Coded Illumination for Improved Lensless Imaging

Yucheng Zheng, *Student Member, IEEE* and M. Salman Asif *Senior Member, IEEE*

***Abstract*—Mask-based lensless cameras can be flat, thin, and light-weight, which makes them suitable for novel designs of computational imaging systems with large surface areas and arbitrary shapes. Despite recent progress in lensless cameras, the quality of images recovered from the lensless cameras is often poor due to the ill-conditioning of the underlying measurement system. In this paper, we propose to use coded illumination to improve the quality of images reconstructed with lensless cameras. In our imaging model, the scene/object is illuminated by multiple coded illumination patterns as the lensless camera records sensor measurements. We designed and tested a number of illumination patterns and observed that shifting dots (and related orthogonal) patterns provide the best overall performance. We propose a fast and low-complexity recovery algorithm that exploits the separability and block-diagonal structure in our system. We present simulation results and hardware experiment results to demonstrate that our proposed method can significantly improve the reconstruction quality.**

***Index Terms*—Lensless imaging, coded illumination, image reconstruction.**

I. Introduction

LENSLESS cameras provide novel designs for extreme imaging conditions that require small, thin form factor, large field-of-view, or large-area sensors [1]–[4]. Compared to conventional lens-based cameras, lensless cameras can be flat, thin, light-weight, and potentially flexible because the physical constraints imposed by a lens are relaxed. FlatCam is an example of a lensless camera [1], which belongs to a broader class of coded-aperture cameras that replace the lens with a coded mask [5], [6]. The image formed on the sensor with a coded mask is a linear combination of multiple shifted versions of the scene. To recover the scene image from the sensor measurements, we need to solve a linear inverse problem. The quality of the recovered image depends on the conditioning of the linear system; especially in the absence of any prior knowledge about the scene.

In this paper, we propose a new method that combines coded illumination with mask-based lensless cameras (such as FlatCam) to improve the quality of recovered images, illustrated in Fig. 1. A 2D scene is illuminated with multiple coded patterns during data acquisition. We capture sensor measurements for each illumination pattern and use a fast recovery algorithm to reconstruct the scene using all the measurements. The main contributions of this paper are as follows leftmargin=4mm,topsep=0pt,itemsep=0ex,partopsep=1ex,parsep=1pt

- We propose a new framework to combine coded illumination with lensless imaging.
- We propose a fast and low-complexity algorithm that exploits separability of the mask and illumination patterns and avoids storing all the measurements or creating large system matrices during reconstruction.
- We design shifting dots and orthogonal patterns that provide best overall performance in terms of quality and computational complexity.
- We present simulation results to show that our method can improve the image reconstruction quality under different system conditions.
- We present real experiments using a prototype we built to evaluate the performance of our imaging system and algorithm in the real environments.

Our main objective is to demonstrate that we can significantly improve the conditioning of a lensless imaging systems and the quality of the reconstruction by using coded illumination. Our experiments show that the image quality improves in almost all cases as we increase the number of illumination patterns. Our method also has an inherent trade-off between imaging speed and the quality of the recovered images. In this paper, we mainly implement and discuss the design of FlatCam [1] with separable masks and illumination patterns. Nevertheless, the ideas presented here can be used to improve the imaging performance of other lensless systems, such as DiffuserCam [3], fresnel zone aperture [7], or phlatcam [8].

The proposed framework can be used to build large-area, high-resolution, and compact cameras that can potentially be used for under-the-display fingerprint and vein imaging. Such cameras can use the display panels for coded illumination [9]–[11]. Lensless microscopes [3], [12], [13] can also benefit by combining coded illumination with data capture. Another potential application for such lensless cameras is in different wearable devices. In particular, our inspiration came from a project in which we seek to embed lensless sensors on a soft robot that will be used to assist infants in rehabilitation. The goal is to track infant arm motion using lensless sensors on a wearable sleeve while they reach objects. In all these cases, the objects in the scene can potentially be illuminated with a built-in display, a mini projector, or an add-on light source combined with a shifting mask,.

II. Related Work

The mask-based lensless camera such as FlatCam [1] is an extended version of pinhole cameras. Although a pinhole camera is able to image the scene directly on its sensor, it often suffers from noise [14]. Coded aperture-based cameras alleviate this problem by using multiple pinholes placed in a designed pattern [1], [2], [5], [6], [15]. In contrast to conventional lens-based cameras that capture images of the

This work was supported in part by NSF Awards 2046293, 2133084 and AFOSR Award FA9550-21-1-0330.

Y. Zheng and M. Asif are with the Department of Electrical and Computer Engineering, University of California, Riverside, CA, 92521 USA (e-mail: yzhen069@ucr.edu; sasif@ucr.edu).

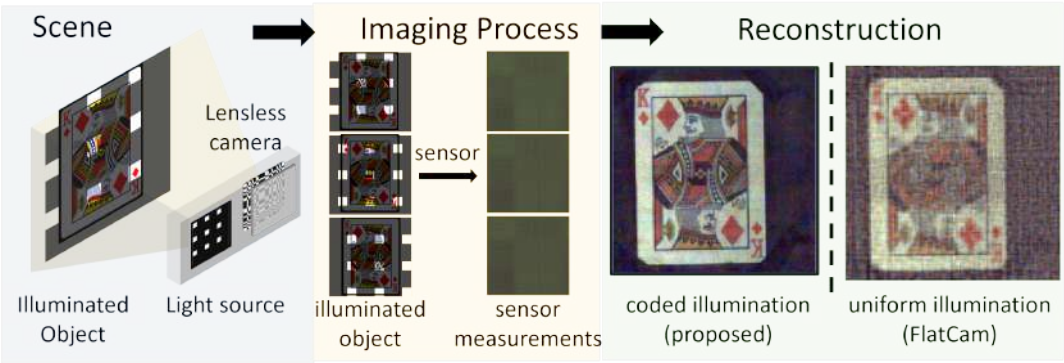


Fig. 1: An overview of our proposed method. We project a sequence of binary illumination patterns onto the object and capture the sensor measurements corresponding to each illumination pattern. The reconstruction result using multiple coded illumination patterns significantly outperforms the conventional method where the scene is illuminated by uniform illumination.

scene directly, coded mask-based cameras capture linear measurements of the scene and perform reconstruction by solving a linear inverse problem. Coded aperture-based cameras can also recover the depth information of a scene [3], [12], [16]–[20]. The main advantage of FlatCam is the thin and flat form factor, which also makes the system ill-conditioned and affects the quality of reconstructed images.

Signal recovery from ill-conditioned and under-determined systems is a classical and long-standing problem in signal processing. A number of methods have been proposed to tackle these problems over the decades [21]–[23]. An ill-conditioned system is unstable as its solution can change dramatically with tiny perturbations; thus, such a system rarely generates good results in a signal recovery problem. An under-determined system has fewer measurements than the number of unknowns; thus, it admits infinitely many solutions, and one can hardly determine the true solution using only the measurements. A standard approach to deal with ill-conditioned and under-determined systems is to add a signal-dependent regularization term in the recovery problem, which constrains the range of solutions. Popular methods include adding sparse and low-rank priors on the signals [21], [22], [24]–[26] and natural-image-like generators prior [27]–[29]. Another approach is to capture multiple, diverse measurements of the scene that makes the modified imaging system well-condition and the reconstruction more accurate [19], [20]. Recently, a number of methods have been proposed that use deep networks to reconstruct images from lensless measurements [8], [30]–[32]. Some of these methods provide an exceptional improvement over traditional optimization-based methods. Nevertheless, deep learning-based methods in general, and end-to-end methods in particular, provide a huge variation in performance for simulated and real data (mainly because of mismatch in the simulated/actual mask-sensor-projector configuration and scenes). In contrast to deep learning methods, our method seeks to improve the conditioning of the underlying linear system and offer better generalization and robust results for arbitrary scenes without the need for any learning from data.

Multiplexed illumination analysis is discussed in [33], which shows how overall reconstruction SNR can be improved using coded illumination with Hadamard codes. Plenoptic imaging and noise analysis for reconstruction from multiplexed measurements are discussed in [34]. Even though this paper is focused on lensless imaging with coded illumination, the noise analysis we discuss in Sec. VI-A follows similar arguments as [33], [34] to show the relationship between reconstruction error and overall system conditioning and singular values.

Our proposed approach is an active imaging approach combining coded modulation or structured illumination method with coded aperture imaging [35]–[37]. Structured illumination schemes are commonly used for imaging beyond diffraction in microscopy. These schemes use multiple structured illumination patterns to down-modulate high spatial frequencies in a sample into a low-frequency region that can be captured by the microscope [35], [38], [39]. Structured light is also widely used in multiplexing scene recovery to improve image quality and SNR [33], [40]–[42]. Other active imaging approach includes time-of-flight sensors [43], [44] that estimate the 3D scene by sending out infrared light and measuring its traveling time in reflection. Coded diffraction imaging is used to recover complex-valued wavefront from Fourier measurements [45], [46]. In coded diffraction imaging, the signal of interest gets modulated by a sequence of illumination patterns before the K-space measurements were captured [47]–[49]. Ptychography is another related method for capturing high-resolution microscopy images by capturing multiple images of the scene using a sequence of coded illumination patterns [50], [51]. Dual photography [52] and compressive sensing [53], [54] schemes also use coded illumination to sample the scenes. A dual photography system can create a high-resolution image of the scene by scanning the entire scene one pixel at a time, but it requires a fast laser projector. Single-pixel camera collects thousands of multiplexed measurements of the scene and solves a regularized optimization problem to reconstruct the image. In our method, we use lensless camera to capture tens of sensor frames with shifting coded illumination patterns and get better reconstruction with improved system conditioning.

