

FLAT FOCUS: DEPTH OF FIELD ANALYSIS FOR THE FLATCAM LENSLESS IMAGING SYSTEM

Jasper Tan, Vivek Boominathan, Ashok Veeraraghavan, and Richard Baraniuk

Rice University, Houston, TX, USA

ABSTRACT

Lensless imaging systems, such as the recently proposed FlatCam, offer numerous advantages over lens-based systems such as a thin form-factor, low cost, and higher light throughput. However, little work has been done in analyzing these systems' depth of field characteristics. A depth-dependent calibration step is necessary to obtain the image from the FlatCam measurements, and this calibration determines the system's depth of field. In this paper, we characterize the FlatCam's depth of field properties and show that (a) for scene depths on the order of tens of centimeters, it is possible to perform depth-selective refocusing from a single captured image and (b) for sufficiently large scene depths, calibrating for one depth can provide a very large depth of field.

Index Terms— Lensless imaging, depth of field, coded aperture, selective refocusing

1. INTRODUCTION

Recent years have seen the emergence of a new generation of imaging systems: lensless cameras. While the majority of imaging devices rely on optical lenses to focus scene light into images, lensless systems instead encode the scene into its measurements by replacing the lens with a light modulator atop the image sensor. Researchers have used various types of light modulators for these lensless systems such as liquid crystal modulators [1], programmable spatial-light-modulators [2], and chrome-on-quartz photomasks [3]. Lensless imaging has also been integrated into other research areas such as compressive sensing [4] and microscopy [5, 6].

Lensless imaging systems offer numerous advantages over lens-based cameras. Lenses force constraints on both the physical geometry of the device and the geometric mapping of the scene to the image. Flexible or curved lens-based cameras, for example, are currently not possible due to the rigidity and bulk of lenses, which are eliminated in lensless systems. Also, with lensless systems, the light modulating element atop the sensor can be configured in infinitely many ways, creating numerous possible mappings from the scene to the sensor. Furthermore, lensless systems can achieve lower manufacturing cost, better wavelength scaling, and higher light throughput compared to lens-based cameras [7].

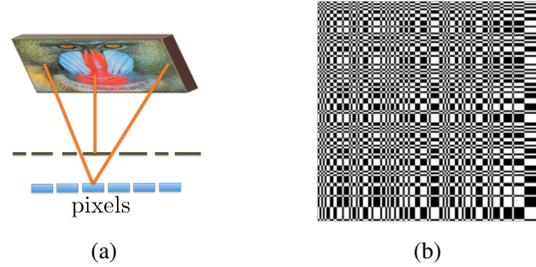


Fig. 1: (a) A simple 1D model of FlatCam. A single pixel receives light rays (orange) coming from multiple light sources. The mask selects which light rays to pass and which to block. (b) [3] suggests that the outer product of two Maximum Length Sequences [8], shown here, has robust reconstruction properties for the mask's apertures.

The depth of field characteristics of lens-based cameras are well studied. Focusing is achieved by altering the distance between the lens aperture and the sensor. A wide or narrow depth of field is obtained by shrinking or opening the lens aperture, respectively. However, the depth of field properties of lensless systems are not as well understood.

This paper aims to characterize the depth of field properties of FlatCam [3], a lensless imaging system that uses a mask of multiple pinhole-like openings or apertures atop a bare image sensor. A major factor in this system's depth of field characteristics is the reconstruction procedure necessary to obtain the scene image from the sensor measurements. The parameters of FlatCam's reconstruction algorithm, obtained via a calibration step, determine which scene depths are in focus in the image. Each calibration procedure is performed for a certain scene depth, thus optimizing the reconstruction for focusing on that depth. Multiple calibration procedures can be performed to obtain reconstruction parameters for multiple scene depths.

