle, Washington 98195, USA
[4]Institute for Nano-Engineered Systems, University of Washington, Seattle, Washington 98195, USA
*vashok@rice.edu

Received 17 August 2023; revised 13 October 2023; accepted 14 November 2023; published 3 January 2024

Foveated imaging provides a better tradeoff between situational awareness (field of view) and resolution, and is critical in long wavelength infrared regimes because of the size, weight, power, and cost of thermal sensors. We demonstrate computational foveated imaging by exploiting the ability of a meta-optical frontend to discriminate between different polarization states and a computational backend to reconstruct the captured image/video. The frontend is a three-element optic: the first element, which we call the "foveal" element, is a metalens that focuses s-polarized light at a distance of f_1 without affecting the p-polarized light; the second element, which we call the "perifovea" element, is another metalens that focuses p-polarized light at a distance of f_2 without affecting the *s*-polarized light. The third element is a freely rotating polarizer that dynamically changes the mixing ratios between the two polarization states. Both the foveal element (focal length = 150 mm; diameter = 75 mm) and the perifoveal element (focal length = 25 mm; diameter = 25 mm) were fabricated as polarization-sensitive, all-silicon, meta surfaces resulting in a large-aperture, 1:6 foveal expansion, thermal imaging capability. A computational backend then utilizes a deep image prior to separate the resultant multiplexed image or video into a foveated image consisting of a high resolution center and a lower-resolution large field of view context. We build a prototype system and demonstrate 12 frames per second real-time, thermal, foveated image and video capture..

https://doi.org/10.1364/OPTICA.502857

1. INTRODUCTION

Nature has optimized the mammalian eye to have a large field of view (FoV) for situational awareness and high resolution in the central area for target detection with high precision. In the human eye, the fovea (small central region covering a 2° FoV) carries half as much information as the perifovea (large outer region covering more than 30° FoV), thereby dedicating precious bandwidth to the central region [see Fig. 1(a)]. Such foveation has long been desired in artificial vision systems as well, since high resolution in the central area is desirable for target detection, while large FoV is essential for contextual awareness. This is particularly true at long wave infrared (LWIR) wavelengths where detectors tend to be low resolution making the simultaneous capture of a high resolution target and large FoV context especially challenging.

Existing solutions for foveated imaging such as foveated sensors [1,2] and foveating dynamic optics [3] do not work for thermal imaging because of the unique challenges at those wavelengths. Foveated sensor designs [2] require an array of cameras, which can be prohibitively expensive in the thermal wavelengths, while requiring depth-dependent calibration between the various sub-views. Single sensor solutions such as the 3D-printed eagle eye by [1] directly trade-off the sensor area for multiple focal lengths. Since thermal sensors tend to be low resolution (smaller than one megapixel), such approaches are infeasible. Foveated dynamic optics such as [3] have never been demonstrated in the thermal regime, primarily because of the challenges in dynamic control of optical properties in materials such as germanium and zinc selenide (materials used typically used to fabricate LWIR lenses). A closely related approach by [4] leverages compound optics to obtain a single composite image that superimposes the low resolution and the high resolution images. This seminal work showcased a conceptual idea along with simulated examples on how a single sensor can be leveraged to simultaneously obtain images with two FoVs. While being a promising solution, the inverse problem of recovering two images from a single image is severely under-constrained. In this paper, we leverage recent advances in two rapidly emerging fields—meta-optics and computational imaging—to demonstrate a light-weight, first-of-its-kind, real-time, thermal, foveated image and video capture in the wild.

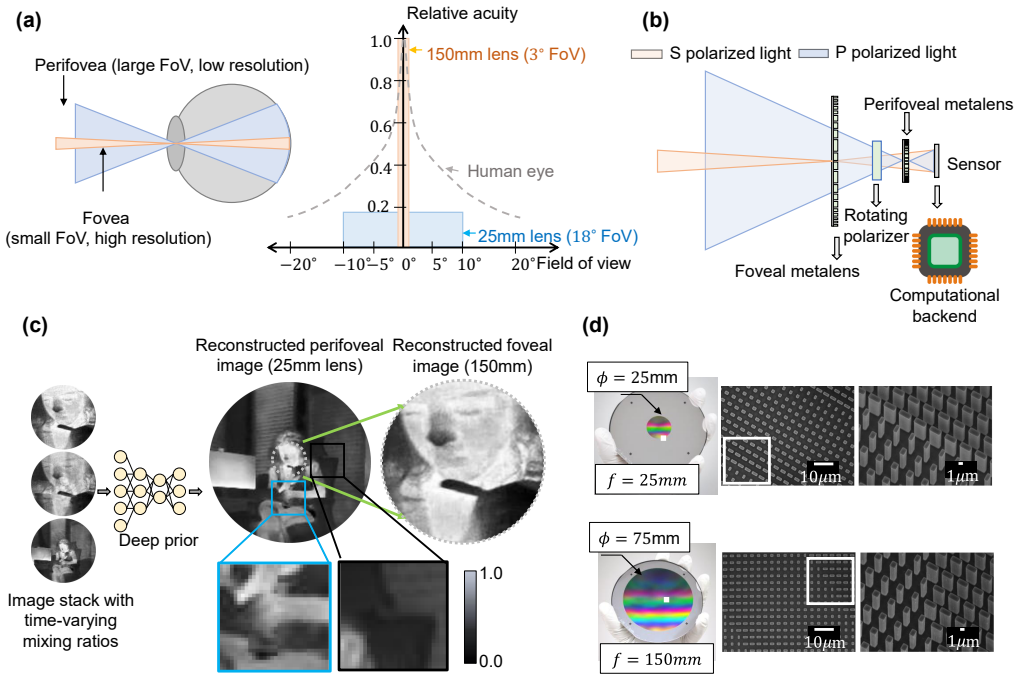

Fig. 1. Overview of proposed foveated imaging system. (a) Our proposed foveated metalens is inspired by the human eye, which has high acuity (high resolution) over a small field of view (fovea) and a low acuity over a large field of view (perifovea). (b) Our optic consists of a foveal element at a focal length of 150 mm, a perifoveal element at 25 mm, and a freely rotating polarizer. The foveal element only modulates the s-polarized light, while the perifoveal element only modulates the p-polarized light. The sensor measures a linear combination of the two images, which is then unmixed with a computational backend. (c) Reconstruction with the computational backend for a simulated example. (d) Scanning electron microscopic images of the two lenses.