A random illumination patterns-based lensless imaging method with simulations was presented in [55]. In contrast, we design shifting dots and orthogonal patterns that are significantly superior to random patterns in terms of quality

of reconstruction and computational complexity. We provide detailed simulations and experimental results on real data captured with a custom-built prototype.

III. Methods

A. Separable Imaging Model

FlatCam [1] consists of an amplitude mask placed on top of a bare sensor, and every sensor pixel records a linear combination of the entire scene. Suppose the sensor plane is at the origin of the 3D Cartesian coordinates (u, v, z) and an amplitude mask is placed parallel to the sensor at distance d. We can model the measurement recorded at sensor pixel (u, v) as

$$y(u,v) = \int x(u',v',z)\,\varphi(u',v',u,v,z)\,du'dv'dz, \tag{1}$$

where $x(u', v', z)$ represents the intensity and $\varphi(u', v', u, v, z)$ represents the sensor response of a point source at (u', v', z). In this paper, we assume that the scene consists of a single plane at a known depth; therefore, we can ignore the depth parameter and represent the sensor measurements as

$$\begin{aligned} y(u,v) &= \int x(u',v')\,\varphi(u',v',u,v)\,du'dv' \\ &\Rightarrow \quad \mathbf{y} = \mathbf{\Phi x}, \end{aligned} \tag{2}$$

where $\mathbf{x}$ denotes the scene intensity vector, $\mathbf{\Phi}$ denotes the system matrix, and $\mathbf{y}$ denotes the sensor measurement vector. The computational and memory complexity of the general imaging model in (2) makes it unsuitable for systems with a large number of scene and sensor pixels. We can overcome this challenge in a number of ways; for instance, we can use a separable model as in FlatCam [1], [12] or a convolutional model as in DiffuserCam [3]. We use a separable system in this paper.

A separable mask pattern that is aligned with the sensor grid yields a separable imaging system, which can be represented as

$$y(u,v) = \int\int x(u',v')\,\varphi_L(u',u)\,du'\;\varphi_R(v',v)\,dv'. \tag{3}$$

The product of $\varphi_L(u', v)$ and $\varphi_R(v', v)$ represents the separable system response for point sources along u, v axes. Let us assume that X represents an $n \times n$ image of the scene intensities at a fixed plane and Y denotes $m \times m$ sensor measurements, then we can represent the separable system in (3) as

$$Y = \Phi_L X \Phi_R^\top, \tag{4}$$

where Φ_L, Φ_R denote the system matrices for u, v axes, respectively. We assume square shapes for the scene and sensor to keep our discussion simple, but the ideas can be extended to arbitrary shapes.

B. Coded Illumination and Reconstruction

The effect of illumination can be modeled as an element-wise product between the scene and the illumination patterns. In our experiments, we use a laser projector placed next to the lensless camera to illuminate the object. In other applications, such as under-the-display cameras, an LED screen can be used for illumination. To simplify the recovery process, we further assume that the illumination patterns are separable and drawn from columns of $n \times k$ matrices P_L and P_R. Let us denote a pattern as $P_{i,j} = p_{Li} p_{Rj}^\top$, where p_{Li} and p_{Rj} are ith and jth columns of P_L and P_R, respectively. We can describe sensor measurements for any given illumination pattern $P_{i,j}$ as

$$Y_{i,j} = \Phi_L (P_{i,j} \odot X) \Phi_R^\top + E_{i,j}, \tag{5}$$

where $\odot$ represents element-wise multiplication operator and $E_{i,j}$ denotes measurement noise.

To recover image X from the sensor measurements $Y_{i,j}$, we can solve the following ℓ_2-regularized least-squares problem:

$$\underset{X}{\operatorname{argmin}} \sum_{i,j} \|Y_{i,j} - \Phi_L (P_{i,j} \odot X) \Phi_R^\top\|_2^2 + \lambda \|X\|_2^2, \tag{6}$$

where $\lambda > 0$ is a regularization parameter. An optimal solution of (6) must satisfy the following conditions (which can be derived by setting the gradient to zero) with $\mathbf{Q} = \sum_{i,j} (\Phi_L^\top Y_{i,j} \Phi_R) \odot P_{i,j}$:

$$\begin{aligned} \mathbf{Q} &= \underbrace{(\Phi_L^T \Phi_L \odot P_L P_L^T)}_{\mathbf{A}_L} X \underbrace{(\Phi_R^T \Phi_R \odot P_R P_R^T)}_{\mathbf{A}_R} + \lambda X, \\ \Rightarrow \mathbf{Q} &= \mathbf{A_L} X \mathbf{A_R} + \lambda X, \end{aligned} \tag{7}$$

where $\mathbf{A_L}$, and $\mathbf{A_R}$ are $n \times n$ matrices. The solution of (7) can be written in closed form using the eigen-decomposition of $\mathbf{A}_L, \mathbf{A}_R$ [1] as

$$\hat{X} = \mathbf{V_L}[(\mathbf{V_L}^\top \mathbf{Q} \mathbf{V_R}) ./ (\mathbf{s_L} \mathbf{s_R}^\top + \lambda \mathbf{1}\mathbf{1}^\top)] \mathbf{V_R}^\top, \tag{8}$$

where $\mathbf{V_L}, \mathbf{V_R}$ denote the eigenvectors and $\mathbf{s_L}, \mathbf{s_R}$ denote the eigenvalues of $\mathbf{A_L}, \mathbf{A_R}$, respectively, $./$ denotes element-wise division of entries in two matrices, and $\mathbf{1}$ denotes a vector with all ones.

A naïve approach would require storing all the measurements $Y_{i,j}$, which increases the storage complexity of the system proportional to the number of illumination patterns. The procedure described above avoids this cost, as we can recursively update an estimate of all the matrices and vectors needed for image recovery without any additional storage overhead. Thus, the storage cost of our method remains constant regardless of the number of illumination patterns. Every captured sensor measurement requires some processing, so the computational cost per recovered image increases linearly with the number of illumination patterns.

The required storage space for all the parameters is $\mathcal{O}(n^2)$ because $\mathbf{Q}$, $\mathbf{A_L}$, and $\mathbf{A_R}$ are $n \times n$ matrices. We only need to compute $\mathbf{A_L}, \mathbf{A_R}$ once, each of which costs $O(mn^2 + kn^2)$. Eigendecomposition of $n \times n$ matrices is $O(n^3)$. The most expensive step in our method is computing $\mathbf{Q}$, which we can perform by in-place addition of $(\Phi_L^\top Y_{i,j} \Phi_R) \odot P_{i,j}$ as we acquire measurements for all i, j. In this manner, we never need to store any of the captured measurements. The complexity of updating $\mathbf{Q}$ is $\mathcal{O}(k^2(nm^2 + mn^2))$.

C. Choice of Illumination Patterns

One of our goals is to select the $n \times k$ illumination pattern matrices P_L, P_R that maximize the quality of reconstruction for fixed Φ_L, Φ_R. The quality of reconstruction in (8) directly depends on the conditioning of the $\mathbf{A_L}, \mathbf{A_R}$ matrices in (7), which in turn depends on the mask and illumination patterns. One possible approach to improve the conditioning of $\mathbf{A_L}, \mathbf{A_R}$ is to make them diagonal or diagonally dominant [23], which we can achieve by enforcing the same structures in $P_L P_L^\top, P_R P_R^\top$.

In principle, we can make $P_L P_L^\top$ diagonal or even identity by using P_L as an identity matrix, which requires $k = n$. This would be equivalent to scanning the entire scene by illuminating one pixel at a time. We can also make $P_L P_L^\top$ identity by selecting P_L as any orthogonal matrix, which also requires $k = n$. In a practical scenario, we can only use a small number of illumination patterns; therefore, $k \ll n$. Below we discuss how we can get diagonally dominant $\mathbf{A_L}, \mathbf{A_R}$ using small values of k.

We propose to use illumination patterns that constitute an orthogonal basis over $k \times k$ blocks and repeat the same patterns across the entire scene. The simplest example of such patterns is a dot pattern in which two adjacent dots are placed k scene pixels apart. We can then shift the dot pattern across horizontal and vertical directions, one pixel at a time, to capture k^2 shifting dots patterns. These shifting dots patterns are separable and orthogonal over every $k \times k$ block. More generally, we can use any sequence of orthogonal separable patterns over $k \times k$ blocks. Let us assume the separable illumination patterns can be drawn from P_L, P_R that are defined as

$$P_L = P_R = \begin{bmatrix} \psi_k \\ \vdots \\ \psi_k, \end{bmatrix} \Rightarrow P_L P_L^\top = P_R P_R^\top = \begin{bmatrix} I_k & \dots & I_k \\ \vdots & \ddots & \vdots \\ I_k & \dots & I_k \end{bmatrix}, \quad (9)$$

where ψ_k and I_k denote $k \times k$ orthogonal and identity matrices, respectively. The resulting $P_L P_L^\top, P_R P_R^\top$ matrices (shown above) will be block matrices with $k \times k$ identity blocks, and the $\mathbf{A_L}, \mathbf{A_R}$ matrices (shown in Fig. 2) will be block matrices with $k \times k$ diagonal blocks.

