Lensless Imaging: A Computational Renaissance

Vivek Boominathan,[R] *Jesse K. Adams,*[R] *M. Salman Asif,*[R] *Ben Avants,*[R] *Jacob T. Robinson,*[R]
Richard G. Baraniuk,[R] *Aswin C. Sankaranarayanan,*[C] *Ashok Veeraraghavan*[R]
[R]Rice University, [C]Carnegie Mellon University

The basic design of a camera has remained unchanged for centuries. To acquire an image, light from the scene under view is focused onto a photosensitive surface using a lens. Over the years, the photosensitive surface has evolved from a photographic film to an array of digital sensors. However, lenses remain an integral part of modern imaging systems in a broad range of applications.

Unfortunately, lenses also introduce a number of limitations. First, while image sensors are typically thin, cameras end up being thick due to the lens complexity and the large distance required between the lens and sensor to achieve focus. For example, the thinnest mobile cameras today are approximately 5 mm thick, with the thickness increasing at larger lens aperture sizes. Second, lenses for visible light can be manufactured with inexpensive materials such as glass and plastic, but lenses for wavelengths farther into the infrared and ultraviolet are either extremely expensive or infeasible. Third, lens-based cameras invariably require post-fabrication assembly, resulting in manufacturing inefficiencies.

In this paper, we review a variety of alternate imaging approaches that completely eschew lenses. The primary task of a lens in a camera is to shape the incoming light wavefront so that it creates a focused image on the sensor. In the absence of a lens, a sensor would simply record the average light intensity from the entire scene. *Lensless imaging systems* dispense with a lens by using other optical elements to manipulate the incoming light. The sensor records the intensity of the manipulated light, which may not appear as a focused image. However, when the system is designed correctly, the image can be recovered from the sensor measurements with the help of a computational algorithm. Figure 1 shows the processes for capturing/reconstructing images in lensed and lensless systems. The simplest lensless imaging system is the pinhole camera. It is inefficient, however, since the small pinhole restricts the amount of light reaching the sensor. Coded aperture cameras improve the light efficiency using a mask with an array of pinholes. The sensor measurements become a superposition of the images formed by each aperture, and the computational recovery algorithm's task is to reorganize the measurements to recover the image.

There are many benefits to going lensless:

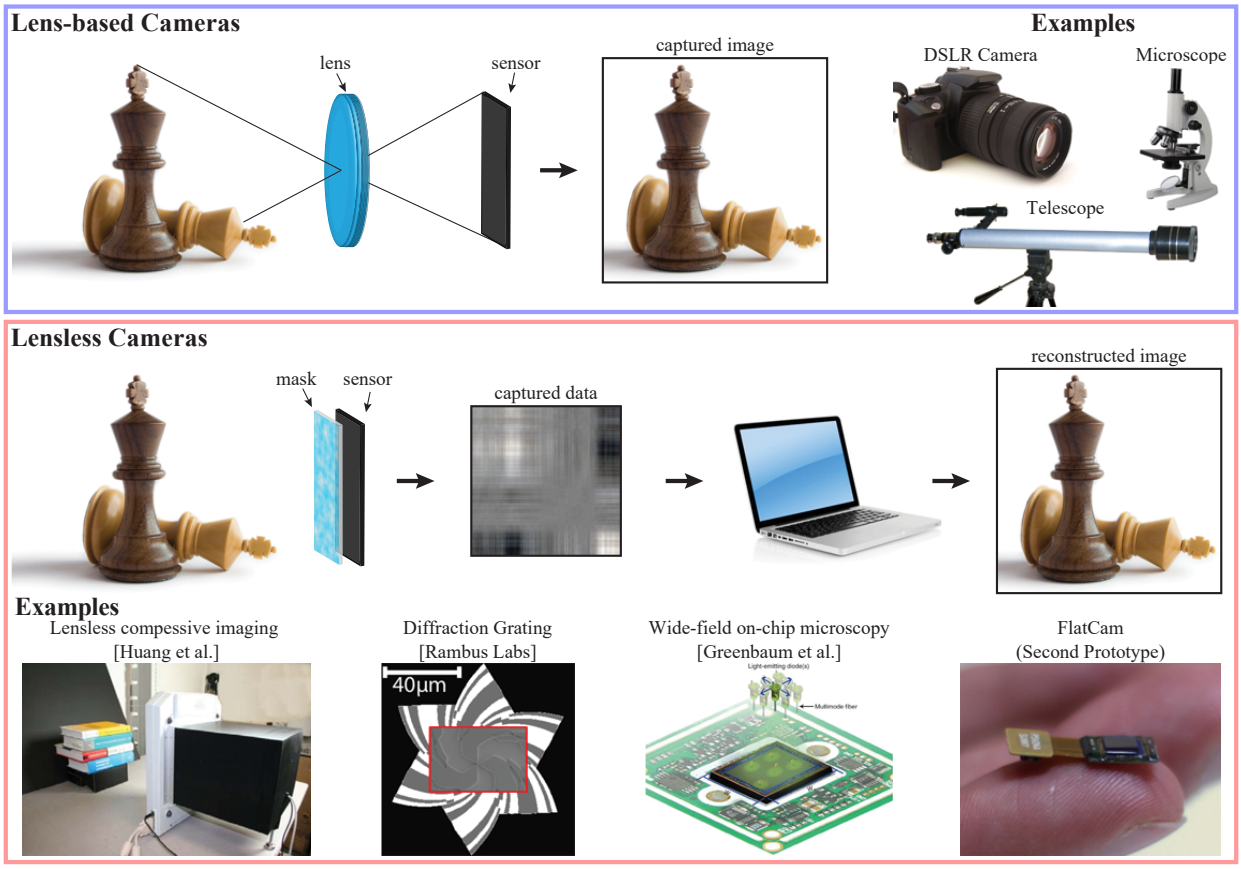

Fig. 1: *Lensed vs Lensless imaging. (Top) Illustration of a lens-based camera where a lens maps the scene on to a sensor to form a clear image. A few examples of lens-based systems are shown. (Bottom) The process of capturing an image using a lensless camera. An additional step of computation is required to reconstruct a clear image from the muddled sensor data. A few examples of lensless cameras ([11], [16], [24] and our prototype based on [31]) are also shown.*

- **Scalable fabrication.** Lensless cameras can be directly fabricated using traditional semiconductor fabrication technology. For example, a multiple-aperture mask can be fabricated either directly in one of the metal interconnect layers or on a separate wafer thermal compression that is bonded to the back side of the sensor, as is typical for back-side-illuminated image sensors [1]. Thus, lensless cameras can benefit from all of the scaling advantages of semiconductor fabrication, resulting in a low-cost, high-yield, high-performance device. In contrast, conventional cameras require inefficient post-fabrication assembly of the lens system.
- **Thin form factor.** Since the standoff distance between a multiple-aperture mask and the image sensor array need only be a few tens to hundreds of microns, an entire lensless camera can be only a few

tens to hundreds of microns thick—resulting in potentially the thinnest cameras ever produced.

- **Wavelength scaling.** Lensless cameras have been used for X-ray, gamma-ray, and astronomical imaging for decades. Lensless imaging in the visible, short-wave infrared (SWIR), and thermal wavebands is relatively new. Moreover, the technology can also be expanded to the mm-wave, terahertz, and other bands with minimal modifications, providing unmatched spectral flexibility.
- **Low cost.** While the cost of high-resolution cameras has fallen rapidly in the visible range, it remains high outside the visible range (e.g., infrared). One reason is that lenses for these wavelengths must be manufactured using expensive materials. By doing away with lenses and the need for post-fabrication assembly, lensless cameras promise significant cost reductions for imaging outside the visible spectrum.
- **Non-planar geometries.** Lensless cameras can be adapted to arbitrary sensor geometries, including not just planar but also cylindrical, spherical, and even flexible sensors. The compact form of spherical lensless cameras promise unmatched maneuverability in constrained environments such as endoscopy.
- **Light throughput.** Lensless cameras can be designed to have very large input apertures, which translates into improved light efficiency and much larger field-of-view than conventional lens-based systems.
- **Three-dimensional (3D) imaging.** Lensless imaging systems can extract 3D and refocusing information in addition to 2D imaging. Although this ability is not yet competitive with existing lens based techniques such as light-field and time of flight, the extracted 3D information may still be useful in some contexts such as gesture identification.