Meta-optics with sub-wavelength feature sizes for arbitrary phase control [5,6] have rapidly grown as alternatives to traditional optics [7–13], over a wide range of wavelengths and novel applications, including reflectors [14–16], vortex beam generators [17–20], holographic masks [21–23], gratings [24,25], optical convolutional neural networks [26], and polarization optics [27,28]. Meta-optics have two distinct advantages over traditional optics: (a) multiplexing (combining multiple phase functions in a single surface) and (b) polarization control (meta-optics can be manufactured with different phase functions for orthogonal states of polarization), and these properties have been used previously for polarimetry [29], depth sensing [30], and tunable focus [31,32].

Here, we leverage the polarization sensitivity of meta-optics to design a three-element optic that achieves large-aperture foveated imaging at the LWIR wavelengths. The first element, which we call the "foveal" element, is a metalens that encodes a phase function for a convex lens of focal length f_1 for s-polarization, whereas the p-polarized light is transmitted without any change (or a lens with focal length ∞). The second element, which we call the "perifoveal" element, is another metalens encoding a phase function for a convex lens of focal length f_2 for p-polarization, whereas the s-polarization that was modulated by the previous surface is left unaltered. The third element is a freely rotating polarizer that dynamically changes the mixing ratios between the two polarization states. Ideally, a polarizer with a fast changing binary state would enable capturing foveal and perifoveal images. However, such a system is hard to implement in practice. A freely rotating polarizer is easier to implement, and enables real-time capture of data. The optical setup results in a linear expansion of the FoV by $f_1 : f_2$ by the foveal lens, which we call the foveal expansion. The schematic of the proposed optical system is shown in Fig. 1(b). S- and p-polarized light independently focused by these two elements multiplexes on the image sensor, resulting in a video sequence that is not human-interpretable, but can be leveraged to reconstruct the two images with computational approaches.

Over the last decade, computational imaging has emerged as a powerful tool, where signal processing [33–42] and machine learning [43–47] algorithms co-designed with such multiplexed imaging systems can undo the effects of multiplexing and reconstruct images and videos. The freely rotating linear polarizer better-conditions the inverse problem by dynamically changing the fraction of s-polarized and p-polarized images that is multiplexed on the image sensor, meaning each frame in the captured video has a different weighted combination of the two images (one with a foveal element and another with a perifoveal element). The two images are then recovered by a computational backend that takes the sequence of images as input, and simultaneously estimates the time-varying mixing ratios, as well as the low resolution, large FoV and high resolution, narrow FoV images. Most conventional deep-learning-based reconstruction algorithms require large amounts of in-domain training data to be successful, something that is challenging when using LWIR meta-optics due to the poor signal to noise ratio (SNR) of LWIR sensors, as well as a paucity of training data. Our reconstruction algorithm consists of a deep generative prior [48] that has the distinct advantage of not requiring any training data, while producing high quality results. The computational backened as well as a simulated example is shown in Fig. 1(c).

We validated our approach by designing and fabricating the metalenses (optimized for 10 μm wavelength) in an all-silicon platform using direct laser writing [8]. The foveal element was designed with a diameter of 75 mm for a focal length of 150 mm.

The perifoveal element was designed with a diameter of 25 mm for a focal length of 25 mm. This resulted in a foveal expansion of 6:1. Images of the metalenses, as well as their scanning electron microscopic (SEM) images are shown in Fig. 1(d). Our three-element optic demonstrates videos captured in the wild at real time (>12 frames per second) enabling high resolution thermal videos of buildings, moving cars, and dynamic humans.

2. PROPOSED METHOD

A. System Overview

Figure 1 shows an overview of the foveated computational metaoptical system optimized for imaging at 10 μm wavelength. The phase functions of the two metalenses are

$$\left.\begin{aligned}\phi_{1,P}(x,y) &= \tfrac{2\pi}{\lambda_0}\left(\sqrt{x^2+y^2+f_1^2}-f_1\right)\\ \phi_{1,S}(x,y) &= \text{constant}\end{aligned}\right\}\ \text{Foveal metalens,} \tag{1}$$

$$\left.\begin{aligned}\phi_{2,P}(x,y) &= \text{constant}\\ \phi_{2,S}(x,y) &= \tfrac{2\pi}{\lambda_0}\left(\sqrt{x^2+y^2+f_2^2}-f_2\right)\end{aligned}\right\}\ \text{Perifoveal metalens,} \tag{2}$$

where $\lambda_0 = 10$ μm is the design wavelength. Since the physical phase of each meta-atom can only modulate between 0 and 2π, we implemented a modulo 2π wrapped phase function for each lens. The foveal lens has a focal length of $f_1 = 150$ mm and a diameter of 75 mm, while the perifoveal lens has a focal length of $f_2 = 25$ mm and a diameter of 25 mm. The lenses are placed at a distance of f_1 and f_2, respectively, from the sensor. The sensor captures images 12 frames per second.