Recall that $\mathbf{A_L}, \mathbf{A_R}$ are system matrices for the linear system we need to solve in (7). We can permute the rows and columns of $\mathbf{Q}$, which is equivalent to permuting rows of $\mathbf{A_L}$ and columns of $\mathbf{A_R}$, without affecting the solution of (7). Let us represent the resulting permuted equations as

$$\widetilde{\mathbf{Q}} = \widetilde{\mathbf{A}}_\mathbf{L} X \widetilde{\mathbf{A}}_\mathbf{R} + \lambda X. \quad (10)$$

We illustrate the permuted system in Fig. 2, where $\widetilde{\mathbf{A}}_\mathbf{L}, \widetilde{\mathbf{A}}_\mathbf{R}$ represent $n \times n$ block diagonal matrices, with k blocks along the diagonal each of size $\frac{n}{k} \times \frac{n}{k}$. As we increase the value of k, the system matrices $\widetilde{\mathbf{A}}_\mathbf{L}, \widetilde{\mathbf{A}}_\mathbf{R}$ become diagonally dominant and the overall conditioning of the system improves.

We can exploit the block diagonal structure of the system matrices to solve the system in (10) in a reliable and computationally efficient manner. Note that the separable, block diagonal system can be divided into k^2 independent systems, each involving an $\frac{n}{k} \times \frac{n}{k}$ patch in $\mathbf{X}$. We can solve all these systems in parallel to speed up recovery. The overall complexity of the inversion also reduces from $O(n^3)$ to $O(n^3/k)$. Furthermore, the conditioning of the overall system now depends on the conditioning of each $\frac{n}{k} \times \frac{n}{k}$ block in $\widetilde{\mathbf{A}}_\mathbf{L}, \widetilde{\mathbf{A}}_\mathbf{R}$. As long as all the blocks are well-conditioned, we can recover the underlying signal accurately.

IV. Simulations

A. Simulation Setup

To validate the performance of the proposed algorithm, we simulate a lensless imaging system where a coded-mask is placed on top of an image sensor. We use a separable maximum length sequence (MLS) mask pattern. The size of each mask feature is 60μm, and the sensor-mask distance is 2mm. The sensor pitch in the simulation is 11.72μm and the total number of pixels on the sensors is 512×512. We simulate a 128×128 planar scene that is 40cm away from the sensor, and the height/width of the scene is 12cm. The simulated sensor noise includes photon noise and read noise [19], and the noisy sensor measurements can be described as

$$\mathbf{Y}_n = \frac{G}{F}(\text{Poisson}(\frac{F}{G}\mathbf{Y}) + N(0, \sigma^2)), \quad (11)$$

where $\mathbf{Y}$ and $\mathbf{Y}_n$ refers to original and noisy measurements, F stands for the full-well capacity, and G represents the gain value. The variance $\sigma = F \times 10^{-R/20}$ and R is the dynamic range. We show the reconstruction results on a few example scenes using different illumination patterns; additional results can be found in the supplementary material.

B. Effect of Illumination on Reconstruction

We first evaluate the conditioning of different illumination patterns by observing the singular values of the system matrices in (7). The matrices that have flat singular value spectrum provide better recovery performance [1], [3], [23]. We tested different types of binary, separable illumination patterns for this experiment. We generate different instances of matrices P_L, P_R and use outer products of all pairs of columns to generate the illumination patterns. We ensure that the union of all the patterns should illuminate all the pixels (i.e., if we add columns of P_L, P_R, they should be nonzero everywhere). **Uniform:** One pattern that illuminates all the pixels simultaneously; P_L, P_R are vectors of all ones. **Random:** P_L and P_R are $k \times n$ binary random matrices that generate k^2 patterns. **Orthogonal:** We tested two types of orthogonal patterns (shifting dots and repeated Hadamard) that yield identical system matrices in (10). **Shifting dots:** P_L, P_R are $k \times n$ matrices, each of which consists of $k \times k$ identity matrices stacked on top of each other (as described in (9)). The base illumination pattern consists of dots separated by k pixels along the horizontal and vertical directions. We generate a total of k^2 illumination patterns, each of which is a shifted version of the base pattern. The summation of all the patterns will give us a uniform illumination pattern. **Repeated Hadamard:** As an extension to shifting dots, P_L, P_R are $k \times n$ matrices, each of which consists of the same $k \times k$ orthogonal Hadamard matrix stacked on top of each other. We can use grayscale or

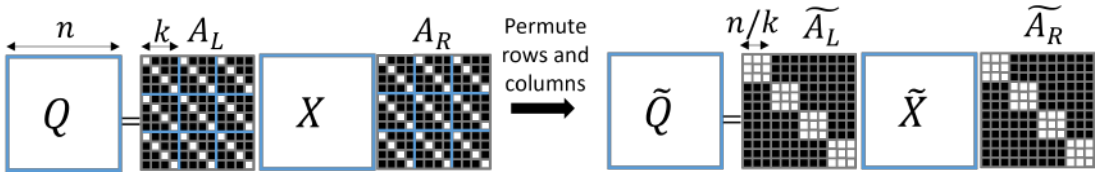
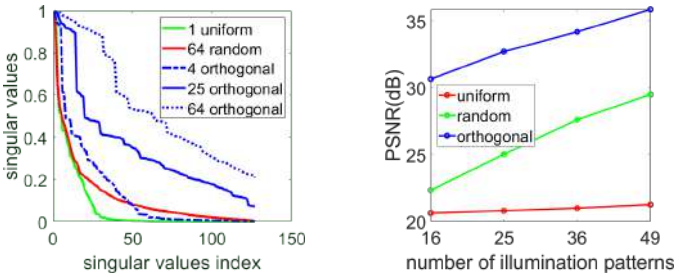
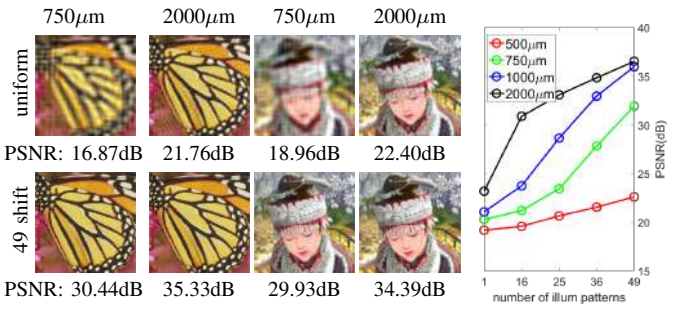


Fig. 2: Illustration of system in (7) when the illumination patterns form orthogonal basis over $k \times k$ image patch. (Left) A_L and A_R are $n \times n$ block matrices with diagonal blocks of size $k \times k$. (Right) Permuting rows and columns results in block diagonal matrices with block size $\frac{n}{k} \times \frac{n}{k}$. The recovery performance of this system depends on the conditioning of each block. We can solve the block diagonal system by recovering $\frac{n}{k} \times \frac{n}{k}$ patches in X independently, in parallel.



(a) System singular values. **(b)** Average PSNR values.

Fig. 3: Recovery performance of the imaging system. (a) Singular values of the system matrices with uniform, 64 random, and 4, 25, and 64 orthogonal shifting dots patterns. (b) Average PSNR of 8 test images reconstructed with different numbers of illumination patterns. Red dashed line shows results with uniform illumination. Reconstruction quality improves as we increase the number of illumination patterns. The orthogonal patterns outperform random and uniform patterns.



(a) Examples of reconstructed images with sensor-to-mask distance at 750μm and 2000μm. **(b)** Average PSNR of all test images.

Fig. 4: Simulation results for reconstruction with different sensor-to-mask distances in uniform illumination and shifting dots illumination patterns. The reconstruction quality improves as the sensor-to-mask distance increases.

color patterns to illuminate the scene. In real experiments, the calibration of the projector and nonlinearity of color/intensity ranges pose additional challenges.

To evaluate the effect of illumination on the lensless imaging system, we observe the decay of singular values of the system matrices in (7) as we increase the number of illumination patterns. Figure 3a plots the singular values for different illumination patterns. The singular values of the original system matrix with one uniform illumination decay sharply. We tested various patterns and found that the shifting dots or orthogonal patterns provide the best overall conditioning (flat SVD curve) for a given budget of measurements. As we increase the number of illumination patterns, the singular values spectrum becomes flatter, which corresponds to a system that is nearly orthogonal. In principle, we can use a dot projector to create shifting dots patterns, but we need a programmable projector for Hadamard-like patterns. In our experiments, we use a laser projector for illumination.

