FlatCam: Replacing Lenses with Masks and Computation

M. Salman Asif[†,*], Ali Ayremlou[†], Aswin Sankaranarayanan[‡], Ashok Veeraraghavan[†], and Richard Baraniuk[†]
[†]ECE Department, Rice University, Houston, TX
[‡]ECE Department, Carnegie Mellon University, Pittsburgh, PA

Abstract

We present a thin form-factor lensless camera, FlatCam, that consists of a coded mask placed on top of a bare, conventional sensor array. FlatCam is an instance of a coded aperture imaging system in which each pixel records a linear combination of light from multiple scene elements. A computational algorithm is then used to demultiplex the recorded measurements and reconstruct an image of the scene. In contrast with vast majority of coded aperture systems, we place the coded mask extremely close to the image sensor that can enable a thin system. We use a separable mask to ensure that both calibration and image reconstruction are scalable in terms of memory requirements and computational complexity. We demonstrate the potential of our design using a prototype camera built using commercially available sensor and mask.

1. Introduction

Over the years, progress in miniature cameras has enabled new applications in endoscopy, pill cameras, and in vivo microscopy. However, the popular strategy of minimizing camera volume reduces the amount of light collected at the sensor as the lens aperture and the sensor size become small. Consequently, ultra-miniature imagers that scale down the optics and sensors suffer from low light collection.

In this paper, we present a camera architecture, *FlatCam*[1], which is inspired by coded aperture imaging principles pioneered in astronomical x-ray and gamma-ray imaging [9, 11, 6]. In the FlatCam design, we envision a camera that has a large photosensitive area but a thin form factor. The FlatCam achieves the thin form factor by dispensing with a lens and replacing it with a coded, binary mask placed almost immediately atop a bare conventional sensor array. The image formed on the sensor can be viewed as a superposition of many pinhole images. Thus, the light collection ability of such a coded aperture system is proportional to the size of the sensor and the transparent regions (pinholes) in the mask. In contrast, the light collection ability of a miniature, lens-based camera is limited by the lens aperture size, which is restricted by the requirements on the device thickness.

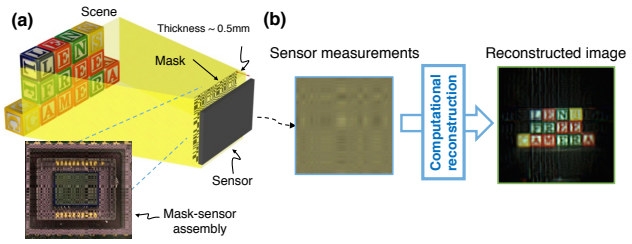

Figure 1: FlatCam architecture. (a) Every point in the scene casts an image on the sensor, all of which add up to produce the sensor measurements. **(b)** An example of sensor measurements and the reconstructed image.

An illustration of the FlatCam design is presented in Fig. 1. Light from any point in the scene passes through a coded mask and lands on a conventional image sensor. The mask consists of opaque and transparent features (to block and transmit light, respectively); each transparent feature can be viewed as a pinhole. Light gets diffracted and modulated by the mask features such that light from each scene point casts a unique mask shadow on the sensor, and this mapping can be represented using a linear operator. A computational algorithm then inverts this linear operator to recover the original light distribution of the scene from the sensor measurements.

Our FlatCam design has many attractive properties besides its thin profile. First, since it reduces the thickness of the camera but not the area of the sensor, it collects more light than miniature, lens-based cameras with same 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.

We demonstrate the potential of the FlatCam using a prototype built in our laboratory with commercially available sensor and mask. We built our prototype by placing mask at about 0.5mm from the sensor surface. Figure 2 illustrates sensor measurements and reconstructed images using our prototype FlatCam.

*Corresponding author. Email: sasif@rice.edu
[1]A longer version of the paper with additional results is available in [1].

2. Related work

Pinhole cameras, the progenitor of lens-based cameras, are typical example of imaging without a lens. However, a small pinhole, which is required for sharp image formation, allows very little light to reach the sensor, resulting in noisy, low-quality images. Indeed, lenses were introduced into cameras to increase the aperture size, and thus the light throughput, without degrading the image sharpness.

Coded aperture cameras extend the idea of a pinhole camera by using masks with multiple *pinholes* [9, 11]. The primary goal of coded aperture cameras is to increase the light throughput compared to a pinhole camera. Coded-aperture cameras have traditionally been used for imaging wavelengths beyond the visible spectrum (e.g., x-ray and gamma-ray imaging), for which lenses or mirrors are expensive or infeasible. [9, 11, 6, 3]. In recent years, coded aperture-based systems using compressive sensing principles [5, 10, 2] have been studied for image super-resolution [17], spectral imaging [20], and video capture [16]. Mask-based lens-free designs have also been proposed for flexible field-of-view selection in [21], compressive single-pixel imaging using a transmissive LCD panel [15], and for separable coded masks [8].