In this paper, we review the past, present, and future of lensless imaging as a shining example of the opportunities afforded by *computational imaging*, a design framework that uses computational algorithms to replace or augment imaging hardware (in this case replacing the lens). After reviewing classical and contemporary approaches to lensless imaging, we introduce and analyze a mathematical model that exposes the key issues underlying these architectures. The bulk of the paper consists of a case study of the *FlatCam*, a particular mask-based lensless imager we have developed.

EARLY LENSLESS IMAGING SYSTEMS

Pinhole cameras

The very first cameras were lensless. Pinhole cameras, also known as the *camera obscura*, were discovered centuries before the invention of lenses and photography. Pinhole cameras have been well known since Alhazen (965-1039 AD) and Mozi (c. 370 BCE). However, the first photograph using a pinhole camera was captured in 1850. Pinhole cameras offer a simple and elegant architecture for lensless imaging that consists of a single aperture in front of a sensor. Light from an object passes

through the pinhole and forms its image on the sensor. However, a tiny pinhole is required to produce sharp images, which results in very low light throughput. As a consequence, a pinhole camera requires very long exposure times to acquire images at high quality. Indeed, lenses were introduced into cameras for precisely the purpose of increasing the size of the aperture, and thus the light throughput, without degrading the sharpness of the acquired image.

BOX 1 (including Fig. 9) TO BE HERE

Coded aperture cameras

Coded aperture cameras extend the idea of a pinhole camera by replacing the small, single aperture with a mask containing multiple apertures [2]–[4]. Coded aperture cameras were originally invented for imaging with X-rays and gamma-rays, wavelengths of light that are not easily amenable to lens-based imaging (see Box 1). In a general coded aperture system, sensor measurements represent a superposition of the images formed behind each pinhole. The primary motivation for a coded mask is to increase the light throughput while retaining the ability to reconstruct high-resolution images. For instance, if the mask contains P pinholes, then the sensor image is the sum of P overlapping images of the scene. The signal-to-noise ratio in such an image is approximately $\sqrt{P}$ times better than a single pinhole image [2], [3].

In contrast to a single-pinhole camera, the sensor measurements of a coded aperture camera do not resemble an image of the scene. Rather, each light source in the scene casts a unique shadow of the mask onto the sensor, encoding information about locations and intensities. Consider a single light source on a dark background; the image formed on the sensor will be a shadow of the mask. If we change the angle of the light source, then the mask shadow on the sensor will shift. If we change the depth of the light source, then the size of the shadow will change. We can represent the relationship between the scene and the sensor measurements as a linear system that depends on the pattern and placement of the mask. Inverting this system using an appropriate computational algorithm will recover an image of the scene.

The design of the mask plays an important role in coded-aperture imaging. An ideal pattern would maximize the light throughput while providing a well-conditioned scene-to-sensor transfer function to facilitate inversion. In this regard, several mask designs have been proposed in coded aperture literature, including Fresnel zone plate, random pinhole patterns, uniformly redundant arrays (URAs) [3], and their extensions. URAs are particularly useful because of two key properties: (1) almost half of the mask is open, which boosts the signal-to-noise ratio; (2) the autocorrelation function of the mask is close to a delta function, which aids in calibration and image recovery. URA patterns are closely related to the

Hadamard-Walsh functions and the maximum length sequences that are maximally incoherent with their cyclic shifts [5].

Zone plates

A zone plate can also be used to focus light and form an image using diffraction [6], [7]. A zone plate consists of concentric transparent and opaque rings (or zones). Light hitting a zone plate diffracts around the opaque regions and interferes constructively at the focal point. Zone plates can be used in place of pinholes or lenses to form an image. One advantage of zone plates over pinholes is their large transparent area, which provides better light efficiency. In contrast with lenses, zone plates can be used for imaging wavelengths where lenses are either expensive or difficult to manufacture [8], [9].

CONTEMPORARY LENSLESS IMAGING SYSTEMS

Recent advances in sensor technology (in particular, the conversion from analog film to digital CCD and CMOS sensor arrays), image reconstruction models and algorithms, and computing resources have made lensless imaging a burgeoning field. Here we briefly review some of the recent research in this area.

Lensless imaging using programmable apertures

Programmable mask-based lensless imaging designs have recently been proposed in [10]–[12]. The camera proposed in [10] consists of a sensor and layers of programmable spatial light modulators (SLMs) whose transmittances are controllable in space and time. By applying different patterns in each layer, the incoming light can be manipulated in a number of ways. For example, the camera can track a moving object by shifting a pinhole in one of the layers, select and capture disjoint regions in the scene, or perform computations on the scene and record the results directly on the sensor.

The lensless camera in [11] (first example of lensless in Figure 1) uses compressive sensing principles to capture and recover images. It consists of a single programmable SLM and a single pixel detector. It captures multiple measurements of the scene by changing the mask pattern. The scene is then reconstructed by solving a sparse recovery program. Using multiple pixel detectors, this design can reconstruct a higher resolution image for a planar or a sufficiently distant scene [13].

The camera in [12] consists of a sensor array and an SLM implementing a separable mask pattern. This camera can reconstruct the scene using a single sensor image, but the reconstruction quality improves using multiple sensor images with different mask patterns. In the development of this camera, the authors showed that traditional techniques [3] of using URA and modified URA (MURA) aperture patterns fail due to significant diffraction effects in the visible spectrum.

Ultra-miniature lensless imaging with diffraction gratings

Ultra-miniature cameras (approximately $100\,\mu$m width and thickness) have been implemented in [14]–[17] using integrated diffraction gratings and CMOS image sensors. The pixels in [14] use diffraction gratings over a photodiode in order to be sensitive to the angle of incident light. The angle selectivity is achieved due to a phenomenon called the Talbot effect [18] and enables the camera to perform lensless 3D imaging in the near-field. The gratings were fabricated as metal wiring layers over the photodiodes.

The phase gratings in [16] are designed such that they impose spiral-shaped diffraction patterns (second example for lensless in Figure 1) on the sensor array. The diffraction pattern is etched on a refractive medium placed above the sensor. The spiral pattern can also be viewed as the point spread function of these imaging systems. Similar to a coded aperture system, the image formed on the sensor is a superposition of shifted and scaled spiral patterns. However, in contrast to an amplitude mask, a phase grating-based mask has improved light efficiency, since it blocks much less light. While an image of the scene can be recovered using a computational algorithm, the primary purpose of these small-size and low-cost designs is distributed monitoring and inspection (for example, in the internet-of-things).

Lensless microscopy via shadow and diffraction imaging

Lensless cameras have also been successfully demonstrated for several microscopy and lab-on-chip applications. We can divide the lensless microscopes into two broad categories: contact-mode shadow imaging-based microscopes [19]–[21] and diffraction-based lensless microscopes [22]–[27]. In a shadow imaging-based microscope, a microscopic sample is placed extremely close to a sensor array (ideally within $1\,\mu$m) so that diffraction is minimized. Light from an illumination source passes through the sample and casts a shadow on the sensor with unit magnification. The shadow image represents the image of the microscopic sample under observation. It is also possible to capture multiple images of a sample with subpixel shifts for the purpose of digital superresolution. The on-chip microscope in [20] demonstrated imaging of red blood cells at a resolution of 600 nm by combining multiple low-resolution shadow images of blood flowing in a microfluidic channel.

Diffraction-based lensless microscopes allow a significant distance between the sample and the sensor plane. Light scattered by the sample interferes with itself and creates an interference pattern on the sensor (third example of lensless in Figure 1). These interference patterns can be digitally processed to reconstruct an image of the sample [24], [25]. The on-chip microscope in [25] demonstrated imaging of red blood cells at a resolution less than $7\,\mu$m with a field-of-view of 20.5 mm^2 . Since the optical sensor records only the intensity of the interference patterns and loses the phase information, image reconstruction relies on computational methods for phase retrieval [28], [29].