B. Metalens Design

To implement the polarization-dependent phase response in the two meta-optics we designed a scatterer with a polarization-dependent response using a rectangular footprint of each nanopost. We first calculated the phase and amplitude response for pillars using rigorous coupled wave analysis (RCWA) [49], assuming a pillar height of 10 μm, period of 4 μm, and varying rectangular footprint defined by the sidewidths, x and y. We then selected a scatterer that would have a phase response covering a range of 2π for s-polarization, while not altering the phase for light with p-polarization. We note that meta-optics for imaging in LWIR have been demonstrated in recent times with polarization insensitivity [50], broadband response [51], large aperture [52], and operation in ambient temperatures [8]. However, to the best of our knowledge, our system is the first to demonstrate results on polarization-sensitive metalenses at LWIR wavelengths. Further details about the optimization procedure are available in Supplement 1.

C. Metalens Fabrication

Each metalens was fabricated on a 500 μm thick double-side polished silicon wafer, lightly doped with boron, giving a sheet resistivity of $10-20\ \Omega-$ cm. We used direct-write lithography (Heidelberg DWL 66^+) to define the location of the metalens aperture in a negative photoresist layer. A 240 nm thick aluminum layer was deposited via electron beam evaporation (CHA solution) and lifted off to form the metal mask surrounding the designated aperture of the metalens, reducing the noise in the experiments. The metalens scatterer layout was aligned and patterned into the open circular aperture using direct-write lithography with a positive photoresist. We utilized deep reactive-ion etching (SPTS DRIE) to transfer the metalens pattern into the silicon layer with a scatterer depth of 10 μm and highly vertical sidewalls.

D. Optimization Approach

Let $I_1(x,y)$, $I_2(x,y)$ be the images formed by the foveal and perifoveal metalenses, respectively. Assuming the two images are static between $t=t_1$ and $t=t_2$, the resultant image on the sensor is

$$I_{\text{meas.}}(x,y) = \alpha_t I_1(x,y) + (1-\alpha_t) I_2(x,y), \tag{3}$$

where $\alpha_t = \cos^2(\theta_t)$ is the mixing ratio depending on the position of the polarizer at time t. Since the polarizer is freely rotating, its exact position is unknown at each image capture, and hence we estimate its position (and hence mixing ratio) along with the foveal and perifoveal images.

Let $\mathbf{x}^1 = I_1(x,y)$ and $\mathbf{x}^2 = I_2(x,y)$ be the vector representation of perifoveal and foveal images, respectively, to be estimated from the measurements $\mathbf{y}_t$. To regularize the inverse problem, we rely on the inherent regularization offered by convolutional neural networks and solve the following modified optimization problem:

$$\min_{\alpha_t,\theta_1,\theta_2} \sum_{t=t_1}^{t_2} \left\| \mathbf{y}_t - \alpha_t \mathbf{x}^1 - (1-\alpha_t)\mathbf{x}^2 \right\|^2 + \mathcal{R}\left(\mathbf{x}^1, \mathbf{x}^2\right), \tag{4}$$

$$\mathbf{x}^1 = \mathcal{N}_1(\mathbf{n}_1; \theta_1), \tag{5}$$

$$\mathbf{x}^2 = \mathcal{N}_2(\mathbf{n}_2; \theta_2), \tag{6}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ are untrained neural networks, θ_1, θ_2 are weights of the neural networks to be optimized, and $\mathbf{n}_1$ and $\mathbf{n}_2$ are inputs with each entry drawn in a uniformly random manner. This approach of optimizing for the weights of an untrained neural network instead of the image is similar in spirit to deep image prior [48]. Utilizing two separate networks promotes dissimilarity between the two images $\mathbf{x}^1$ and $\mathbf{x}^2$, and hence results in high quality separation [53]. Further details about the network, as well as the regularization function $\mathcal{R}$, are available in Supplement 1.

E. Experimental Setup

Our setup consists of two, four inch metalenses mounted on optomechanical systems (Thorlabs LMR4) on a rail system. A linear polarizer (Thorlabs WP25M-IRC) was mounted on a fast rotation stage (Thorlabs ELL14) and asynchronously rotated at 2 Hz. For some of our experiments, we placed a 10 μm spectral filter with a bandwidth of 500 nm to obtain sharp images. All the results in the paper were captured with an exposure duration of 40 ms. Images in Figs. 2(a) and 2(c), Figs. 3, and 4 were captured with an Infratec VarioCAM HD 1800 camera, while the images in Figs. 2(b) and 2(d), and Fig. 5 were captured with a FLIR A655sc camera. The Infratec VarioCAM HD 1800 camera was fitted with a 0.5× relay lens, and the FLIR A655sc was fitted with a 0.3× relay lens. The relay system was required to overcome the mechanical constraints of the camera, which prevented us from placing the 25 mm lens at the appropriate distance from the sensor. We found that the relay had no effect on the final quality of the acquired