Existing coded aperture-based lensless systems have two main limitations: First, a majority of coded aperture systems place the mask significantly far away from the sensor (e.g., 65mm distance in [8]). In contrast, our FlatCam design offers a thin form factor. For instance, in our prototype with a visible sensor, the spacing between the sensor and the mask is only 0.5mm. Second, the masks in some designs have transparent features only in a central region that is much smaller than the sensor size. In contrast, almost half of the features (spread across the entire surface) in our mask are transparent. As a consequence, the light throughput of our designs are many orders of magnitude larger as compared to previous designs. Furthermore, the lensless cameras proposed in [15, 8] use programmable spatial light modulators (SLM) and capture multiple images while changing the mask patterns. In contrast, we use a static mask in our design, which can potentially be fixed on the sensor during fabrication or the assembly process.

In last few decades, a number of micorlens-based thin imaging systems have also been developed. The TOMBO architecture [19], inspired by insect compound eyes, reduces the camera thickness by replacing a single, large focal-length lens with multiple, small focal-length microlenses. Each microlens and the sensor area underneath it can be viewed as a separate low-resolution, lens-based camera, and a single high-resolution image can be computationally reconstructed by fusing all of the sensor measurements. Similar architectures have been used for designing thin infrared cameras [18]. The camera thickness in this design is dictated by the geometry of the microlenses; reducing the camera thickness requires a proportional reduction in the sizes of the microlenses and sensor pixels. As a result, microlens-based cameras currently offer only up to a four-fold reduction in the camera thickness [4].

Recently, miniature cameras with integrated diffraction gratings and CMOS image sensors have been developed [12, 13]. These cameras have been successfully demonstrated on tasks such as motion estimation and face detection. While these cameras are indeed ultra-miniature in total volume (100 micron sensor width by 200 micron thickness), they retain the large thickness-to-width ratio of conventional lens-based cameras. Because of the small sensor size, they suffer from reduced light collection ability. In contrast, in our visible prototype below, we used a 6.7mm wide square sensor, which increases the amount of light collection by about three orders of magnitude, while the device thickness remains approximately similar (500 micron).

3. FlatCam design

Our FlatCam design places an amplitude mask (with transparent and opaque features that transmit and block light, respectively) almost immediately in front of the sensor array (see Fig. 1). While we focus on a single mask for exposition purposes, the concept extends to multiple amplitude masks in a straightforward manner.

The mask pattern determines the light throughput of the system and the complexity of system calibration and inversion. In our designs, half of the binary mask features are transparent, which halves our light collection ability compared to the maximum limit. However, the main advantage of the FlatCam design is that it allows us to use a large sensor arrays for a given device thickness constraint, thereby significantly increasing the light collection capabilities of devices under thickness constraints.

3.1. Image formation

Light from the scene is modulated and diffracted by the mask and recorded on the image sensor. By assuming that the image formed on the sensor is a superposition of light sources in the scene, we can describe the transfer function between the scene image and the sensor measurements as

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

Here, x denotes the pixelated scene image, y denotes the sensor measurements, Φ denotes the transfer matrix, and e denotes the sensor noise and any model mismatch. Since each sensor pixel measures multiplexed light from multiple scene pixels, each row of Φ encodes how strongly each scene pixel contributes to the intensity measured at a particular sensor pixel. In other words, any column in Φ denotes the image formed on the sensor if the scene contains a single, point light source at the respective location.

3.2. Separable mask pattern

The (linear) relationship between the scene irradiance x and the sensor measurements y is contained in the multiplexing matrix Φ. Discretizing the unknown scene irradiance into $N \times N$ pixels and assuming an $M \times M$ sensor array, Φ is an $M^2 \times N^2$ matrix. Given a mask and sensor, we can obtain the entries of Φ either by modeling the transmission of light from the scene to the sensor or through a calibration process. Clearly, even for moderately sized systems, Φ is prohibitively large to either estimate (calibration) or invert (image reconstruction), in general. For example, to describe a system with a megapixel resolution scene and a megapixel sensor array, Φ will contain on the order of $10^6 \times 10^6 = 10^{12}$ elements.