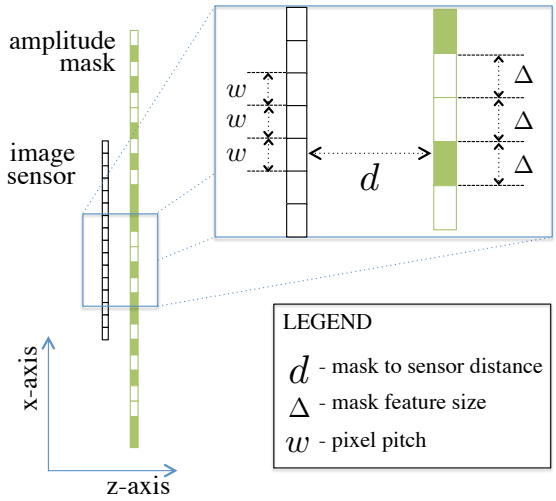

Fig. 2: *Schematic of a lensless imager using a single amplitude mask.*

A Mathematical Model for Lensless Imaging

A simple mathematical model can be used to explain, characterize, and analyze the operation of a variety of lensless imagers.

Lensless imaging architecture

Consider the imaging architecture in Figure 2, which consists of an amplitude mask placed in front of an image sensor. Both the sensor and the mask are assumed to be planar and parallel to each other. The mask is placed a distance d (typically measured in microns) in front of the sensor; hence, we can assume the sensor is placed on the plane $z = 0$ and the mask on the plane $z = d$. Assume without loss of generality that the mask is binary-valued and consists of opaque and transparent elements that either block or transmit light. An important variable is the smallest feature size on the mask, Δ; intuitively, the binary mask is constructed using opaque or transparent building blocks of size $\Delta \times \Delta$. Denote the pixel pitch, or the size of individual sensor pixels, by w. Given this basic setup, we can characterize the spot size produced by a mask element and characterize when the spot can be well-approximated using geometric (ray) optics.

Image formation

We characterize image formation using the geometric optics model. While this approach largely ignores diffraction, the resulting model is useful for the design and analysis of well-conditioned imaging architectures. Furthermore, the calibration procedure that we detail in subsequent sections can account for

un-modeled diffraction effects. For the simplicity of notation, we assume a simplified 2D world imaged by a 1D mask and sensor. The extension to a 3D world imaged by a 2D mask and sensor is straightforward except where stated otherwise.

For a suitably defined scene irradiance vector $\mathbf{x} \in \mathbb{R}^N$, the scene-to-sensor mapping can be described using the linear set of equations

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\Phi \in \mathbb{R}^{M\times N}$ is the measurement matrix, $\mathbf{y} \in \mathbb{R}^M$ is the image formed on the sensor, and $\mathbf{e}$ is measurement noise. This model can be interpreted in two different ways: (1) Each sensor measures a weighted, linear combination of light from multiple scene locations, and each row in Φ encodes the weights for the respective sensor. For a scene at infinity, the weights for two different sensor pixels simply differ by a translation of the mask pattern. As a consequence, the matrix Φ has a *Toeplitz* structure. (2) Every light source in the scene casts a shadow of the mask on the sensor. Thus, the image formed on the sensor is a superposition of shifted and scaled versions of the mask. The shift and the scaling of the mask pattern encodes the angle and distance of the light source onto the sensor. These properties are invaluable in the design of masks that provide near-optimal recovery under noise. Given the image formation model in (1), our tasks are to formulate an inversion algorithm that recovers the scene $\mathbf{x}$ from the sensed image $\mathbf{y}$ and design mask patterns that achieve optimal recovery performance. We study both problems in the subsequent sections.

Image reconstruction

Given the sensor measurements $\mathbf{y} \in \mathbb{R}^M$ and the measurement matrix Φ, recovering $\mathbf{x} \in \mathbb{R}^N$ depends mainly on the rank of the matrix Φ and its condition number. When $\text{rank}(\Phi) = N$ and the matrix is well-conditioned, we can obtain an estimate of $\mathbf{x}$ by solving the least-squares problem

$$\min_{\mathbf{x}} \|\Phi\mathbf{x} - \mathbf{y}\|_2^2, \tag{2}$$

which has the closed form solution $\widehat{\mathbf{x}}_{\text{LS}} = \Phi^+\mathbf{y} = \mathbf{x} + \Phi^+\mathbf{e}$, where $\Phi^+ = (\Phi^T\Phi)^{-1}\Phi^T$ is the pseudoinverse of Φ. When Φ is not well-conditioned, the least squares estimate $\widehat{\mathbf{x}}_{\text{LS}}$ suffers from noise amplification. When Φ is rank-deficient, the matrix becomes singular and an estimate cannot be achieved.

In the ill-conditioned and rank-deficient cases, we can use an image prior to regularize the inverse problem. Specifically, instead of solving (2), we solve

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda\,\mathcal{R}(\mathbf{x}), \tag{3}$$

where $\|\mathbf{y} - \Phi\mathbf{x}\|^2$ quantifies the data fidelity, $\mathcal{R}(\mathbf{x})$ is a regularization term that enforces an image prior, and $\lambda > 0$ controls the trade off between fidelity and regularization. A popular choice for the regularizer that is useful for noise-suppression is *Tikhonov regularization* (a.k.a. ridge regression) via $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_2^2$.

Natural signals such as images and videos exhibit a host of geometric properties including sparse gradients and sparse coefficients in certain transform domains (e.g., Fourier or wavelets). By enforcing these geometric properties, we can suppress noise amplification as well as obtain unique solutions even when Φ is rank-deficient (i.e, $M < N$). A pertinent example for image reconstruction is the *total-variation* (TV) model, where the regularizer $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_{\text{TV}}$ corresponds to the TV of the image, which is computed from its gradients. Writing the scene $\mathbf{x}$ as the 2D image $\mathbf{x}(u, v)$ and defining $\mathbf{g}_u = D_u\mathbf{x}$ and $\mathbf{g}_v = D_v\mathbf{x}$ as the u- and v-components, respectively, of the spatial gradient of the image, the TV of the image is given by

$$\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_{\text{TV}} = \sum_{u,v} \sqrt{\mathbf{g}_u(u, v)^2 + \mathbf{g}_v(u, v)^2}.$$

The minimization (3) with a TV prior is convex and produces images with sparse gradients. A host of efficient techniques have been developed to obtain the solution. A range of even more realistic image models have been developed (e.g., [30]), but the resulting optimization might not be convex.

FLATCAM: A LENSLESS IMAGING CASE STUDY

To illustrate the design tradeoffs involved in a practical lensless camera design, we review the FlatCam [31], which was inspired by the coded aperture imaging principles pioneered in astronomical X-ray and Gamma-ray imaging [2]–[4], [32] (see Box 1).

Architectural overview

The FlatCam design achieves a large photosensitive area with a thin form factor by replacing the lens with a coded, binary mask. The thickness of the camera is minimized by placing the mask almost immediately on top of a bare conventional sensor array. The image formed on the sensor can be viewed as a superposition of many pinhole images. An illustration of the FlatCam design is presented in Figure 3. Light from all points in the scene passes through a coded mask and forms a multiplexed image on the sensor. A computational algorithm is used to recover the original light distribution of the scene from the sensor measurements.