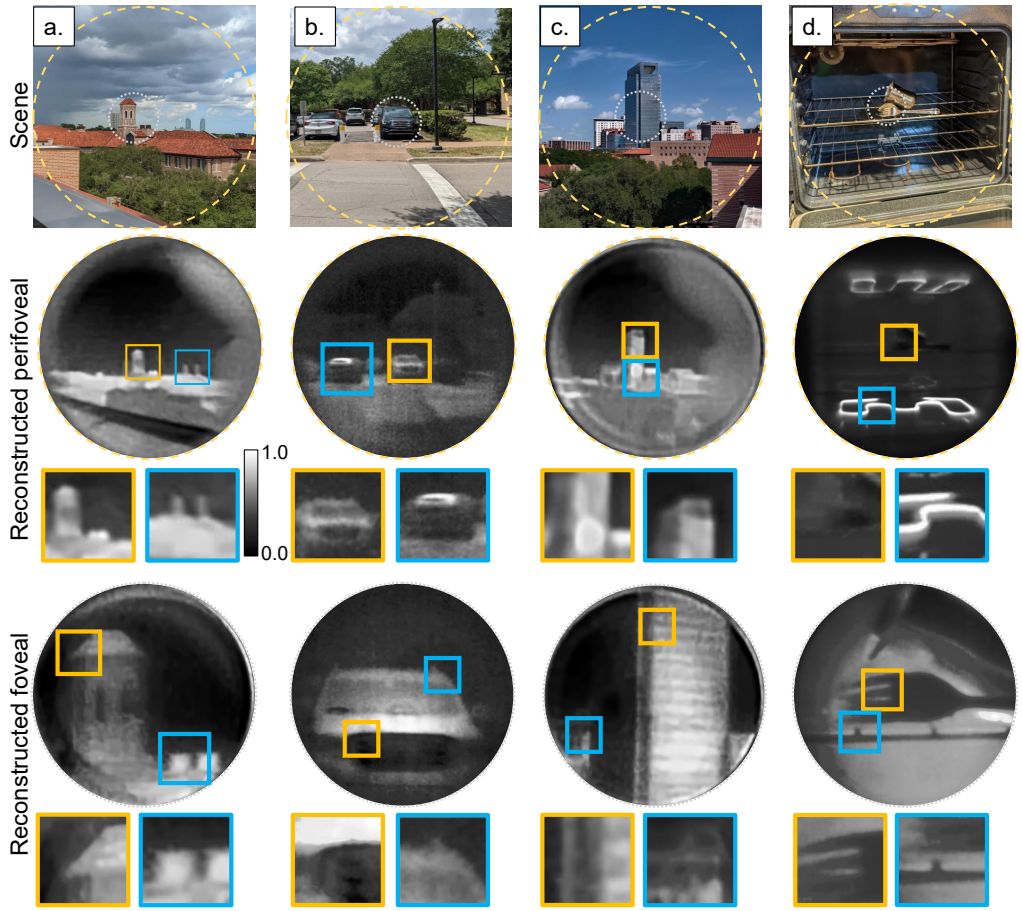

Fig. 2. Static imaging in the wild. We captured images of several outdoor and indoor scenes including (a) rooftop, (b) cars in parking lot, (c) tall buildings, and (d) a hot oven. For scenes with objects at relatively lower temperature (a), (c), we removed the narrowband spectral filter to increase SNR. Each of the scenes has a resolution of 290 × 290 pixels with each pixel measuring 24 μm. Our computational backend was able to recover very fine features such as windows in the rooftop building, the car grill, windows in the tall skyscraper, and the fork in the oven.

images, except for a magnification. Further details are available in Supplement 1.

3. RESULTS

A. Image Reconstruction of Still Life in the Wild

Our optical setup produces high resolution foveated images in diverse indoor and outdoor settings. The powerful computational backend is capable of reconstructing images from raw measurements in diverse settings. Figure 2 shows reconstruction of several outdoor (a), (b), (c) and indoor (d) scenes with a wide range of temperature variations. We note that the relative FoV of the foveal lens can be changed in a content-adaptive manner, very similar to a zoom system. The perifoveal image can be used to identify a region of interest (ROI) where higher resolution is required, and then the foveal lens can be moved/tilted appropriately to focus on this region. In our examples, we moved the foveal lens with respect to the perifoveal lens in a way to visualize parts of the scene with interesting features, such as the tall tower in (a), foreground car in (b), building panes in (c), and fork in (d). In each case, 8 s of video data was captured with the polarizer rotating freely. We removed the spectral filter for Figs. 2(a) and 2(c) to increase measurement SNR. We note that, while our metalens is designed only for 10 μm, we can still capture images under broadband light as demonstrated recently [8]. The advantages of dual focal length are evident—the perifoveal image provides a context of the surroundings, while the foveal image provides high quality details of the central region. In particular, note the clear grating-like structure in Fig. 2(c) due to window frames, as well as the parallel prongs of the fork in Fig. 2(d).

B. Video Reconstruction of Dynamic Scenes at 12 fps

Our optical setup is capable of capturing dynamic scenes at high spatial and temporal resolutions, a result previously unreported in the metalens literature. To demonstrate this, we imaged three people exhibiting different forms of motion, as shown in Fig. 3(a). The person on the left was rocking left to right, the person in the middle was sitting on a chair and waving his hand with small range of motion, and the person on the right was waving his hand with a large range. We reconstructed a video sequence in a sliding window manner, where a continuous sequence of eight measured images was used to recover a pair of perifoveal and foveal video frames. Note the perifoveal image clearly showing waving of the right person's hand, but nearly no motion of the person sitting in the center. The foveal, however, clearly shows the waving motion, along with all fingers separately. Figure 3(b) shows a parking lot with several cars. The video sequence was captured by moving the camera from the right end of the parking lot to the left end.