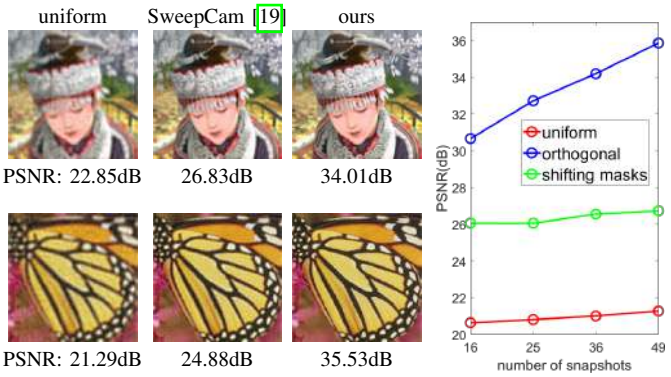
We present simulation results for reconstruction of 8 test images using different types and number of illumination patterns in Fig. 3b. We observe that orthogonal patterns outperform other patterns in terms of PSNR. Additional simulation results are available in the supplementary material.

C. Effect of Sensor-to-Mask Distance

The sensor-to-mask distance of a lensless imaging system greatly influences the conditioning of the system and the quality of reconstruction. We present simulation results to evaluate the performance with different sensor-to-mask distances. We keep the sensor size fixed at 512×512 pixels, and test the sensor-to-mask distance at 500μm, 750μm, 1000μm, and 2000μm. The reconstruction results for two test images are presented in Fig. 4, along with the PSNR plot for different numbers of shifting dots patterns. We observe that multiple coded illumination patterns outperform the results with a single uniform pattern at all the sensor-to-mask distances. Even if the sensor-to-mask distance is 750μm, the results with 25 coded illumination patterns are comparable to the uniform illumination results at 2000μm. Additional simulation results are available in the supplementary material.

D. Comparison with Multishot Lensless Methods

Multishot lensless methods are also applied in [19], [20], where multiple frames of lensless measurements are captured with a shifting or programmable mask on top of the image sensor. We present the simulation results comparing with a shifting mask-based multishot lensless imaging system [19] in Fig. 5. We keep the sensor size fixed at 512×512 and sensor-to-mask distance at 2000μm. The mask feature size is 60μm in all cases. We simulate the imaging process for SweepCam [19] with the mask at multiple shifting positions between -15 and $+15$ mask feature pixels (i.e., -900μm to $+900\mu$m physical shift). The number of captured frames for SweepCam [19] are the same as our method with coded illumination patterns. The simulation results in Fig. 5 show that our method with the same number of coded illumination patterns provides significantly better results compared to SweepCam.



(a) Example results with uniform illumination, 49 shifting masks, and 49 shifting dots.

(b) Average PSNR of all test images.

Fig. 5: Simulation results for reconstruction using uniform illumination, shifting masks from SweepCam [19] and shifting dots illumination patterns using our method at different number of measurements instances. Our method outperform uniform the other methods.

V. Experiments

A. Experiment Setup

We build a prototype with a lensless camera and a Sony MP-CL1 laser projector, shown in Fig. 6. The lensless camera prototype consists of an image sensor with a coded mask on top of it. The mask has a separable MLS pattern described in [1]. The mask has 511×511 square features, each of length/width 60μm. The sensor-mask distance is 2mm. We use a Sony IMX249 sensor that has 1920×1200 pixels with 5.86μm pixel pitch. We bin 2×2 sensor pixels and record 512×512 measurements from the center of the sensor. The effective sensor pitch is 11.72μm and the effective sensor area is nearly 6×6 mm. The target objects are 40cm away from the camera, and the projector illuminates 12×12 cm area on the scene plane. Finally, we reconstruct 128×128 pixels in the illuminated area, which results in the effective sampling interval of 120mm/128 = 0.93mm per pixel in the reconstructed images.

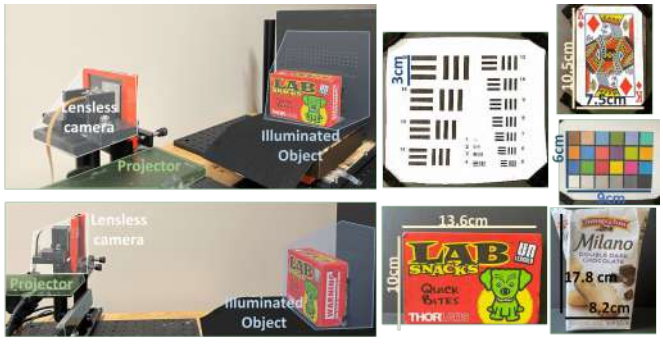


Fig. 6: The experiment setup and five test scenes (annotated and scaled to proportional size). The projector is placed right next to the lensless camera. The target scenes/objects are 40cm away from the camera.

In our experiment, the pixel grid of the scene, illumination patterns P_L, P_R, the system matrices Φ_L, Φ_R must be correctly aligned; otherwise, we will get artifacts in the reconstruction. To avoid any grid mismatch, we use the same projector to calibrate the system matrices and generate the illumination patterns in our experiments. We estimate the system matrices Φ_L, Φ_R following the Hadamard pattern-based calibration procedure in [1]. Let us denote the Hadamard patterns as an $n \times n$ orthogonal matrix $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$, where each $\mathbf{h}_i$ represents a Hadamard pattern of length n. To calibrate the system for an $n \times n$ pixel grid, we project n horizontal and n vertical (rank-one) Hadamard patterns on a flat surface in the scene and record their response on the sensor. Every horizontal pattern can be represented as an $n \times n$ rank-one matrix $X_i = \mathbf{h}_i \mathbf{1}^\top$, where $\mathbf{1}$ denotes a vector of all ones. The corresponding sensor response can be represented as a rank-one matrix $Y_i = \Phi_L(\mathbf{h}_i \mathbf{1}^\top)\Phi_R^\top \equiv \mathbf{u}_k \mathbf{v}^\top$, where $\mathbf{u}_i = \Phi_L \mathbf{h}_i$ and $\mathbf{v} = \Phi_R \mathbf{1}$. We can concatenate all the $\mathbf{u}_i$ as columns in a matrix as $\mathbf{U} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_n] = \Phi_L \mathbf{H}$ and estimate $\Phi_L = \mathbf{U}\mathbf{H}^\top$. We can repeat the same procedure with vertical Hadamard patterns $\mathbf{1}\mathbf{h}_i^\top$ and estimate Φ_R. Finally, we conduct the experiments with coded illumination while the position and angle of the projector are fixed. This ensures that the pixel grids of the projector and the transfer matrices are identical.

Fig. 7: Experimental results on five test scenes with different numbers of shifting dots patterns. The reconstruction results using multiple coded illumination patterns outperform the standard lensless camera with uniform illumination. The quality of reconstruction gradually increases as the number of illumination patterns increases.

B. Effect of Illumination Patterns

We present experimental results of our method on five different scenes in Fig. 7 and Fig. 8. The first three rows are printed card, color board, and a resolution target on sheets

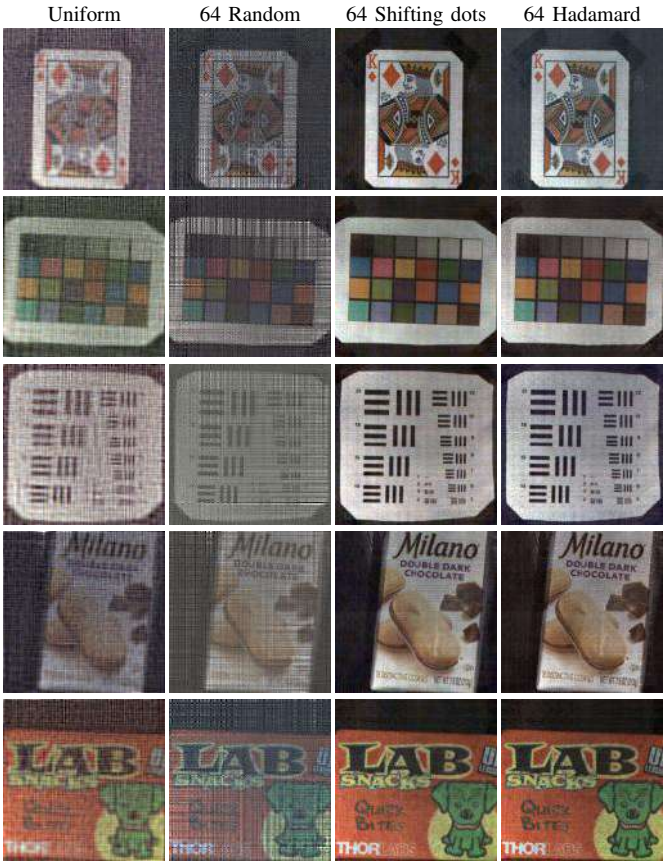
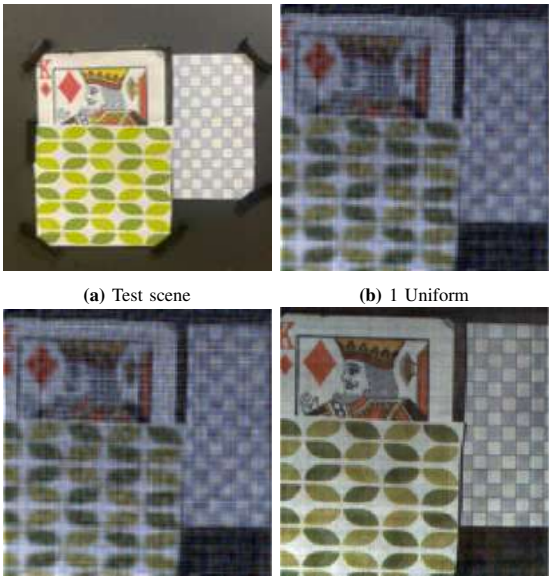


Fig. 8: Experimental results on five test scenes with different types of illumination patterns. The orthogonal patterns (shifting dots, hadamard) outperform random patterns and uniform illumination.