The FlatCam design has many attractive properties besides its slim profile. First, since it reduces the thickness of the camera but not the area of the sensor, it can collect more light than a miniature, lens-based camera of the same thickness. The light collection ability of FlatCam is proportional to the size of

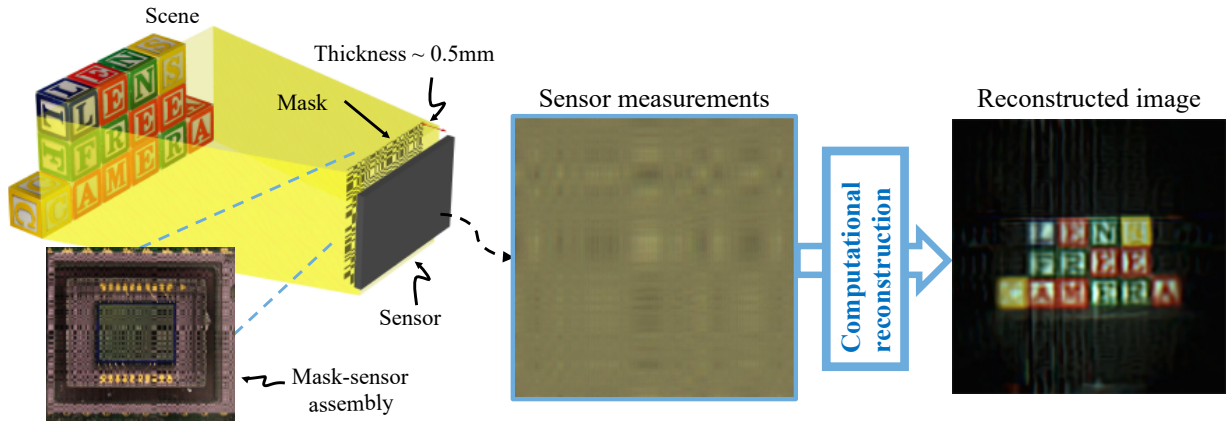

Fig. 3: *FlatCam architecture. Every light source within the camera field-of-view contributes to every pixel in the multiplexed image formed on the sensor. A computational algorithm reconstructs the image of the scene. (Left) Inset shows the mask-sensor assembly of our prototype, in which a binary, coded mask is placed 1.2 mm away from an off-the-shelf digital image sensor. (Middle) An example of sensor measurements. (Right) Image reconstructed by solving a computational inverse problem of the form (3). Figure modified from [31].*

the sensor and the transparent regions (pinholes) in the mask. In contrast, the light collection ability of a lens-based camera is limited by the lens aperture size, which is restricted by the requirements on the device thickness. Second, the mask can be created from inexpensive materials that operate over a broad range of wavelengths. Third, the mask can be fabricated simultaneously with the sensor array, creating new manufacturing efficiencies.

Mask design and calibration

Separable masks: The FlatCam uses a separable mask pattern, i.e., the 2D mask pattern is the outer product of two 1D patterns. Such a pattern drastically reduces the storage and computational footprint of the measurement matrix Φ. When the mask pattern is separable, the imaging equation (1) can be rewritten as

$$Y = \Phi_L X \Phi_R^T + E, \tag{4}$$

where X is an $N \times N$ matrix containing the scene radiance; Y in an $M \times M$ matrix containing the sensor measurements; Φ_L and Φ_R are matrices representing 1D convolution along the rows and columns of the scene, respectively; and E denotes the sensor noise and model mismatch. For a megapixel scene/image and a megapixel sensor, Φ_L and Φ_R each have only 10^6 elements, as opposed to 10^{12} elements in Φ. A similar idea has been recently proposed in [12] with the design of doubly Toeplitz mask.

Mask design: The mask pattern should be chosen to make the matrices Φ_L and Φ_R as numerically stable as possible, which ensures a stable recovery of the image X from the sensor measurements Y. In the context of image reconstruction using signal priors (for example, TV prior discussed above), random matrices enjoy stable recovery guarantees. Hence, we construct the separable mask pattern as the outer-product of two 1D pseudo-random sequences.

Calibration: The low-dimensionality of Φ_L and Φ_R in (4) support a simple and efficient calibration scheme. Instead of modeling the convolution shifts and diffraction effects for a particular mask-sensor arrangement, we directly estimate the system matrices from training data. To align the mask and sensor, we adjust their relative orientation such that a separable scene in front of the camera yields a separable image on the sensor. For a perfectly aligned system, displaying a horizontal/vertical line on a screen in front of the camera results in an image containing a set of sharp horizontal/vertical stripes. We first achieve sharpness by rotating the mask relative to the screen. Then, we align the sensor and mask so that the stripes on the sensor image are parallel to the image axis. To calibrate a system that can recover an image X with dimensions $N \times N$, we estimate the left and right matrices Φ_L and Φ_R using the sensor measurements of $2N$ known calibration patterns projected on a screen as depicted in Fig. 4. Our calibration procedure relies on an important observation: If the scene X is separable, i.e., $X = \mathbf{a}\mathbf{b}^T$ where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$ then, for an ideal system,

$$Y = \Phi_L \mathbf{a}\mathbf{b}^T \Phi_R^T = (\Phi_L \mathbf{a})(\Phi_R \mathbf{b})^T.$$

In essence, the image formed on the sensor is a rank-1 matrix, and using a truncated singular value decomposition (SVD), we can obtain estimates of $\Phi_L \mathbf{a}$ and $\Phi_R \mathbf{b}$ up to a signed, scalar constant. We take N separable pattern measurements for calibrating each of Φ_L and Φ_R. In practice, we average several measurements of each calibration pattern to reduce the effects of sensor noise.

Prototypes

We have built two different FlatCam prototypes. The first prototype consists of a Point Grey Flea3 with a Sapphire EV76C560 CMOS sensor which has a 5.3 μm pixel size and measured Chief Ray Angle (CRA) of 25°. (The CRA of a sensor determines the cone of light that can enter a pixel.) The diffractive mask is chrome on quartz glass placed adjacent to the infrared filter of the sensor (mask-to-sensor distance: 1.2 mm). The pattern on the mask is an outer product of two length-1024 pseudorandom sequences of smallest feature size 25 μm. Sample reconstructions using this prototype are shown in Figure 5 (Top). Reconstructions of a dynamic scene are shown in Figure 5 (Bottom); here, we operated the camera with

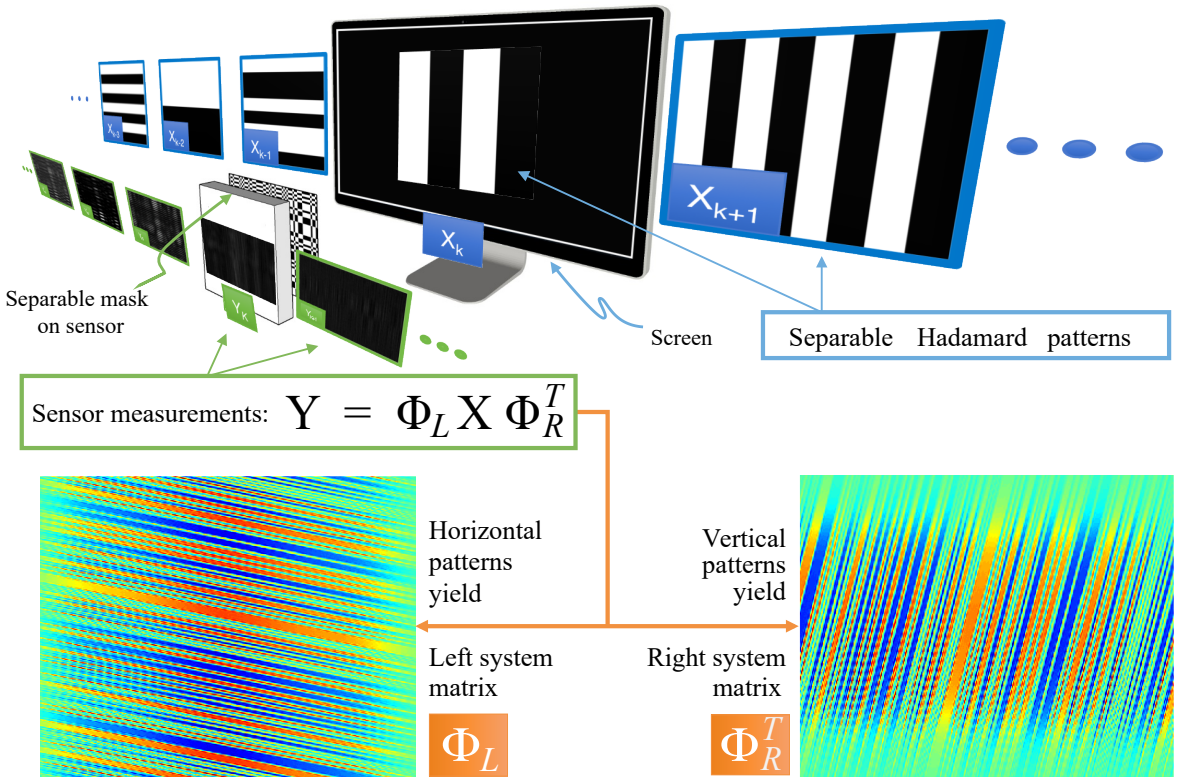

Fig. 4: *Calibration for measuring the left and right matrices Φ_L and Φ_R corresponding to a separable mask. (Top) Separable patterns displayed on a screen in front of the camera. The patterns are orthogonal, one-dimensional Hadamard codes that are repeated along either the horizontal or vertical direction. (Bottom) Estimated left and right matrices. Figure modified from [31].*

a 3 ms exposure and recovered videos of 60 frames-per-second with each frame of the video recovered independently as a stand-alone image.