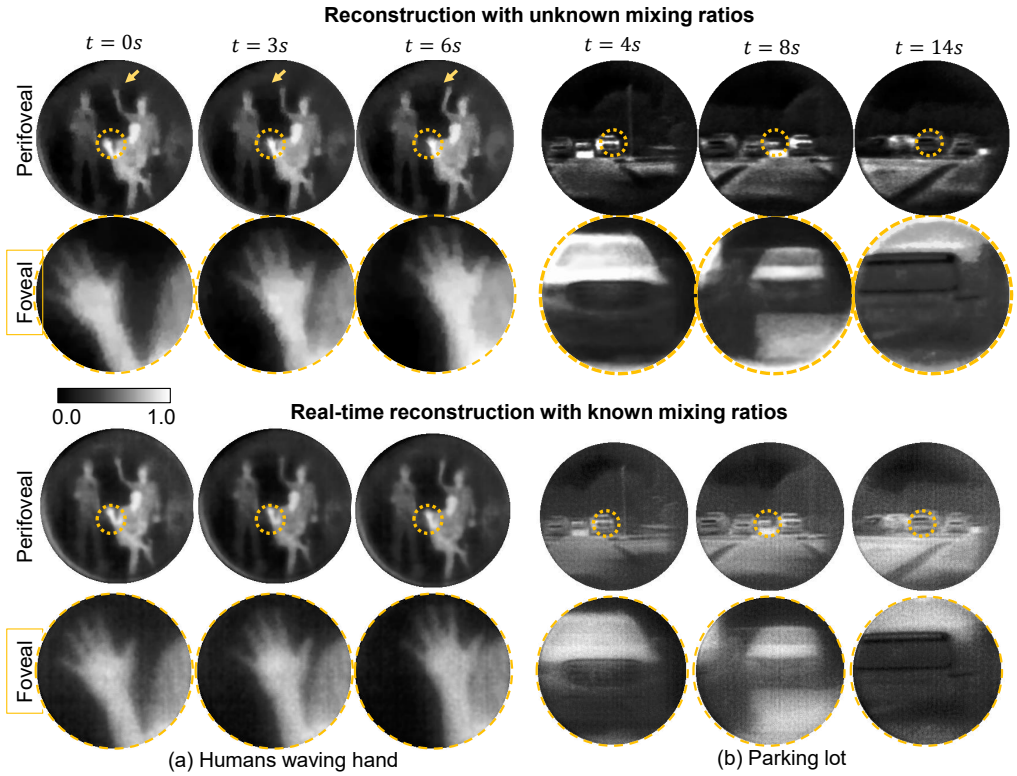

Fig. 3. Capturing videos of dynamic scenes. Our optical setup enables simultaneous reconstruction of small and large FoV images of dynamic scenes with high accuracy. We leveraged a sliding window reconstruction where we assume that the scene does not change over eight frames (half the period of the polarizer) and then reconstruct a frame. Each of the scenes has a resolution of 290 × 290 pixels with each pixel measuring 24 µm. Here we show reconstructed images from two videos with dynamic motion including (a) humans waving their hands and (b) a parking lot with several cars. If the mixing ratios are known *a priori*, then we can employ a simple least-squares-based solution that enables reconstruction in real-time, shown in the third and fourth rows. See Visualization 1, Visualization 2, and Visualization 4.

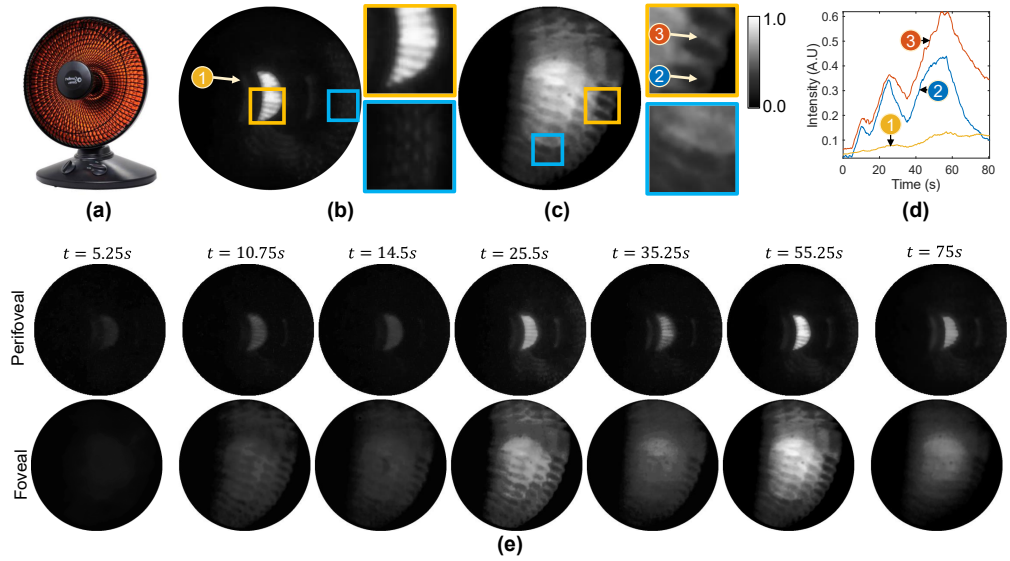

Fig. 4. Imaging thermodynamics at high spatial and temporal resolutions. We switched a coil-based heater (a) on for 5 s, then switched off for 5 s, on for 10 s, off for 10 s, on for 20 s, and finally switched off in sequence, resulting in heating and cooling periods, and then reconstructed (b) perifoveal and (c) foveal images. We can observe the complete heater, including the high frequency components of the parabolic reflector in the perifoveal images, as well as the coil in the foveal images. (d) Transients over three marked areas show (1) indirect heating of the heating hub, (2) indirect heating close to the coil, and (3) direct heating of the coil. See Visualization 3 and Visualization 4.

During the capture, the cars in the front were static, while a car in the background turned left and exited. The perifoveal images capture the complete scene with all three cars, while the foveal images show details of each car in each snapshot. If the mixing ratios α_t are known *a priori*, such as in the case when the polarizer rotates synchronously with the sensor exposure duration, then

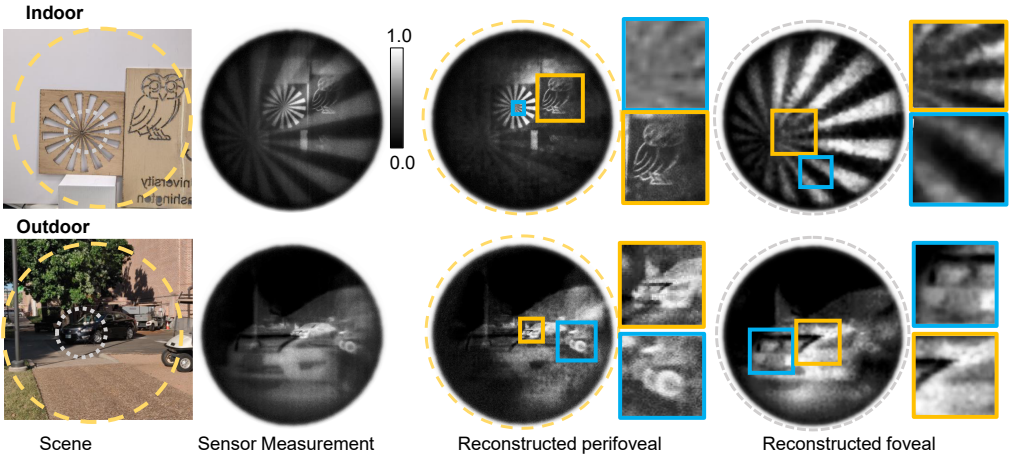

Fig. 5. Snapshot recovery without polarizer. Our setup is capable of working without a rotating polarizer albeit with a small loss in reconstruction quality. The figure above shows an indoor and outdoor example. A single sensor measurement is shown with notable contributions from both lenses, followed by the reconstructed perifoveal and foveal images.