(a) Test scene **(b)** 1 Uniform **(c)** 49 Uniform **(d)** 49 Shifting dots

Fig. 9: Sample results for imaging performance with one uniform, 49 uniform, and 49 shifting dots illumination patterns. Capturing 49 shots requires the same data acquisition time, but the results for 49 shifting dots patterns are significantly better than 49 uniform patterns.

of paper. Since these scenes are planar, they fit the 2D scene assumption of our model perfectly. The last two rows are real objects with depth variations and may cause depth mismatch in the reconstruction.

From the results in Fig. 7, we observe that a single uniform illumination provides poor reconstruction. As we increase the number of illumination patterns, the resolution of the reconstruction images improves significantly. For example, in the third row, the horizontal and vertical features that cannot be resolved with the uniform illumination are easily resolved with the 49 patterns. Also, we observe in the fourth row that, the letters on the cookie bag that are completely unrecognizable with uniform illumination become very clear with 49 illumination patterns. The last two rows also suggest that our method is robust to small variations in depth. From the results in Fig. 8, the repeated orthogonal patterns(shifting dots and Hadamard) outperform other patterns, as expected from the singular values in Fig. 3a.

Note that the neighboring pixels have a similar response on the sensor in the lensless imaging system; therefore, neighboring pixels are harder to resolve compared to two pixels that are far from each other. The shifting dots pattern is a dot array where the illuminated pixels are maximally separated; therefore, two neighboring pixels in the scene are not illuminated at the same time. Also, as we increase the number of shifting dots patterns, the distance between adjacent illuminated pixels also increases, and that provides better reconstruction. Hadamard patterns provide similar separation because of their orthogonal structure.

Note that if we increase the number of measurements with coded illumination, the system conditioning and the quality of reconstructed images improve. On the other hand, if we increase the number of measurements for the uniform illumination pattern, the system conditioning and the quality of reconstructed images remains unchanged. To show this effect, we provide experimental results comparing the quality of reconstruction with multiple measurements captured under uniform and coded illumination. We present the results using 49 uniform and 49 shifting dots illumination patterns in Fig. 9. Note that the 49 uniform and 49 shifting dots illuminations use the same amount of exposure/capture time. We observe that shifting dots illumination provides significantly better reconstruction compared to uniform illumination. We observe that capturing 49 shots with uniform illumination reduces noise slightly compared to one uniform illumination, but 49 shifting dots provide significantly superior results. Note that multiple shots with uniform illumination can provide slightly better SNR because noise variance reduces due to averaging. The improvement with shifting dots patterns is the result of the better conditioning of the system that multiple uniform illumination shots cannot provide.

In summary, the ill-conditioned system matrices with uniform illumination pattern cause various artifacts. Capturing measurements from coded illumination improves the conditioning of the overall system and the resulting reconstructed images contain much fewer artifacts and noise. More illumination patterns demand more acquisition time, which enforces a trade-off between the quality of reconstruction and data acquisition time.

C. Effective Resolution and MTF

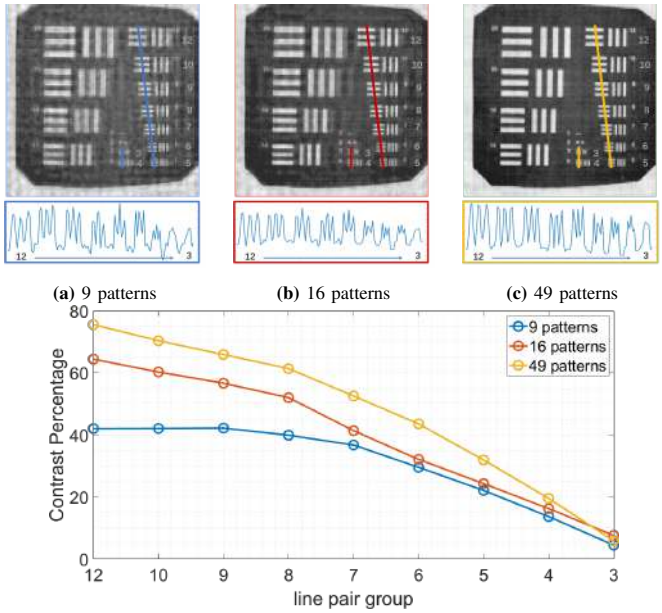
We use the resolution target image in Fig. 10 to analyze the spatial resolution of our method empirically. The distance between two printed white stripes in group 20 (upper-left) is 7.5mm (0.13 lp/mm) and the distance between two white stripes in group 5 is 1.9mm (0.52 lp/mm). The resolution of our imaging system is directly determined by the sampling interval of the illumination patterns. We place the target scene 40cm away from the camera and projector. The resolution of the MP-CL1 projector is 1280×720 and the overlap between the camera FOV and the maximum illuminating area at 40cm throw distance is 22×13cm. The projector pixel pitch is 0.17mm, which determines the achievable resolution of our method with the MP-CL1 projector. In our experiments, the width of every reconstructed pixel is 0.93mm, and the angular sampling of our system is $0.27°$. We can select smaller sampling intervals pitch by dividing the illuminating area into more pixels, but the minimal size cannot be smaller than projector pixel.

We present the modulation transfer function (MTF) of the reconstruction from different numbers of illumination patterns in Fig. 10. MTF measures how well we can discern the intensity of bright and dark pixels in line pairs under different conditions [56]. To plot the MTF, we manually select every group of horizontal and vertical line pairs after subtracting a fixed DC background, then average every group along the columns and rows, respectively, and finally compute the contrast percentage as $\frac{I_{\max}-I_{\min}}{I_{\max}+I_{\min}} \times 100$. We observe from the MTF plots that the contrast ratio for all the resolution groups improves as we increase the number of illumination patterns. We also plot the intensity values of the horizontal line stripes from multiple groups in Fig. 10, where the peaks and valleys correspond to the white and black stripes of the horizontal line pairs. We observe that the line pairs of group 12 can be distinguished clearly while the line pairs of group 3 cannot be distinguished.

D. Compressive Sensor Measurements

One potential application of coded illumination with lensless imaging is to expand the space of sensor measurements without increasing the number of sensor pixels. This can be especially useful when sensors capture fewer measurements than the number of pixels we seek to recover. Compressive sensing often deals with such scenarios where the number of sensor pixels is smaller than the number of unknown scene pixels [21], [24]. The system becomes under-determined and the reconstruction quality becomes worse. In lensless imaging systems, the number of sensor measurements is often larger than the number of reconstruction pixels, but that comes at the expense of low resolution of reconstructed images or larger and more costly sensors.

To validate the robustness of our method in the case of an under-determined system, we performed an experiment by binning the sensor measurements at increasing factors. The results are presented in Fig. 11. In our experiments, we capture 512×512 measurements and bin them to 256×256, 128×128, or 64×64 pixels by averaging the neighboring pixels in post-processing. The reconstruction image has 128×128 pixels, which implies the 64×64 sensor measurements yield an under-determined system with one illumination. We present the result of two scenes with different binning factors. We observe that as we bin a larger number of pixels, the system with a single uniform illumination becomes ill-conditioned and the quality of reconstruction degrades. On the other hand, the system with 49 shifting dots patterns provides stable reconstruction even when the measurements are binned to 64×64 pixels.



(d) Modulation transfer function plot. The vertical axis shows the contrast ratio in percentage.

Fig. 10: Resolution analysis of coded illumination. Top images in (a,b,c) show resolution target reconstructed using 9, 16, and 49 shifting dots patterns. Bottom plots in (a,b,c) show the intensity of a line from group 12 to group 3. The MTF (modulation transfer function) plot for different numbers of shifting dots illumination patterns is shown in (d). To compute MTF, we manually select each group of horizontal and vertical line pairs after subtracting a fixed DC background, then average every group along the columns and rows, respectively, and finally compute the contrast ratio.

E. Deep Network-based Denoising vs Reconstruction

Deep learning-based methods have been widely used for image recovery and enhancement tasks [30], [31], [57]. For instance, UNet [58] is used for image denoising and removing artifacts from reconstructed images [20], [30]. Deep learning is also used as the priors to improve the reconstruction results [28], especially when the number of measurements is small. UNet-based networks have been used to recover photorealistic images from lensless measurements in [30]. In our experiments, we observed that deep learning-based methods that learn to recover images directly from sensor measurements perform very well on simulated data but provide catastrophic results on real data. In contrast, the simple least-squares method we discussed in (8) provides stable results in all cases because coded illumination provides a well-conditioned system. We present simulation and experimental results below.