One way to reduce the complexity of Φ is to use a separable mask for the FlatCam system. If the mask pattern is separable (i.e., an outer product of two one-dimensional patterns), then the imaging system in (1) can be rewritten as

$$Y = \Phi_L X \Phi_R^T + E, \tag{2}$$

where Φ_L, Φ_R denote matrices that correspond to one-dimensional convolution along the rows and columns of the scene, respectively, X is an $N \times N$ matrix containing the scene radiance, Y in an $M \times M$ matrix containing the sensor measurements, and E denotes the sensor noise and any model mismatch. For a megapixel scene and a megapixel sensor, Φ_L and Φ_R have only 10^6 elements each, as opposed to 10^{12} elements in Φ. Similar idea has been recently proposed in [8] with the design of doubly toeplitz mask. In our implementation, we also estimate the system matrices using a separate calibration procedure, which also becomes significantly simpler for a separable system.

In addition to simplifying the calibration task, separability of the coded mask also significantly reduces the computational burden of image reconstruction. Iterative methods for solving the optimization problems require repeated applications of the multiplexing matrix and its transpose. Continuing our numerical example from above, for a non-separable, dense mask, both of these operations would require on the order of 10^{12} multiplications and additions for mega-pixel images. With a separable mask, however, the application of the forward and transpose operators requires only on the order of 2×10^9 scalar multiplications and additions—a tremendous reduction in computational complexity.

3.3. Image reconstruction

Given a set of $M \times M$ sensor measurements Y, our ability to invert the system (2) to recover the desired $N \times N$ image X primarily depends on the rank and the condition number of the system matrices Φ_L, Φ_R. If both Φ_L and Φ_R are well-conditioned, then we can estimate X by solving a simple least-squares problem:

$$\widehat{X}_{\text{LS}} = \arg\min_X \ \|\Phi_L X \Phi_R^T - Y\|_2^2, \tag{3}$$

which has a closed form solution. However, if Φ_L, Φ_R are not well-conditioned, the least-squares estimate $\widehat{X}_{\text{LS}}$ suffers from noise amplification. A simple approach to reduce noise amplification is to add an ℓ_2 regularization term in the least-squares problem in (3)

$$\widehat{X}_{\text{Tik}} = \arg\min_X \ \|\Phi_L X \Phi_R^T - Y\|_2^2 + \tau \|X\|_2^2, \tag{4}$$

where $\tau > 0$ is a regularization parameter. The solution of (4) can be explicitly written using the SVD of Φ_L, Φ_R as

$$\widehat{X}_{\text{Tik}} = V_L[(\Sigma_L U_L^T Y U_R \Sigma_R)./(\sigma_L \sigma_R^T + \tau \mathbf{1}\mathbf{1}^T)]V_R^T, \tag{5}$$

where σ_L and σ_R denote the diagonal entries of Σ_L^2 and Σ_R^2, respectively, and $A./B$ denote element-wise division of matrices A and B. Thus, once the SVDs of Φ_L and Φ_R are computed and stored, reconstruction of an $N \times N$ image from $M \times M$ sensor measurements involves a fixed cost of two $M \times N$ matrix multiplications, two $N \times N$ matrix multiplications, and three $N \times N$ diagonal matrix multiplications.

4. Experimental results

We built this FlatCam prototype using a Sony ICX285 CCD color sensor inside a Point Grey Grasshopper 3 camera. The sensor has 1086×1384 pixels, each 6.45μm wide, arranged in an RGB Bayer pattern. The physical size of the sensor array is approximately 6.7mm $\times$ 8.9mm. We used a custom-made chrome-on-quartz photomask, one side of which is covered with a pattern defined using a thin chrome film. The transparent regions of the mask transmit light, while the chrome film regions of the mask block light. We created the binary mask pattern using the outer product of two M-sequences [14]; since the M-sequences contain ± 1 entries, we replaced every -1 with a 0 in the resulting outer product. We printed the mask such that each element corresponds to a 30μm square box (transparent, if 1; opaque, if 0) on the printed mask. Even though the binary mask is not separable as is, we can represent the sensor image using the separable system described in (2) by subtracting the row and column mean from the sensor images.

We opened the camera body to expose the sensor surface and placed the quartz mask on top of it using mechanical posts such that the mask touches the protective glass (hot mirror) on top of the sensor. Thus, the distance between the mask and the sensor is determined by thickness of the glass, which for this sensor is 0.5mm. In most of our experiments, the exposure time was fixed at 10ms, but we adjusted it according to the scene intensity to avoid excessively bright or dark sensor images. For the static scenes we averaged 20

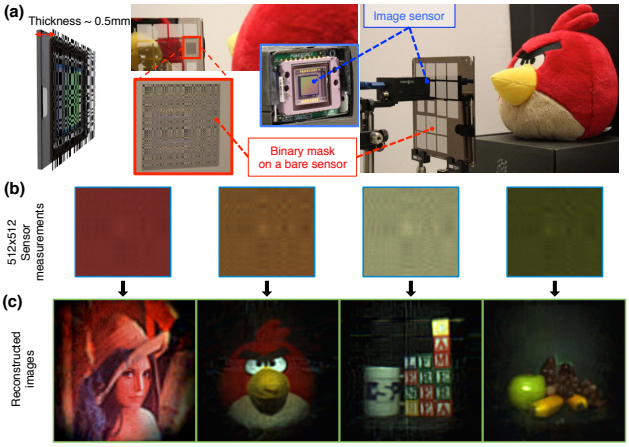

Figure 2: FlatCam prototype and results. **(a)** CCD sensor with a separable M-sequence mask. **(b)** Sensor measurements are linear combinations of the light from all scene locations. **(c)** Reconstructed 512×512 color images by processing each color channel independently.

sensor images to create a single set of measurements to be used for reconstruction.