we can leverage a least-squares-based reconstruction (details in Supplement 1) without any deep prior. This enables a real-time reconstruction, visualized in the second row in Fig. 3. The computational time per frame was less than 20 ms of CPU time. This computation can be performed asynchronously while the camera is capturing images, thereby enabling a real-time reconstruction at 12 fps. Thanks to the low computational complexity of OpenCV [54] functions, the computational backend can be implemented on a low power computational platform such as a Raspberry-Pi, creating a compact imaging system. The complete video sequence is available in Visualization 1 and Visualization 2.

As with dynamic motion, our setup can also image thermodynamic phenomena at high resolution. Figure 4(a) shows a portable heater going through three cycles of switching on and off (5 s, 10 s, and 20 s), producing a distinct transient visualized in Fig. 4(d). The effect of this heating cycle shows up in three forms. At spatial point 1, which is indirectly heated, the rise and fall in relative intensity are small, and smooth. At point 2, which is also indirectly heated but close to the heating coil, the transients are stronger than at point 1. At point 3 on the coil, the transients are the strongest. Point 1 is easy to distinguish in the perifoveal image that provides an overall view of the heater, including the high frequency components of the parabolic reflector, but does not resolve the fine coil structure. In contrast, the foveal image clearly shows the coils and enables evaluation of transients at very fine-grained spatial resolutions. The thermodynamics video sequence is available in Visualization 3.

C. Single Image Reconstruction

While polarization control enables high quality image separation, it may not always be feasible to add a polarizer in the optical system, such as imaging with very high frame rates. Our deep-prior-based computational backend is sufficiently powerful to work in such scenarios, where we can only capture a single frame with a combination of images from both lenses. There, we leverage the similarity between the downsampled foveal image, and the central crop of the perifoveal image to regularize the inverse problem. Given the perifoveal image $\mathbf{x}^1$ and the foveal image $\mathbf{x}^2$, we have

$$C\mathbf{x}^1 \approx D\mathbf{x}^2, \tag{7}$$

where C is the cropping operator, and D is the downsampling operator. We then solve the modified optimization problem

$$\min_{\mathbf{x}^1,\mathbf{x}^2} \| \mathbf{y} - (\mathbf{x}^1 + \mathbf{x}^2)\|^2 + \eta_1 \| C\mathbf{x}^1 - D\mathbf{x}^2\|^2 + \mathcal{R}(\mathbf{x}^1, \mathbf{x}^2), \tag{8}$$

where η_1 is the weight of penalty for similarity between the cropped perifoveal image and the downsampled foveal image. Further details about the regularizer are available in Supplement 1.

Figure 5 shows an indoor example with a wooden stencil and an outdoor example with a parked car. The indoor scene has a perifoveal image consisting of the sector star target, as well as an owl, while the foveal image zooms into the sector star. The foveal image clearly shows the sector at high resolution. The sensor measurement for the outdoor scene has distinct contributions from both perifoveal (small car, and golf cart) and foveal (enlarged car) images. The reconstructed images show the surroundings in the perifoveal image including the wheels of a golf cart, while the foveal image shows the features of the car. Note that there is some overlap in the reconstructions. The ill-posedness of the inverse problem produces artifacts in both perifoveal and foveal images. It is possible to remove such artifacts with more advanced computational approaches including a trained neural network [55], which we leave for future work.

4. DISCUSSION

We have demonstrated a first-of-its-kind optical system based on polarization-sensitive metalenses that enables foveation. The optical setup along with a powerful computational backend is capable of imaging in the wild at real-time rates, opening up a wide range of possibilities in applications that require compact optical systems along with large field of view and high spatial resolution.

A. Limitations and Future Directions

Since our optical system relies on polarization to separate the two images, scenes with strongly polarized light (such as purely linearly polarized) will produce images only with one of the two lenses. In practice, most real images are composed of unpolarized or partially polarized light, and hence our approach is expected to work for a broad range of scenes.

Our metalens-based foveated imager currently produces sharp images only for a narrow range of wavelengths. This limitation is common across most metalenses [56]. Further, due to a simple hyperbolic phase function, the captured images are sharp at the center of the lens, but tend to be distorted towards the edges of the FoV (see Supplement 1 for further details). However, by co-optimizing the meta-optics and computational backend, the metalenses can be made to focus over a much broader range of wavelengths and over a much larger FoV as shown in the visible [44]. Foveated imaging with broadband performance will constitute a promising future direction. Our metalenses have moderate transmission efficiency of approximately 60%, due to a lack of antireflection (AR) coating and absorption in the silicon substrate. Combined with the usage of a polarizer, this invariably results in a reduced sensitivity (throughput approximately 18%). Our lab prototype achieved a noise-equivalent temperature difference (NETD) of 152 mK, while an optical system equipped with an off-the-shelf 25 mm objective achieved 71 mK, thereby resulting in a 2× lower sensitivity. This can be overcome with more advanced manufacturing procedures with AR coatings on the metalenses or using transparent materials in the LWIR range, such as As_2S_3 or GeSbSe thin-films on CaF_2 substrate [57].