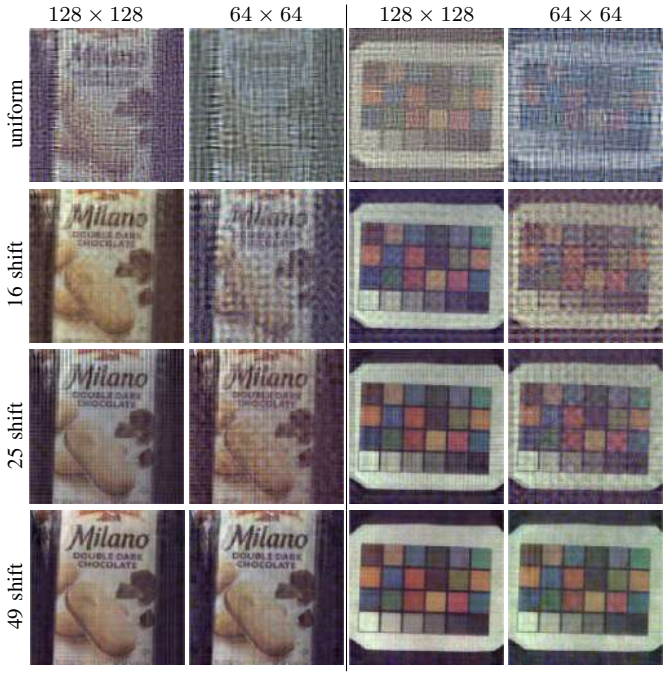


Fig. 11: Experimental results for reconstructing the 128×128 images from 64×64 and 128×128 measurements. Single (uniform) illumination-based method fails to recover images as the number of measured pixels reduce. Our method with 49 shifting dots patterns recovers near-perfect reconstruction at different levels of binning (compression) factors.

We performed some experiments to compare the performance of four methods with coded and uniform illumination: least squares (LS) in (8), LS with a trained UNet that is used as a refinement network, LS with pretrained refinement network in FlatNet [30], and end-to-end trained FlatNet. We present results for simulated sensor measurements in Fig. 12 and on real data captured using our prototype in Fig. 13.

The description of the four methods is as follows. We used 1000 natural RGB images from [59] to train the deep networks. leftmargin=4mm,topsep=0pt,itemsep=0ex,partopsep=1ex,parsep=1pt

- **LS.** Images are reconstructed using the least-square (LS) method with ℓ_2-regularization (8).
- **LS + trained UNet.** We trained a UNet using 1000 training images that learns to map the LS reconstruction into a refined image.
- **Trained FlatNet.** FlatNet [30] contains two steps: inversion step and refinement step that are trained jointly in an end-to-end manner. The inversion step maps the sensor measurements into a coarse image estimate, while the refinement step maps the coarse estimate to a cleaner image. We train both steps of the FlatNet in an end-to-end manner with simulated measurements for uniform and coded illuminations. We created the synthetic data by simulating the noisy lensless measurements for training images using the calibrated transfer matrices Φ_L, Φ_R described in Sec.V-A. Note that we need to train a separate network for different types/number of illumination patterns.
- **LS + pretrained FlatNet.** The refinement network in FlatNet has same architecture as our trained UNet. For the sake of comparison, we also used the pretrained refinement network from FlatNet to refine the LS solution.

Figure 12 shows images reconstructed from simulated measurements of selected test images. We observe that LS, LS+UNet, and trained FlatNet provide good reconstruction for simulated data using uniform illumination. In particular, images from trained FlatNet have best visual appearance simulations Fig. 12(d). For coded illumination, the quality of reconstruction improves for all the methods. The pretrained refinement network from FlatNet provides overly smoothed images and distorts spatial details.

Figure 13 shows images reconstructed from measurements captured by real prototype. Trained FlatNet fails to recover images from real data captured with uniform illumination in Fig. 13. In contrast, LS and LS+UNet provide a low-resolution reconstruction with uniform illumination. For coded illumination, trained FlatNet reconstructions have high quality that is similar to the results provided by LS and LS + trained UNet. Figure 13 further shows that UNet can partially reduce the artifacts and noise in the images reconstructed from the uniform illumination, but fails to improve the spatial resolution. In contrast, coded illuminations significantly improves the resolution and quality of the reconstructed images. Additional results can be found in the supplementary material.

Our main takeaway from these experiments is that deep networks cannot always recover missing details if the system is ill conditioned. In particular, deep network-based refinement/denoising can improve the appearance of images but cannot recover missing details. End-to-end trained networks, which recover images directly from multiplexed sensor measurements, perform well when train and test settings are identical, but provide catastrophic results in case of any mismatch. Coded illuminations improve the conditioning of the system, which benefits all the recovery methods.

VI. Discussion

We propose a framework for combining coded illumination with lensless imaging. We present extensive simulation and real experiment results to demonstrate that we can get significantly improved reconstruction with multiple coded illumination patterns compared to original uniform illumination.

Advantages: leftmargin=4mm,topsep=0pt,itemsep=0ex,partopsep=1ex,p

- Qualitative experiment results show that our method has robust performance even if the original system is severely ill-conditioned (small sensor-mask distance and small number of measurements).
- In space-limited applications such as under-the-display sensing, where the sensor-to-mask distance has to be small, our proposed method can offer significantly better reconstruction compared to uniform illumination.
- Our proposed reconstruction algorithm is efficient in both space and time. For shifting dots (and any other orthogonal) pattern over $k \times k$ blocks, the system can be transformed into k^2 independent small problems that can be solved in parallel for faster and better reconstruction.
- Shifting dots patterns have a simple structure but provide the best results in our experiments. Shifting dots patterns

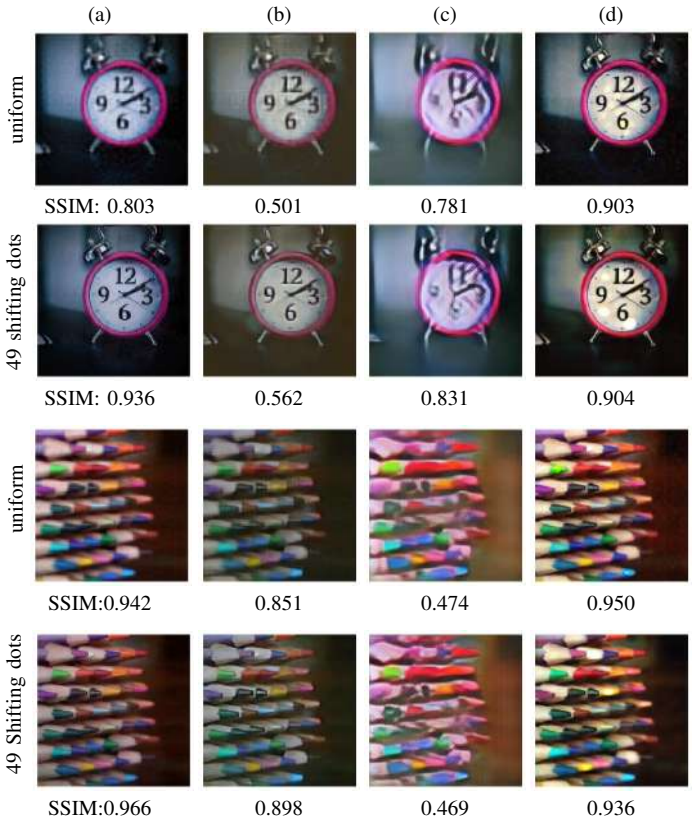


Fig. 12: Reconstruction results for simulated measurements with 49 uniform and shifting dots illumination patterns. Images in four columns show (a) LS solution, (b) LS solution with trained UNet refinement, (c) LS solution with pretrained FlatNet refinement, and (d) trained FlatNet that reconstructs image directly from measurements. For each image, we show the SSIM value underneath.

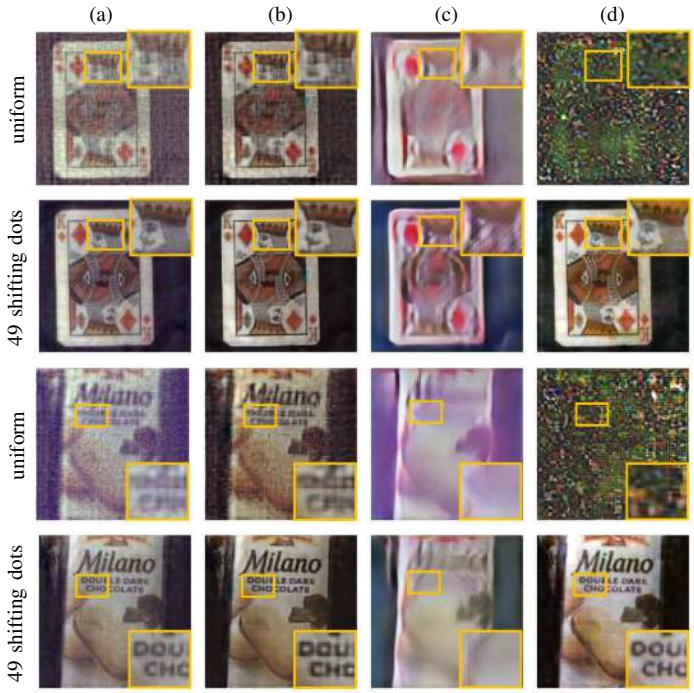


Fig. 13: Reconstruction results for real-hardware measurements with 49 uniform and shifting dots illumination patterns. Images in four columns show (a) LS solution, (b) LS solution with trained UNet refinement, (c) LS solution with pretrained FlatNet refinement, and (d) trained FlatNet that reconstructs image directly from measurements.

can be easier to implement because all the patterns are simply the shifted copies of the same base pattern.
- Programmable orthogonal patterns would provide similar gains as shifting dots with potentially better performance in strong ambient light (at the expense of complex hardware).