The second prototype was assembled with a diffractive mask and spacer attached directly to the surface of an Omnivision OV5647 CMOS sensor (fourth example of lensless in Figure 1). The Omnivision sensor has pixels of size 1.4 μm and measured CRA of 28°. The diffractive mask was fabricated by depositing a thin-film of chrome on fused silica that was then patterned with photoresist and etched to leave the desired pattern. The mask was then diced, aligned to the CMOS pixel array and attached with optical epoxy (mask to sensor distance 500 μm). The pattern on the mask is the outer product of length-1296 and 972 pseudorandom sequences of smallest feature size of 2.8 μm. The smaller feature size and pixel pitch of this prototype enable reconstructions at a higher resolution. However, a drawback of the smaller pixel pitch is a sensor with poorer SNR performance which results in noisier measurements and reconstruction compared to our first prototype (see Figure 6).

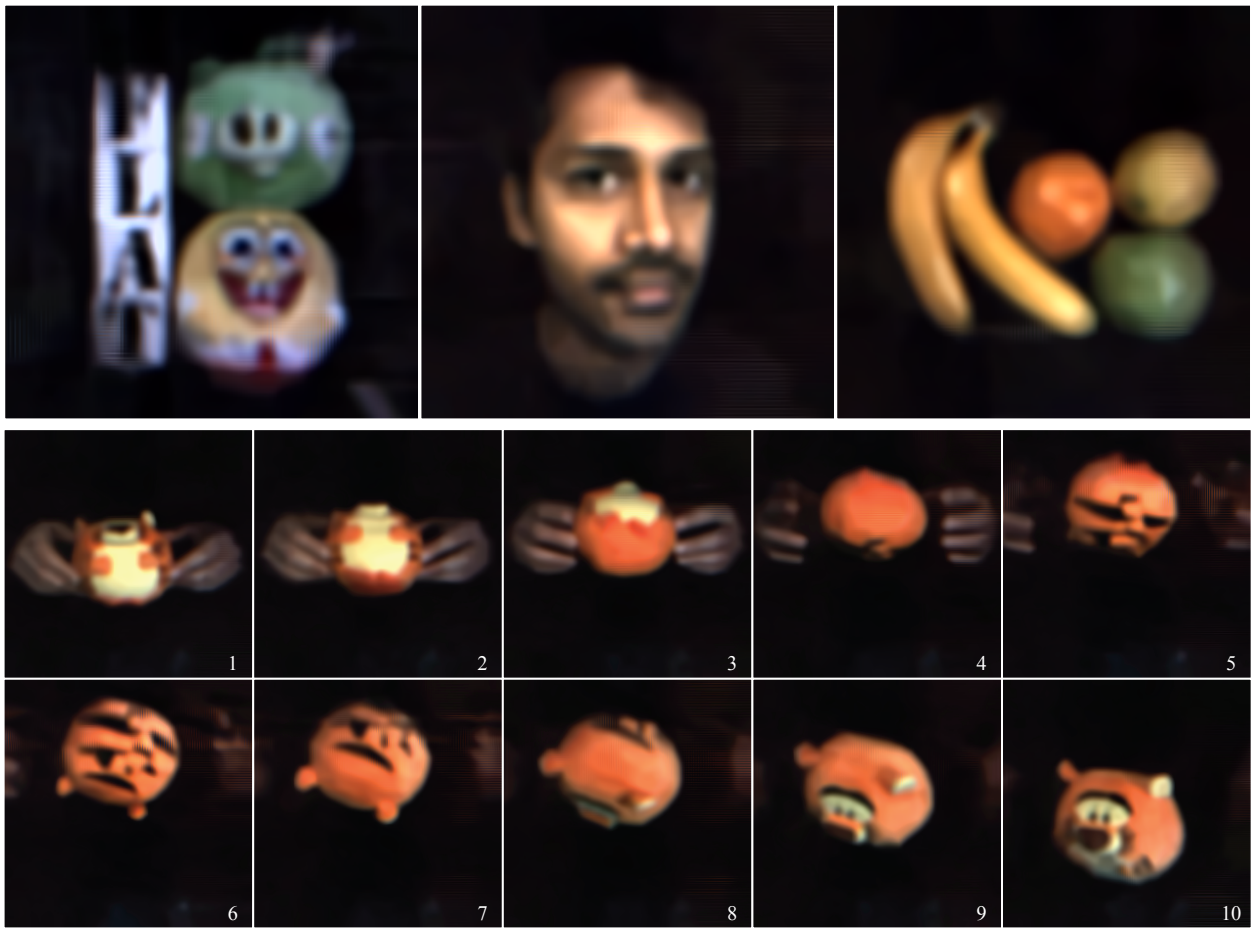

Fig. 5: *(Top) Reconstruction of three static scenes using the Flea FlatCam prototype. (Bottom) Sample frames from video reconstruction of a toy performing a backflip aided by human hands. The video was recorded at 60 frames-per-second.*

The remainder of the examples below were obtained using the higher-quality Flea3 sensor prototype.

Programmable masks

In many applications, a camera has the opportunity to acquire several images of a scene, and both folk wisdom and theory tell us that averaging the acquisitions should suppress noise. In contrast to a lens-based camera, a lensless camera could be equipped with a programmable mask that changes for each acquisition in order to provide diversity in the acquisitions. Presumably, such a camera should not only suppress noise but also average out imperfections in the measurement operator Φ (recall that Φ is never perfectly conditioned in a coded aperture system).

To demonstrate the potential of a programmable mask FlatCam, we simulated multi-image capture using

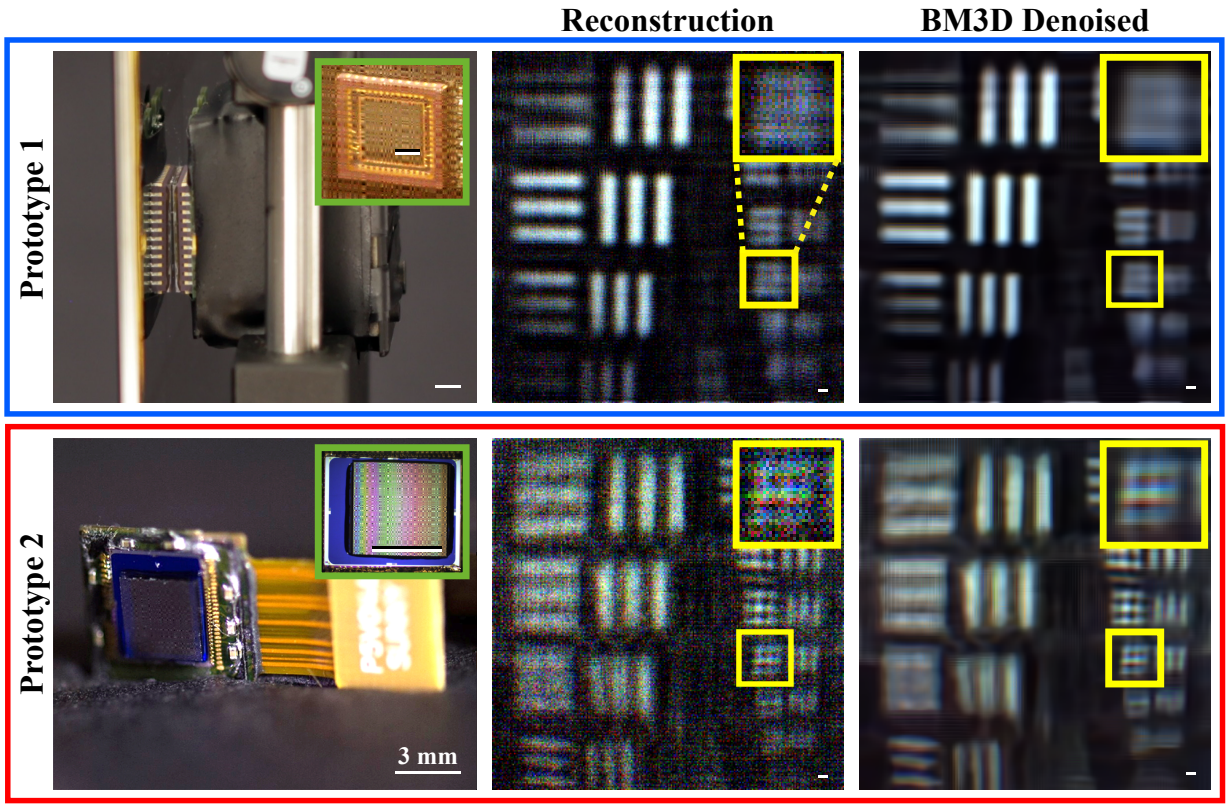

Fig. 6: *Flatcam prototype comparison. Top left is our first prototype with chrome mask placed directly in front of the Flea3 sensor; bottom left is our second prototype with the Omnivision sensor directly epoxied to mask (insets show close-up of the sensors and masks). To the right are reconstruction and BM3D denoised reconstructions for each prototype. The smaller feature size and pixel pitch of the Omnivision prototype provide superior resolution at the cost of more noisier image reconstruction.*