Funding. Vannevar Bush Faculty Fellowship; Air Force Office of Scientific Research (FA9550-22-1-0060); National Science Foundation (CCF-1911094, IIS-1652633, IIS-1739574, IIS-1838177, IIS-2107313, IIS-1730574, NNCI-1542101, NNCI-2025489, NSF-2127235, NSF-2127235); Office of Naval Research (MURI N00014-20-1-2787, N00014-18-1-2047, N00014-18-12571, N00014-20-1-2534).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but can be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content. Raw videos for Visualization 1, Visualization 2 and Visualization 3 are available in Visualization 4.

REFERENCES

1. S. Thiele, K. Arzenbacher, T. Gissibl, H. Giessen, and A. M. Herkommer, "3D-printed eagle eye: compound microlens system for foveated imaging," Sci. Adv. **3**, e1602655 (2017).
2. G. Carles, S. Chen, N. Bustin, J. Downing, D. McCall, A. Wood, and A. R. Harvey, "Multi-aperture foveated imaging," Opt. Lett. **41**, 1869–1872 (2016).
3. E. J. Tremblay, I. Stamenov, R. D. Beer, A. Arianpour, and J. E. Ford, "Switchable telescopic contact lens," Opt. Express **21**, 15980–15986 (2013).
4. G. Carles, J. Babington, A. Wood, J. F. Ralph, and A. R. Harvey, "Superimposed multi-resolution imaging," Opt. Express **25**, 33043–33055 (2017).
5. S. M. Kamali, E. Arbabi, A. Arbabi, and A. Faraon, "A review of dielectric optical metasurfaces for wavefront control," Nanophotonics **7**, 1041–1068 (2018).
6. A. Zhan, S. Colburn, C. M. Dodson, and A. Majumdar, "Metasurface freeform nanophotonics," Sci. Rep. **7**, 1673 (2017).
7. Z. Han, S. Colburn, A. Majumdar, and K. F. Böhringer, "Millimeter-scale focal length tuning with MEMS-integrated meta-optics employing high-throughput fabrication," Sci. Rep. **12**, 5385 (2022).
8. L. Huang, Z. Coppens, K. Hallman, Z. Han, K. F. Böhringer, N. Akozbek, A. Raman, and A. Majumdar, "Long wavelength infrared imaging under ambient thermal radiation via an all-silicon metalens," Opt. Mater. Express **11**, 2907–2914 (2021).
9. E. Arbabi, A. Arbabi, S. M. Kamali, Y. Horie, M. Faraji-Dana, and A. Faraon, "MEMS-tunable dielectric metasurface lens," Nat. Commun. **9**, 812 (2018).
10. S. Vo, D. Fattal, W. V. Sorin, Z. Peng, T. Tran, M. Fiorentino, and R. G. Beausoleil, "Sub-wavelength grating lenses with a twist," IEEE Photon. Technol. Lett. **26**, 1375–1378 (2014).
11. P. R. West, J. L. Stewart, A. V. Kildishev, V. M. Shalaev, V. V. Shkunov, F. Strohkendl, Y. A. Zakharenkov, R. K. Dodds, and R. Byren, "All-dielectric subwavelength metasurface focusing lens," Opt. Express **22**, 26212–26221 (2014).
12. D. Lin, P. Fan, E. Hasman, and M. L. Brongersma, "Dielectric gradient metasurface optical elements," Science **345**, 298–302 (2014).
13. F. Lu, F. G. Sedgwick, V. Karagodsky, C. Chase, and C. J. Chang-Hasnain, "Planar high-numerical-aperture low-loss focusing reflectors and lenses using subwavelength high contrast gratings," Opt. Express **18**, 12606–12614 (2010).
14. A. Arbabi, E. Arbabi, Y. Horie, S. M. Kamali, and A. Faraon, "Planar metasurface retroreflector," Nat. Photonics **11**, 415–420 (2017).
15. Q. Fan, P. Huo, D. Wang, Y. Liang, F. Yan, and T. Xu, "Visible light focusing flat lenses based on hybrid dielectric-metal metasurface reflector-arrays," Sci. Rep. **7**, 45044 (2017).
16. D. Fattal, J. Li, Z. Peng, M. Fiorentino, and R. G. Beausoleil, "Flat dielectric grating reflectors with focusing abilities," Nat. Photonics **4**, 466–470 (2010).
17. Y. Ran, J. Liang, T. Cai, and H. Li, "High-performance broadband vortex beam generator using reflective Pancharatnam–Berry metasurface," Opt. Commun. **427**, 101–106 (2018).
18. F. Yue, D. Wen, J. Xin, B. D. Gerardot, J. Li, and X. Chen, "Vector vortex beam generation with a single plasmonic metasurface," ACS Photon. **3**, 1558–1563 (2016).
19. X. Ma, M. Pu, X. Li, C. Huang, Y. Wang, W. Pan, B. Zhao, J. Cui, C. Wang, Z. Zhao, and X. Luo, "A planar chiral meta-surface for optical vortex generation and focusing," Sci. Rep. **5**, 10365 (2015).
20. Y. Yang, W. Wang, P. Moitra, I. I. Kravchenko, D. P. Briggs, and J. Valentine, "Dielectric meta-reflectarray for broadband linear polarization conversion and optical vortex generation," Nano Lett. **14**, 1394–1399 (2014).
21. H. Ren, X. Fang, J. Jang, J. Bürger, J. Rho, and S. A. Maier, "Complex-amplitude metasurface-based orbital angular momentum holography in momentum space," Nat. Nanotechnol. **15**, 948–955 (2020).
22. Y. Hu, L. Li, Y. Wang, M. Meng, L. Jin, X. Luo, Y. Chen, X. Li, S. Xiao, H. Wang, and Y. Luo, "Trichromatic and tripolarization-channel holography with noninterleaved dielectric metasurface," Nano Lett. **20**, 994–1002 (2019).
23. G. Zheng, H. Mühlenbernd, M. Kenney, G. Li, T. Zentgraf, and S. Zhang, "Metasurface holograms reaching 80% efficiency," Nat. Nanotechnol. **10**, 308–312 (2015).
24. A. Arbabi, E. Arbabi, M. Mansouree, S. Han, S. M. Kamali, Y. Horie, and A. Faraon, "Increasing efficiency of high numerical aperture metasurfaces using the grating averaging technique," Sci. Rep. **10**, 7124 (2020).
25. P. Lalanne, S. Astilean, P. Chavel, E. Cambril, and H. Launois, "Blazed binary subwavelength gratings with efficiencies larger than those of conventional échelette gratings," Opt. Lett. **23**, 1081–1083 (1998).
26. C. M. V. Burgos, T. Yang, Y. Zhu, and A. N. Vamivakas, "Design framework for metasurface optics-based convolutional neural networks," Appl. Opt. **60**, 4356–4365 (2021).
27. A. H. Dorrah, N. A. Rubin, A. Zaidi, M. Tamagnone, and F. Capasso, "Metasurface optics for on-demand polarization transformations along the optical path," Nat. Photonics **15**, 287–296 (2021).
28. A. Arbabi, Y. Horie, M. Bagheri, and A. Faraon, "Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission," Nat. Nanotechnol. **10**, 937–943 (2015).
29. N. A. Rubin, G. D'Aversa, P. Chevalier, Z. Shi, W. T. Chen, and F. Capasso, "Matrix Fourier optics enables a compact full-Stokes polarization camera," Science **365**, eaax1839 (2019).
30. Q. Guo, Z. Shi, Y.-W. Huang, E. Alexander, C.-W. Qiu, F. Capasso, and T. Zickler, "Compact single-shot metalens depth sensors inspired by eyes of jumping spiders," Proc. Natl. Acad. Sci. USA **116**, 22959–22965 (2019).
31. S. Tian, H. Guo, J. Hu, and S. Zhuang, "Dielectric longitudinal bifocal metalens with adjustable intensity and high focusing efficiency," Opt. Express **27**, 680–688 (2019).
32. Z. Yao and Y. Chen, "Focusing and imaging of a polarization-controlled bifocal metalens," Opt. Express **29**, 3904–3914 (2021).

33. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," ACM Trans. Graph. **26**, 70–80 (2007).
34. R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," in *ACM SIGGRAPH* (2006).
35. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," IEEE Signal Process. Mag. **25**(2), 83–91 (2008).
36. A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk, "CS-MUVI: video compressive sensing for spatial-multiplexing cameras," in *IEEE International. Conference on Computational Photography (ICCP)* (2012).
37. Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2011).
38. A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," ACM Trans. Graph. **26**, 69 (2007).
39. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**, 739–745 (2013).
40. A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," Appl. Opt. **47**, B44–B51 (2008).
41. C. Saunders, J. Murray-Bruce, and V. K. Goyal, "Computational periscopy with an ordinary digital camera," Nature **565**, 472–475 (2019).
42. D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. Wong, and J. H. Shapiro, "Photon-efficient imaging with a single-photon camera," Nat. Commun. **7**, 12046 (2016).
43. S.-H. Baek, H. Ikoma, D. S. Jeon, Y. Li, W. Heidrich, G. Wetzstein, and M. H. Kim, "Single-shot hyperspectral-depth imaging with learned diffractive optics," in *IEEE International Conference on Computer Vision (ICCV)* (2021).
44. E. Tseng, S. Colburn, J. Whitehead, L. Huang, S.-H. Baek, A. Majumdar, and F. Heide, "Neural nano-optics for high-quality thin lens imaging," Nat. Commun. **12**, 6493 (2021).
45. Y. Li, M. Qi, R. Gulve, M. Wei, R. Genov, K. N. Kutulakos, and W. Heidrich, "End-to-end video compressive sensing using anderson-accelerated unrolled networks," in *IEEE International Confernce on Computational Photography (ICCP)* (2020).
46. V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," ACM Trans. Graph. **37**, 1–13 (2018).
47. A. Chakrabarti, W. T. Freeman, and T. Zickler, "Rethinking color cameras," in *IEEE International Conference on Computational Photography (ICCP)* (2014).
48. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).
49. V. Liu and S. Fan, "S4: a free electromagnetic solver for layered periodic structures," Comput. Phys. Commun. **183**, 2233–2244 (2012).
50. Q. Fan, M. Liu, C. Yang, L. Yu, F. Yan, and T. Xu, "A high numerical aperture, polarization-insensitive metalens for long-wavelength infrared imaging," Appl. Phys. Lett. **113**, 201104 (2018).
51. M. Meem, S. Banerji, A. Majumder, F. G. Vasquez, B. Sensale-Rodriguez, and R. Menon, "Broadband lightweight flat lenses for long-wave infrared imaging," Proc. Natl. Acad. Sci. USA **116**, 21375–21378 (2019).
52. J. Li, Y. Wang, S. Liu, T. Xu, K. Wei, Y. Zhang, and H. Cui, "Largest aperture metalens of high numerical aperture and polarization independence for long-wavelength infrared imaging," Opt. Express **30**, 28882–28891 (2022).
53. Y. Gandelsman, A. Shocher, and M. Irani, "Double-dip: Unsupervised image decomposition via coupled deep-image-priors," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).
54. G. Bradski, "The OpenCV Library," Dr. Dobb's J. Softw. Tools **120**, 122–125 (2000).
55. X. Zhang, R. Ng, and Q. Chen, "Single image reflection separation with perceptual losses," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).
56. F. Presutti and F. Monticone, "Focusing on bandwidth: achromatic metalens limits," Optica **7**, 624–631 (2020).
57. B. J. Eggleton, B. Luther-Davies, and K. Richardson, "Chalcogenide photonics," Nat. Photonics **5**, 141–148 (2011).