Limitations: leftmargin=4mm,topsep=0pt,itemsep=0ex,partopsep=1ex,parsep=0pt

- The resolution of reconstructed images are limited by the resolution of illumination patterns.
- Coded illumination hardware can bring additional cost and complexity to the system design.
- Our current system cannot capture dynamic scenes because our model requires multiple measurements of every scene under different illumination patterns.
- The shifting dots patterns are simple and easy to implement, but they reduce the light throughput.
- The camera and projector are co-located in our setup; therefore coded illumination does not provide a direct advantage toward depth estimation. Reconstruction of 3D and dynamic scenes with this framework is possible and requires additional work [60].
- Similar to other active illumination systems, the performance of our method will suffer under strong ambient lights (see supplementary).
- Our imaging model does not account for non-Lambertian objects, and any such object can cause artifacts in the reconstruction (see supplementary).

A. Error Analysis in Multi-shot System

In this section, we discuss the effects of noise and system conditioning on the reconstruction error. In our coded illumination method (especially, with shifting dots), each frame of the multi-shot system can have low SNR compared to single-shot system given the same capture time. While the SNR of each frame is worse, the multi-shot systems offer much better conditioning, which in turn provides better reconstruction.

Let us represent the measurements captured by a reference lensless camera under uniform illumination in the following vectorized form: $y = Ax + e$, where A denotes the system matrix, x denotes the unknown image, and e denotes the measurement noise. We denote the measurement signal to noise ratio of this reference system as SNR_{ref}. Note that under the assumption that photon count can be approximated as $\mathcal{N}(\lambda t, \lambda t)$, $\text{SNR}_{ref} \propto \sqrt{\lambda t}$, where λ represents rate of photon arrival and t represents exposure time).

Suppose A is full rank but ill-conditioned with singular values $s_1, \ldots, s_N$, and the least-squares solution is $\hat{x}$. If we capture T measurements with uniform illumination (or increase exposure time of a single measurement by a factor of T), then the measurement SNR will increase by a factor of $\sqrt{T}$ but the singular values remain unchanged. A simple derivation (see details in [33]) can show that mean signal to

mean reconstruction error ratio (SER) can be written as

$$\text{SER}_{unif} = \frac{\mathbb{E}\|x\|_2}{\mathbb{E}\|x-\hat{x}\|_2} = \frac{\sqrt{T}\,\text{SNR}_{ref}}{\sqrt{\text{Tr}(A^\top A)^{-1}}} = \frac{\sqrt{T}\,\text{SNR}_{ref}}{\rho_{ref}}, \tag{12}$$

where $\rho_{ref} = \text{Tr}(A^\top A)^{-1} = \sqrt{\sum_{i=1}^{N} 1/s_i^2}$. The reconstruction error is dominated by the small singular values as they will increase the denominator.

On the other hand, using coded illuminations will provide us a modified system matrix $\tilde{A}$ with singular values $\tilde{s}_1, \ldots, \tilde{s}_N$ that decay at a slower rate (as shown in Fig. 3a). The SNR of coded-illumination measurements can be different from uniform illumination; for example, if we turn on every kth pixel in the illumination pattern, the measurement SNR will reduce by a factor of $\sqrt{k}$. In our equivalent design with shifting dots, we would capture $T = k$ measurements with shifting dots. Therefore, the mean signal to mean reconstruction error ratio (SER) with T shifting dots illumination patterns can be written as

$$\text{SER}_{dots} = \frac{\mathbb{E}\|x\|_2}{\mathbb{E}\|x-\hat{x}\|_2} = \frac{\text{SNR}_{ref}}{\rho_{dots}}, \tag{13}$$

$\rho_{dots} = \text{Tr}(\tilde{A}^\top \tilde{A})^{-1} = \sqrt{\sum_{i=1}^{N} 1/\tilde{s}_i^2}$. Note that if we use T Hadamard patterns instead of shifting dots, then the SNR per measurement would not reduce, and the error bound above will not have the $\sqrt{T}$ factor in the denominator.

In summary, suppose we use same total exposure time (or number of equal-exposure measurements) for uniform and shifting dots patterns, then

$$\frac{\text{SER}_{dots}}{\text{SER}_{unif}} = \frac{1}{T}\frac{\rho_{ref}}{\rho_{dots}}. \tag{14}$$

In practice, the coded illumination-based system offers significantly better conditioning compared to reference system; that is, $\rho_{ref} \gg \rho_{dots}$.

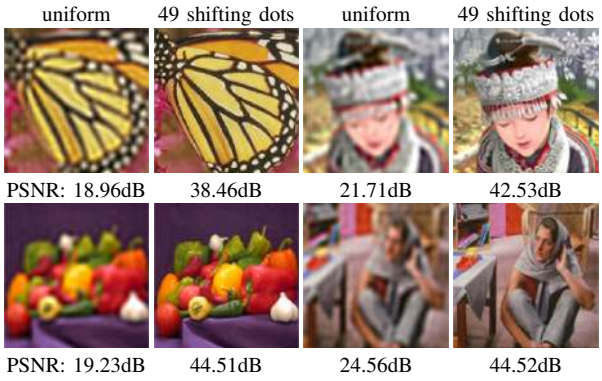
We also show experimental results for multi-shot coded illumination system and uniform pattern system with the same amount of exposure time in Fig. 9(c,d). We can observe that the 49 shifting dots outperform the reconstruction using uniform illumination.

B. Coded Illumination with Nonseparable Systems

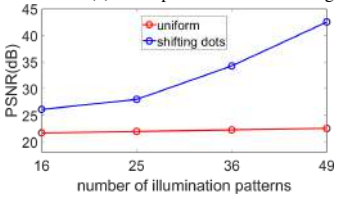
All the lensless cameras can be modeled as a linear system; therefore, the broader finding that adding coded illumination improves the conditioning of the system is valid across all the lensless designs. Our particular choice of separable masks and illumination patterns is indeed specific to FlatCam-inspired designs, where we gain computational advantages as the solution can be written in a separable closed form.

In principle, we can add coded illumination to convolutional models such as DiffuserCam as $\mathbf{Y}_i = \varphi * (\mathbf{P}_i \odot \mathbf{X})$, where φ denotes PSF for the mask, $\mathbf{X}$ denotes the unknown image, $\mathbf{P}_i$ denotes ith illumination pattern, and $\mathbf{Y}_i$ denotes the corresponding sensor measurements. We do not have a closed-form solution for such an imaging model; therefore, reconstruction involves using iterative solver.

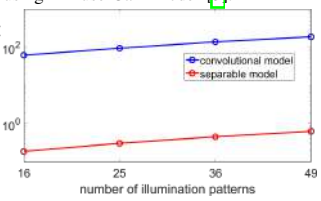
To validate our argument, we also provide simulation results in Fig. 14 combining our coded illumination system with the convolution model proposed in DiffuserCam [3]. The PSFs used in the simulation are also from the DiffuserCam data. We observe in Fig. 14 that the coded illumination pattern also improves the condition of the DiffuserCam.



(a) Example reconstructed images using DiffuserCam model [3].



(b) Average PSNR values.



(c) Average reconstruction time.

Fig. 14: Simulation results with uniform illumination and 49 shifting dots coded illumination patterns using the convolution model in DiffuserCam [3]. (a) Sample reconstructed images in (a) and average reconstruction PSNR vs number of patterns in (b) show that coded illumination patterns improve the imaging performance of DiffuserCam. (c) Average reconstruction time for the convolutional model is an order of magnitude larger than the separable model that has a simple closed-form solution.

Combining a convolution model with coded illumination has some additional practical challenges that do not arise in the separable model we presented earlier in the paper. Two main challenges we encounter are (1) alignment of spatial grids and (2) reconstruction complexity. The alignment process is necessary to avoid any model mismatch. The sampling grid in the convolutional model cannot be arbitrary, as it is determined by the PSF and sensor pixel pitch. The alignment requires additional (tedious) calibration steps to estimate relative sizes/displacements of illumination pixels and scene pixels. We leave the calibration and hardware implementation tasks for future work. In contrast, in our paper, we used a separable model that can be calibrated using the illumination pattern itself (as long as illumination grid yields a separable response on the sensor). Coded illumination-based convolutional models do not have a simple closed-form solution because the combined operator cannot be diagonalized with Fourier transform. We can implement the forward and adjoint operators using Fourier transform, but the overall reconstruction procedure is much slower than the separable model we used in our paper, as shown in Fig. 14c.

Supplementary material. Additional simulation and experimental results are in the supplementary document. The codes are available at https://github.com/CSIPlab/codedcam.

REFERENCES

[1] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2017.

[2] V. Boominathan, J. K. Adams, M. S. Asif, B. W. Avants, J. T. Robinson, R. G. Baraniuk, A. C. Sankaranarayanan, and A. Veeraraghavan, "Lensless imaging: A computational renaissance," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 23–35, 2016.