the sequences of separable measurement matrices $\{\Phi_{L_i}\}_{i=1,..,L}$ and $\{\Phi_{R_j}\}_{j=1,..,R}$. The measurements of the scene X using each pair of measurement matrices are given by

$$Y_{(i,j)} = \Phi_{L_i} X \Phi_{R_j}^T + E_{(i,j)}, \tag{5}$$

which can be stacked into the larger system

$$Y = \begin{bmatrix} Y_{(1,1)} & . & . & . & . \\ . & . & . & . & . \\ . & . & \{Y_{(i,j)}\} & . & . \\ . & . & . & . & . \\ . & . & . & . & Y_{(L,R)} \end{bmatrix}, \Phi_L = \begin{bmatrix} \Phi_{L_1} \\ . \\ \{\Phi_{L_i}\} \\ . \\ \Phi_{L_L} \end{bmatrix}, \Phi_R = \begin{bmatrix} \Phi_{R_1} \\ . \\ \{\Phi_{R_j}\} \\ . \\ \Phi_{R_R} \end{bmatrix} \tag{6}$$

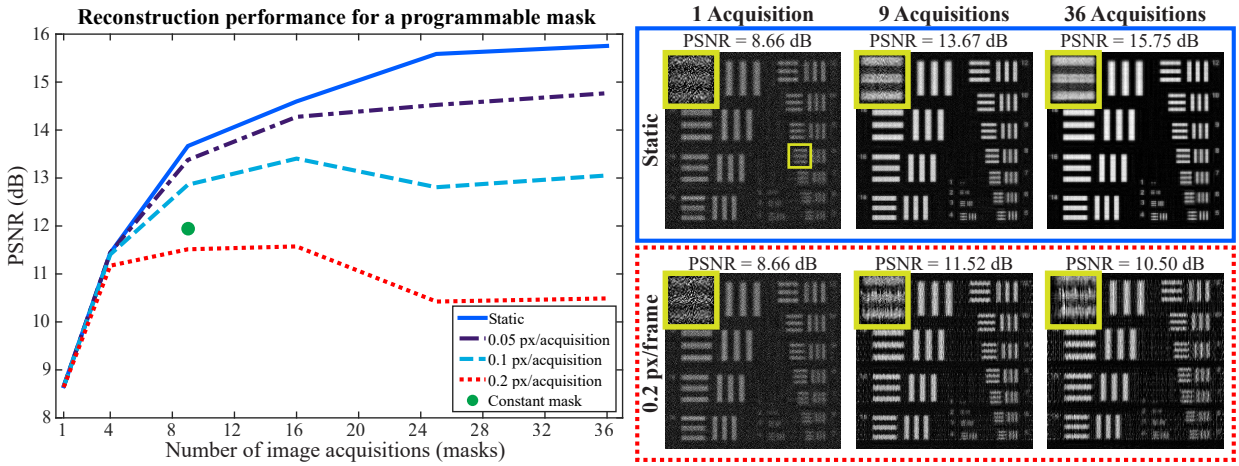

Fig. 7: *Simulation experiment of a FlatCam with a programmable mask. (Left) Reconstruction performance (in terms of peak-signal-to-noise ratio, PSNR) as we increase the number of image acquisitions (masks). A different mask pattern is used for each acqusition. The PSNR increases consistently for static and slow-moving scenes, but after peaking early deteriorates for faster moving scenes due to model mismatch. The green dot indicates the performance when all 9 image acquisitions of a static scene are done using the same (constant) mask. Though the performance is better than a single acquisition, it is still outshined by the programmable mask. (Right) Reconstructed images using 1, 9, and 36 acquisitions for the static scene and fast-moving scene.*

in order to estimate X by solving the least-squares problem

$$\hat{X}_{LS} = \arg\min_{X} ||\Phi_L X \Phi_R^T - Y||_F^2, \tag{7}$$

where $||\cdot||_F$ denotes the Frobenius norm. A regularization term can also be added as in the single-image-capture case.

Figure 7 (left) illustrates the results of a simulation of this approach. We generated a virtual high-resolution scene X, acquired a number of noisy acquisitions according to the model (5), and recovered the image estimate according to (7) with an additional Tikhonov regularizer. The horizontal axis corresponds to the number of acquisitions, each of which used a new mask. The vertical axis corresponds to peak-signal-to-noise ratio (PSNR), which measures the mean-squared error between the image estimate and X. Each acquisition had the same, fixed exposure time, and so we expect the quality of the image estimate to improve as we fuse more acquisitions. The blue curve demonstrates not only this improvement, but also an additional improvement due to changing the mask for each acquisition rather than reusing the same mask repeatedly. In particular, taking 9 acquisitions using 9 different masks attains a PNSR of 13.7 dB, while taking 9 acquisitions with the same mask attains a PSNR of only 11.8 dB (green dot on the figure). This 2 dB gain is testimony to the power of programmable masks.

The careful reader will note, however, that in practice the scene might change during acquisition, which would invalidate the model (5) and commensurately reduce performance. To investigate the effect of scene motion, we repeated the above experiment with three different dynamic scenes that diagonally translate the virtual high-resolution scene X by 0.05, 0.1, and 0.2 pixels per acquisition. For the slow-moving scene, the reconstruction performance improves with the number of acquisitions, just as with a static scene. But for the faster-moving scenes, after peaking early on, the reconstruction performance deteriorates with the number of acquisitions, due to the increased deviation from the model. Figure 7 (right) shows the reconstructed images using 1, 9, and 36 measurements for the static scene and the scene moving at 0.2 pixels per acquisition.

The trade-off between spatial and temporal resolution could be improved by estimating the motion between frames and registering the measurements before reconstructing the image (a difficult, but solvable problem; see [33]). Adaptive measurement schemes also hold promise for balancing this tradeoff.

3D imaging

FlatCam can computationally change its focus to new depths in a scene from a single acquisition. The key is that, for a given mask design, we can calibrate a set of separable measurement matrices $\{\Phi_{L_i}\}_{i=1,..,L}$ and $\{\Phi_{R_j}\}_{j=1,..,R}$, each obtained using a screen at a different depth (recall Figure 4).

Figure 8 (left) shows a heat map of reconstruction PSNR of a simulated 2D scene as a function of the scene distance and the calibration distance of the measurement matrices. We see that the reconstruction quality improves as the calibration depth of the camera approaches the actual scene depth. Moreover, the sensitivity of the reconstruction due to the discrepancy in these depths decreases with increasing scene distance.

Figure 8 (right), shows the reconstruction of a 3D scene at two different depth planes. For a particular fixed mask, we calibrated the measurement matrices with a screen at the distances of 7 cm and 27 cm. We accounted for the field of view of the sensor by adjusting the size of the calibration patterns in accordance with the CRA of the sensor. The resulting two sets of matrices were then used to create focused images at 7 cm and 27 cm from a single acquisition with FlatCam. (The line artifacts in the experimental reconstruction are due to scene illumination leaking into the sensor from the sides that was not accounted for in the calibration procedure. We can reduce the unaccounted light in future prototypes by introducing baffles.)