We reconstructed 512×512 RGB images from our prototype using 512×512 RGB sensor measurements. Since the sensor has 1086×1384 pixels, we first cropped and uniformly subsampled the sensor image to create an effective 512×512 color sensor image; then we subtracted the row and column means from that image. The resulting image corresponds to the measurements described by (2), which we used to reconstruct the desired image X. Some example sensor images and corresponding reconstruction results are shown in Fig. 2. In these experiments, we solved an ℓ_2-regularized least-squares problem in (4), followed by BM3D denoising [7].

5. Conclusion

The mask-based, lens-free FlatCam design proposed here can have a significant impact in an important emerging area of imaging, since high-performance, broad-spectrum cameras can be monolithically fabricated instead of requiring cumbersome post-fabrication assembly. The thin form factor and low cost of lens-free cameras makes them ideally suited for many applications in surveillance, large surface cameras, flexible or foldable cameras, disaster recovery, and beyond, where cameras are either disposable resources or integrated in flat or flexible surfaces and therefore have to satisfy strict thickness constraints. Emerging applications like wearable devices, internet-of-things, and in-vivo imaging could also benefit from the FlatCam approach.

Acknowledgments

This work was supported by NSF grants IIS–1116718, CCF–1117939, and CCF–1527501.

References

[1] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. Baraniuk. Flatcam: Thin, bare-sensor cameras using coded aperture and computation. Available online: arXiv 1509.00116. 1

[2] R. G. Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24(4), 2007. 2

[3] D. J. Brady. *Optical imaging and spectroscopy*. John Wiley & Sons, 2009. 2

[4] A. Brückner, J. Duparré, R. Leitel, P. Dannberg, A. Bräuer, and A. Tünnermann. Thin wafer-level camera lenses inspired by insect compound eyes. *Optics Express*, 18(24):24379–24394, 2010. 2

[5] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006. 2

[6] T. Cannon and E. Fenimore. Coded aperture imaging: Many holes make light work. *Optical Engineering*, 19(3):193–283, 1980. 1, 2

[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 4

[8] M. J. DeWeert and B. P. Farm. Lensless coded-aperture imaging with separable doubly-toeplitz masks. *Optical Engineering*, 54(2):023102–023102, 2015. 2, 3

[9] R. Dicke. Scatter-hole cameras for x-rays and gamma rays. *The Astrophysical Journal*, 153:L101, 1968. 1, 2

[10] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. 2

[11] E. Fenimore and T. Cannon. Coded aperture imaging with uniformly redundant arrays. *Applied optics*, 17(3):337–347, 1978. 1, 2

[12] P. R. Gill, C. Lee, D.-G. Lee, A. Wang, and A. Molnar. A microscale camera using direct fourier-domain scene capture. *Optics letters*, 36(15):2949–2951, 2011. 2

[13] P. R. Gill and D. G. Stork. Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings. In *Computational Optical Sensing and Imaging*, pages CW4C–3, 2013. 2

[14] S. W. Golomb. *Shift register sequences*. Aegean Park Press, 1982. 3

[15] G. Huang, H. Jiang, K. Matthews, and P. Wilford. Lensless imaging by compressive sensing. In *20th IEEE International Conference on Image Processing*, pages 2101–2105, 2013. 2

[16] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady. Coded aperture compressive temporal imaging. *Optics express*, 21(9):10526–10545, 2013. 2

[17] R. F. Marcia and R. M. Willett. Compressive coded aperture superresolution image reconstruction. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 833–836, 2008. 2

[18] M. Shankar, R. Willett, N. Pitsianis, T. Schulz, R. Gibbons, R. Te Kolste, J. Carriere, C. Chen, D. Prather, and D. Brady. Thin infrared imaging systems through multichannel sampling. *Applied optics*, 47(10):B1–B10, 2008. 2

[19] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (TOMBO): concept and experimental verification. *Applied optics*, 40(11):1806–1813, 2001. 2

[20] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51, 2008. 2

[21] A. Zomet and S. K. Nayar. Lensless imaging with a controllable aperture. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 339–346, 2006. 2