[3] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," *Optica*, vol. 5, no. 1, pp. 1–9, Jan 2018.

[4] V. Boominathan, J. T. Robinson, L. Waller, and A. Veeraraghavan, "Recent advances in lensless imaging," *Optica*, vol. 9, no. 1, pp. 1–16, 2022.

[5] T. M. Cannon and E. E. Fenimore, "Coded Aperture Imaging: Many Holes Make Light Work," *Optical Engineering*, vol. 19, p. 283, Jun. 1980.

[6] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Appl. Opt.*, vol. 17, no. 3, pp. 337–347, Feb 1978.

[7] T. Shimano, Y. Nakamura, K. Tajima, M. Sao, and T. Hoshizawa, "Lensless light-field imaging with fresnel zone aperture: quasi-coherent coding," *Appl. Opt.*, vol. 57, no. 11, pp. 2841–2850, Apr 2018.

[8] V. Boominathan, J. K. Adams, J. T. Robinson, and A. Veeraraghavan, "Phlatcam: Designed phase-mask based thin lensless camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1618–1629, 2020.

[9] A. Yang and A. C. Sankaranarayanan, "Designing display pixel layouts for under-panel cameras," *IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI) / Special Issue of ICCP*, vol. 43, no. 7, pp. 2245–2256, July 2021.

[10] P.-H. Yin, C.-W. Lu, J.-S. Wang, K.-L. Chang, F.-K. Lin, C.-J. Chang, and G.-C. Bai, "A 368 × 184 optical under-display fingerprint sensor with global shutter and high-dynamic-range operation," in *2020 IEEE Custom Integrated Circuits Conference (CICC)*, 2020, pp. 1–4.

[11] T. Yokota, K. Fukuda, and T. Someya, "Recent progress of flexible image sensors for biomedical applications," *Advanced Materials*, vol. 33, no. 19, p. 2004416, 2021.

[12] J. K. Adams, V. Boominathan, B. W. Avants, D. G. Vercosa, F. Ye, R. G. Baraniuk, J. T. Robinson, and A. Veeraraghavan, "Single-frame 3d fluorescence microscopy with ultraminiature lensless flatscope," *Science Advances*, vol. 3, no. 12, 2017.

[13] N. Chakrova, R. Heintzmann, B. Rieger, and S. Stallinga, "Studying different illumination patterns for resolution improvement in fluorescence microscopy," *Opt. Express*, vol. 23, no. 24, pp. 31 367–31 383, Nov 2015.

[14] A. Yedidia, C. Thrampoulidis, and G. Wornell, "Analysis and optimization of aperture design in computational imaging," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 4029–4033, April 2018.

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

[16] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007.

[17] M. S. Asif, "Lensless 3d imaging using mask-based cameras," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6498–6502.

[18] Y. Zheng and M. S. Asif, "Joint image and depth estimation with mask-based lensless cameras," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1167–1178, 2020.

[19] Y. Hua, S. Nakamura, M. S. Asif, and A. C. Sankaranarayanan, "Sweepcam — depth-aware lensless imaging using programmable masks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1606–1617, 2020.

[20] Y. Zheng, Y. Hua, A. C. Sankaranarayanan, and M. S. Asif, "A simple framework for 3d lensless imaging with programmable masks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2603–2612.

[21] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[22] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[23] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[24] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[25] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[26] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[27] R. Hyder, V. Shah, C. Hegde, and M. S. Asif, "Alternating phase projected gradient descent with generative priors for solving compressive phase retrieval," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7705–7709.

[28] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning*. PMLR, 2017, pp. 537–546.

[29] P. Hand, O. Leong, and V. Voroninski, "Phase retrieval under a generative prior," in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 9136–9146.

[30] S. Khan, V. Sundar, V. Boominathan, A. Veeraraghavan, and K. Mitra, "Flatnet: Towards photorealistic scene reconstruction from lensless measurements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, oct 2020.

[31] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express*, vol. 27, no. 20, pp. 28 075–28 090, Sep 2019.

[32] K. Monakhova, V. Tran, G. Kuo, and L. Waller, "Untrained networks for compressive lensless photography," *Opt. Express*, vol. 29, no. 13, pp. 20 913–20 929, Jun 2021.

[33] Y. Y. Schechner, P. N. Belhumeur, and S. K. Nayar, "Multiplexing for optimal lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 08, pp. 1339–1354, aug 2007.

[34] G. Wetzstein, I. Ihrke, and W. Heidrich, "On plenoptic multiplexing and reconstruction," *International Journal of Computer Vision*, vol. 101, pp. 384–400, 2012.

[35] M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, "Three-Dimensional Resolution Doubling in Wide-Field Fluorescence Microscopy by Structured Illumination," *Biophysical Journal*, vol. 94, pp. 4957–4970, Jun. 2008.

[36] S. K. Nayar and M. Gupta, "Diffuse structured light," in *2012 IEEE International Conference on Computational Photography (ICCP)*, April 2012, pp. 1–11.

[37] D. Fofi, T. Sliwa, and Y. Voisin, "A comparative survey on invisible structured light," in *IST/SPIE Electronic Imaging*, 2004.

[38] R. Heintzmann and C. G. Cremer, "Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating," in *Optical Biopsies and Microscopic Techniques III*, I. J. Bigio, H. Schneckenburger, J. Slavik, K. S. M.D., and P. M. Viallet, Eds., vol. 3568, International Society for Optics and Photonics. SPIE, 1999, pp. 185 – 196.

[39] M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy." *Journal of microscopy*, vol. 198 Pt 2, pp. 82–7, 2000.

[40] J. Gu, T. Kobayashi, M. Gupta, and S. K. Nayar, "Multiplexed illumination for scene recovery in the presence of global illumination," in *2011 International Conference on Computer Vision*, 2011, pp. 691–698.

[41] K. Mitra, O. S. Cossairt, and A. Veeraraghavan, "A framework for analysis of computational imaging systems: Role of signal prior, sensor noise and multiplexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 1909–1921, 2014.

[42] N. Ratner, Y. Y. Schechner, and F. Goldberg, "Optimal multiplexed sensing: bounds, conditions and a graph theory link," *Opt. Express*, vol. 15, no. 25, pp. 17 072–17 092, Dec 2007.

[43] C. B. S. Burak Gokturk, Hakan Yalcin, "A time-of-flight depth sensor - system description, issues and solutions," in *Conference on Computer Vision and Pattern Recognition Workshop*, June 2004, pp. 35–35.

[44] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, p. 45, 2013.

[45] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[46] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.
[47] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 277–299, 2015.
[48] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, "Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes," *Annual Review of Physical Chemistry*, vol. 59, no. 1, pp. 387–410, 2008, pMID: 18031219.
[49] G. Jagatap, Z. Chen, S. Nayer, C. Hegde, and N. Vaswani, "Sample efficient fourier ptychography for structured data," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 344–357, 2020.
[50] J. M. Rodenburg, "Ptychography and related diffractive imaging methods," ser. Advances in Imaging and Electron Physics, Hawkes, Ed. Elsevier, 2008, vol. 150, pp. 87–184.
[51] L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for fourier ptychography with an led array microscope," *Biomed. Opt. Express*, vol. 5, no. 7, pp. 2376–2389, Jul 2014.
[52] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch, "Dual photography," in *ACM SIGGRAPH 2005 Papers*, ser. SIGGRAPH '05. New York, NY, USA: Association for Computing Machinery, 2005, p. 745–755.
[53] 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 Processing Magazine*, vol. 25, no. 2, pp. 83–91, March 2008.
[54] P. Sen and S. Darabi, "Compressive dual photography," in *Computer Graphics Forum*, vol. 28, no. 2. Wiley Online Library, 2009, pp. 609–618.
[55] Y. Zheng, R. Zhang, and M. S. Asif, "Coded illumination and multiplexing for lensless imaging," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 9250–9253.
[56] C. S. Williams and O. A. Becklund, *Introduction to the Optical Transfer Function*, 1989.
[57] J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," in *Proc. IEEE ICCV*, 2019.
[58] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.
[59] G. Chaladze and L. kalatozishvili, "Linnaeus 5 dataset," Nov 2017.
[60] Y. Zheng and M. S. Asif, "Coded illumination for 3d lensless imaging," *IEEE Open Journal of Signal Processing*, vol. 3, pp. 432–439, 2022.

Yucheng Zheng (Student Member, IEEE) received the B.Sc. degree in electrical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China in 2017. He is currently working toward the Ph.D. degree at the University of California, Riverside, CA, USA. His current research interests include computational imaging, computer vision and signal processing.

M. Salman Asif (Senior Member, IEEE) received his B.Sc. degree from the University of Engineering and Technology, Lahore, Pakistan, and his M.S and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, Georgia, USA. He is currently an Associate Professor at the University of California Riverside, USA. Prior to that he worked as a Postdoctoral Researcher at Rice University and a Senior Research Engineer at Samsung Research America, Dallas. He has received NSF CAREER Award, Google Faculty Award, Hershel M. Rich Outstanding Invention Award, and UC Regents Faculty Fellowship and Development Awards. His research interests include computational imaging, signal/image processing, computer vision, and machine learning.