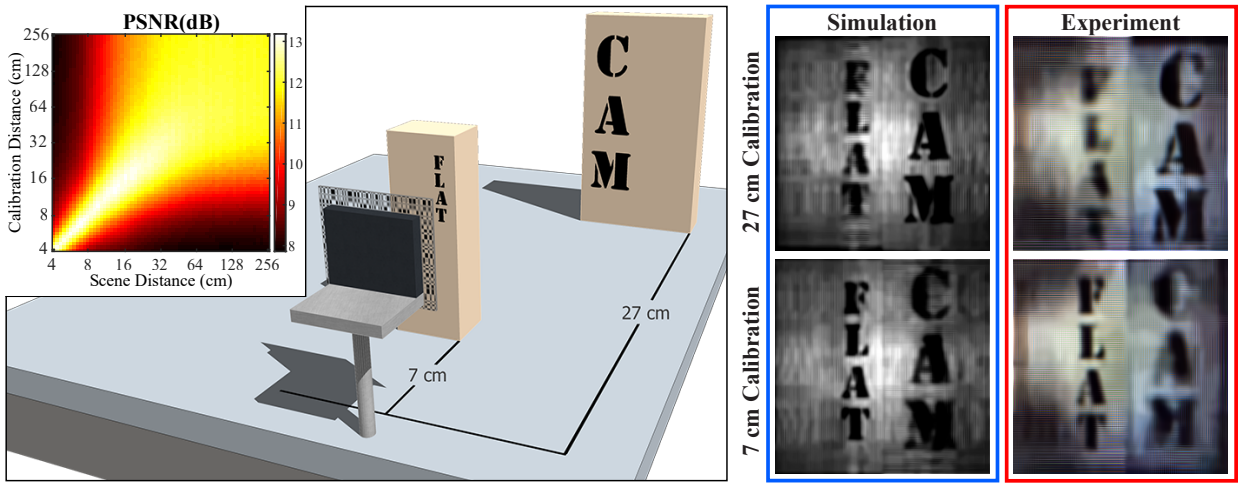

Fig. 8: *3D imaging with FlatCam. (Center) Experimental setup showing "FLAT" and "CAM" at different distances from the camera. (Left) Heat map of reconstruction PSNR of a simulation of the scene as a function of the scene distance and the calibration distance of the measurement matrices. At closer scene distances, the reconstruction is sensitive to the choice of multiplexing matrix at the correct calibration depth; at further scene distances, the sensitivity decreases. (Right) Reconstruction at 7 cm and 27 cm through simulation and our prototype FlatCam. The word "FLAT" placed at 7 cm is in focus when reconstructed using measurement matrices calibrated to depth of 7cm. The word "CAM" placed at 27 cm is in focus when reconstructed using measurement matrices calibrated to depth of 27cm.*

Limitations and Challenges Facing Lensless Imaging

The very first cameras were lensless (pinhole cameras), but the advent of lenses and other advanced optics relegated such systems to niche applications like X-ray and gamma ray imaging. The resurgence of lensless imaging can be attributed to the convergence of four factors: the development of digital CMOS and CCD sensor arrays, efficient and realistic image models and recovery algorithms, powerful computing, and new mask designs (such as the separable mask in the FlatCam).

The further development of lensless imaging, however, will face challenges. As the mask is moved closer to the sensor in any pinhole or coded aperture camera, the angular resolution decreases, resulting in a trade-off between minimal thickness and spatial resolution [29]. Additionally, computationally recovering a scene from less-than-perfectly-conditioned sensor measurements results in noise amplification. Although noise amplification cannot be eliminated, careful design of mask patterns and regularization models can minimize this effect. The necessity for a computational algorithm also results in a time-lag between image acquisition and reconstruction ($\sim$100 ms for FlatCam). Such a delay may be acceptable in certain applications but unacceptable in others such as augmented or virtual reality. There are a number

of avenues for continued research and development that could lead to significantly improved lensless imaging performance, including new architectures for improving spatial resolution, new image models to reduce the demultiplexing noise, and new computational algorithms to support high-speed sensing. Sometimes, size matters. The lensless imaging approach promises to challenge the traditional barriers of size, weight, cost, and performance in a broad range of applications spanning consumer, medical, scientific imaging, machine vision, and remote sensing. Indeed, the future of lensless imaging research and development looks very bright.

BOX 1: CODED APERTURE IN X-RAY AND GAMMA-RAY IMAGING

Coded aperture cameras were originally invented for X-ray astronomy [2], [34], and they have been primarily used for X-ray and gamma-ray imaging since then [3], [4], [32], [35]. For instance, SWIFT space telescope (see Figure 9) is a multi-wavelength space telescope currently in use for observing gamma ray bursts [36].

Image formation in a lens-based camera can be viewed as a one-to-one mapping of points at a focal plane in the scene onto a sensor. A lens is a refractive element that manipulates light wavefronts such that all the light coming from a certain direction in the scene converges to a particular location on the sensor. Visible light can be easily manipulated using transparent materials, such as glass and plastic, that have a large refractive index. Therefore, lenses for visible light are easily available at low cost.

High energy radiation beyond the visible spectrum, such as X-rays and gamma-rays, are routinely acquired in radiology, screening, and astronomy applications. Imaging these radiations enables us to look inside a human body for medical diagnosis, screen luggage at the airports, and observe black holes and supernovae in the cosmos. However, X-rays and gamma-rays are not as easy to manipulate with refractive optics as visible light. Therefore, the methods for imaging high-energy radiations primarily rely on reflection or diffraction optics.

The classical imaging architectures for X-rays and gamma-rays use a collimator in front of a sensor. A collimator typically consists of a thick sheet of lead or other material opaque to the incoming rays with multiple holes. Every sensor pixel behind a hole has a narrow field of view, since only a small cone of light in a particular direction can travel through each hole. Thus, a collimator localizes the directions of the rays reaching the sensor. Light from multiple locations and angles can be recorded by moving the collimator and the detector accordingly. The two primary drawbacks of collimator-based imaging are (i) light throughput is extremely low, since the collimator allows only a fraction of incoming light to reach the sensor, and (ii) the recorded image has a low angular resolution, because every sensor pixel records the average intensity of light over its entire field-of-view.

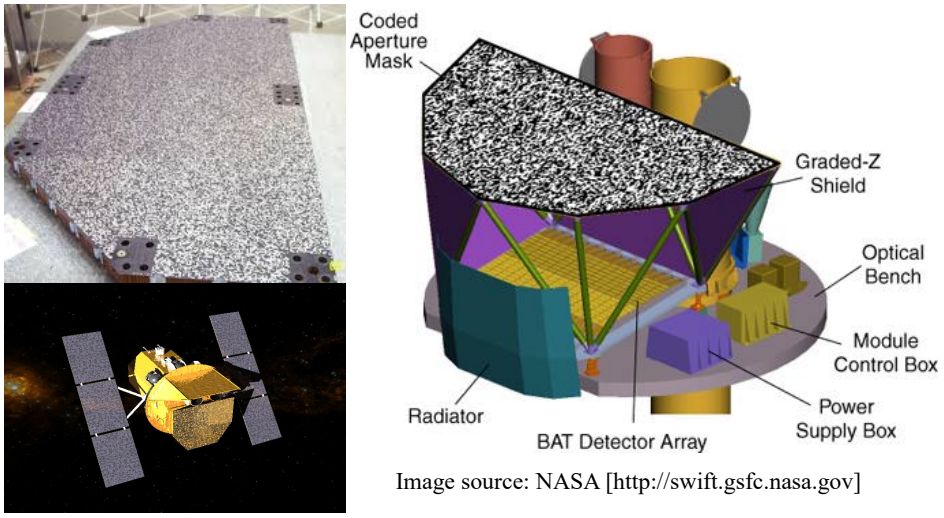

Fig. 9: *SWIFT is a multi-wavelength space observatory dedicated to the study of gamma-ray bursts. Its burst alert telescope (BAT) uses a coded mask to detect gamma ray burst events and compute their coordinates in the sky. The D-shaped coded aperture mask is made of nearly 54,000 lead tiles [36].*

A coded aperture-based imaging system offers better light efficiency and angular resolution as compared to either a pinhole- or collimator-based system. A coded aperture camera consists of a mask with transparent and opaque features placed in front of a sensor. Light from any particular location in the scene casts its unique shadow of the mask on the sensor plane. Therefore, each sensor pixel records a coded multiplexing of light from multiple scene locations. The relationship between the sensor measurements and the scene intensities can be described as a linear system, which can be solved using a computational algorithm.

ACKNOWLEDGMENTS

This work was supported in part by NSF grants CCF-1527501, CCF-1502875, DARPA REVEAL grant HR0011-16-C-0028, ONR DURIP grant N00014-15-1-2878, and ONR grant N00014-15-1-2735.

REFERENCES

[1] V. Dragoi, A. Filbert, S. Zhu, and G. Mittendorfer, "CMOS wafer bonding for back-side illuminated image sensors fabrication," in *2010 11th International Conference on Electronic Packaging Technology & High Density Packaging*, 2010, pp. 27–30.

[2] R. Dicke, "Scatter-hole cameras for X-rays and gamma rays," *The Astrophysical Journal*, vol. 153, p. L101, 1968.

[3] E. Fenimore and T. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Applied Optics*, vol. 17, no. 3, pp. 337–347, 1978.

[4] T. Cannon and E. Fenimore, "Coded aperture imaging: Many holes make light work," *Optical Engineering*, vol. 19, no. 3, pp. 193–283, 1980.

[5] A. Busboom, H. Elders-Boll, and H. Schotten, "Uniformly redundant arrays," *Experimental Astronomy*, vol. 8, no. 2, pp. 97–123, 1998.

[6] J. W. Goodman, *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.

[7] H. H. Barrett, "Fresnel zone plate imaging in nuclear medicine," *Journal of Nuclear Medicine*, vol. 13, no. 6, pp. 382–385, 1972.

[8] J. Kirz, "Phase zone plates for X-rays and the extreme UV," *J. Optical Soc. Am.*, vol. 64, no. 3, pp. 301–309, 1974.

[9] Y. Chu, J. Yi, F. De Carlo, Q. Shen, W.-K. Lee, H. Wu, C. Wang, J. Wang, C. Liu, C. Wang *et al.*, "Hard-X-ray microscopy with fresnel zone plates reaches 40 nm rayleigh resolution," *Applied Physics Letters*, vol. 92, no. 10, p. 103119, 2008.

[10] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2006, pp. 339–346.

[11] G. Huang, H. Jiang, K. Matthews, and P. Wilford, "Lensless imaging by compressive sensing," in *20th IEEE International Conference on Image Processing*, 2013, pp. 2101–2105.

[12] M. J. DeWeert and B. P. Farm, "Lensless coded-aperture imaging with separable doubly-toeplitz masks," *Optical Engineering*, vol. 54, no. 2, pp. 023 102–023 102, 2015.

[13] H. Jiang, G. Huang, and P. Wilford, "Multi-view in lensless compressive imaging," in *Picture Coding Symposium (PCS), 2013*, Dec 2013, pp. 41–44.

[14] A. Wang, P. Gill, and A. Molnar, "Angle sensitive pixels in CMOS for lensless 3D imaging," in *IEEE Custom Integrated Circuits Conference*, 2009, pp. 371–374.

[15] P. R. Gill, C. Lee, D.-G. Lee, A. Wang, and A. Molnar, "A microscale camera using direct fourier-domain scene capture," *Optics Letters*, vol. 36, no. 15, pp. 2949–2951, 2011.

[16] P. R. Gill and D. G. Stork, "Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings," in *Computational Optical Sensing and Imaging*. Optical Society of America, 2013, pp. CW4C–3.

[17] D. Stork and P. Gill, "Lensless ultra-miniature CMOS computational imagers and sensors," in *International Conference on Sensor Technologies and Applications*, 2013, pp. 186–190.

[18] H. T. E. F.R.S., "Lxxvi. facts relating to optical science. no. iv," *Philosophical Magazine Series 3*, vol. 9, no. 56, pp. 401–407, 1836. [Online]. Available: http://dx.doi.org/10.1080/14786443608649032

[19] X. Cui, L. M. Lee, X. Heng, W. Zhong, P. W. Sternberg, D. Psaltis, and C. Yang, "Lensless high-resolution on-chip optofluidic microscopes for caenorhabditis elegans and cell imaging," *Proceedings of the National Academy of Sciences*, 2008. [Online]. Available: http://www.pnas.org/content/early/2008/07/25/0804612105.abstract

[20] S. A. Lee, R. Leitao, G. Zheng, S. Yang, A. Rodriguez, and C. Yang, "Color capable sub-pixel resolving optofluidic microscope and its application to blood cell imaging for malaria diagnosis," *PLoS ONE*, vol. 6, no. 10, pp. 1–6, 10 2011. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0026127

[21] G. Zheng, S. A. Lee, Y. Antebi, M. B. Elowitz, and C. Yang, "The epetri dish, an on-chip cell imaging platform based on subpixel perspective sweeping microscopy (spsm)," *Proceedings of the National Academy of Sciences*, vol. 108, no. 41, pp. 16 889–16 894, 2011. [Online]. Available: http://www.pnas.org/content/108/41/16889.abstract

[22] J. Spence, U. Weierstall, and M. Howells, "Phase recovery and lensless imaging by iterative methods in optical, X-ray and electron diffraction," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 360, no. 1794, pp. 875–895, 2002.

[23] H. Faulkner and J. Rodenburg, "Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm," *Physical Review Letters*, vol. 93, no. 2, p. 023903, 2004.

[24] A. Greenbaum, W. Luo, T.-W. Su, Z. Göröcs, L. Xue, S. O. Isikman, A. F. Coskun, O. Mudanyali, and A. Ozcan, "Imaging without lenses: Achievements and remaining challenges of wide-field on-chip microscopy," *Nature Methods*, vol. 9, no. 9, pp. 889–895, 2012.

[25] A. Greenbaum, Y. Zhang, A. Feizi, P.-L. Chung, W. Luo, S. R. Kandukuri, and A. Ozcan, "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy," *Science Translational Medicine*, vol. 6, no. 267, pp. 267ra175–267ra175, 2014.

[26] J. Rodenburg, A. Hurst, A. Cullis, B. Dobson, F. Pfeiffer, O. Bunk, C. David, K. Jefimovs, and I. Johnson, "Hard-X-ray lensless imaging of extended objects," *Physical Review Letters*, vol. 98, no. 3, p. 034801, 2007.

[27] M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C. M. Kewish, R. Wepf, O. Bunk, and F. Pfeiffer, "Ptychographic X-ray computed tomography at the nanoscale," *Nature*, vol. 467, no. 7314, pp. 436–439, 2010.

[28] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[29] J. Miao, "Coherent diffraction imaging," *Microscopy and Microanalysis*, vol. 20, no. S3, pp. 368–369, 2014.

[30] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," in *SPIE International Symposium on Optical Science, Engineering, and Instrumentation*, 1999.

[31] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. Baraniuk, "Flatcam: Thin, bare-sensor cameras using coded aperture and computation," *arXiv preprint arXiv:1509.00116*, 2015.

[32] P. Durrant, M. Dallimore, I. Jupp, and D. Ramsden, "The application of pinhole and coded aperture imaging in the nuclear environment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 422, no. 1, pp. 667–671, 1999.

[33] A. C. Sankaranarayanan, L. Xu, C. Studer, Y. Li, K. F. Kelly, and R. G. Baraniuk, "Video compressive sensing for spatial multiplexing cameras using motion-flow models," *CoRR*, vol. abs/1503.02727, 2015. [Online]. Available: http://arxiv.org/abs/1503.02727

[34] E. Caroli, J. Stephen, G. Di Cocco, L. Natalucci, and A. Spizzichino, "Coded aperture imaging in X-and gamma-ray astronomy," *Space Science Reviews*, vol. 45, no. 3-4, pp. 349–403, 1987.

[35] D. J. Brady, *Optical Imaging and Spectroscopy*. John Wiley & Sons, 2009.

[36] C. Markwardt, S. Barthelmy, J. Cummings, D. Hullinger, H. Krimm, and A. Parsons, "The swift bat software guide," *NASA/GSFC, Greenbelt, MD*, vol. 6, 2007.