Two key results follow from our analysis. First, for small distances on the order of tens of centimeters or shorter, refocusing can be performed by parametrizing the reconstruction according to the desired depth of focus. Thus, multiple images with different focus depths can be obtained from a single FlatCam capture. Second, for greater distances on the order of a foot and larger, a large depth of field is achievable with

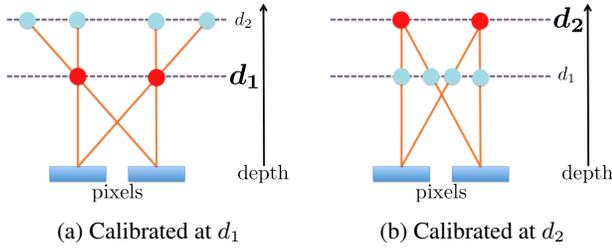


Fig. 2: A 1D model of FlatCam with $M = 2$ sensor pixels and $N = 2$ scene points is shown. The aim is to reconstruct the $N = 2$ scene points. (a) The system is calibrated at d_1 . The $M \times N = 4$ rays intersect at $N = 2$ scene points at this depth plane. This Φ matrix would not represent light sources from depth d_2 as well because the same light rays do not intersect the depth plane at $N = 2$ scene points. (b) To more accurately model scene points in depth d_2 , the elements of Φ should be changed to describe a different set of light rays.

only one set of reconstruction parameters, thereby showing that only one calibration is necessary for imaging at a large range of depths for farther scenes.

2. THE FLATCAM LENSLESS IMAGING SYSTEM

This section provides a brief overview of FlatCam. We refer the interested reader to [3, 7] for a more detailed explanation of the system.

2.1. Linear Combination of Scene Points

In a lens-based system, the optical lens forces each sensor pixel to only interact with a small portion of the scene. Without the lens, an individual sensor pixel would average light from all the points in the scene. By placing a mask of apertures atop a bare image sensor, the sensor pixel only receives intensities from scene points whose light rays to the sensor pixel would intersect an aperture, as illustrated in Figure 1.

Each sensor pixel then sees a weighted or linear combination of all the scene points, where these weights depend heavily on the location of the apertures. For example, the weight of a scene point that has an aperture on its path to the sensor pixel would have a larger weight than one that does not. The measurement of an individual sensor pixel, y_i , can then be written as $y_i = \langle \phi_i, x \rangle$, where x is the vectorized form of the scene points' intensities and ϕ_i is the vector containing the weights applied to each scene point's intensity. Extending this to multiple sensor pixels yields

$$y = \Phi x, \quad (1)$$

where Φ is a matrix with each row containing the weights applied to the scene points for an individual sensor pixel. A separable mask can also be used to yield a separable Φ matrix, decreasing the number of its needed elements [3]. When

FlatCam captures an image, the goal is to recover the scene intensities x from the sensor measurements y . Given Φ , this becomes an inverse problem that can be solved using methods such as truncated SVD [9], Tikhonov regularization, or optimization-based methods.

2.2. Ray Tracing and Dependence on Depth

Φ_{ij} , or the element in the i th row and j th column of the Φ matrix refers to how much of the irradiance of scene point x_j is captured by the sensor pixel y_i . One way to obtain a rough estimate of the elements of Φ is to use ray tracing for every possible pair of sensor pixel and scene point.

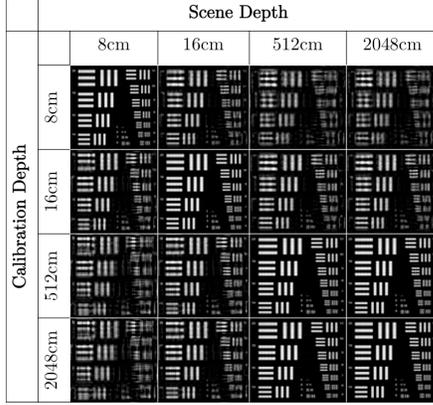
A simple case of ray tracing is to assume a two-dimensional scene that resides on a plane parallel to the image sensor. A grid of locations on this plane is used to define the locations of the scene points, x . The perpendicular distance between the scene plane and the image sensor is the *scene depth*. A ray is then traced from each scene point x_j to each sensor pixel y_i . In the simplest case, we can assign Φ_{ij} to be 1 if the ray intersects a pinhole and 0 otherwise. In reality, Φ would contain values between 0 and 1 because of light diffraction, sensor characteristics, and other physical effects.

In this way, each element of Φ describes the travel of a ray of light from the scene, through the FlatCam mask, and into the sensor. Let M be the number of elements of y and N be the number of elements of x . While there are $M \times N$ rays being described in Φ , they each only intersect the scene plane at one of N scene points, which make up the N pixels in the reconstructed image. However, at any other scene depth plane aside from the initial one, these rays will intersect at more than N points. Attempting to reconstruct N scene points at another scene depth gives an image that is blurred. This is illustrated in Figure 2.

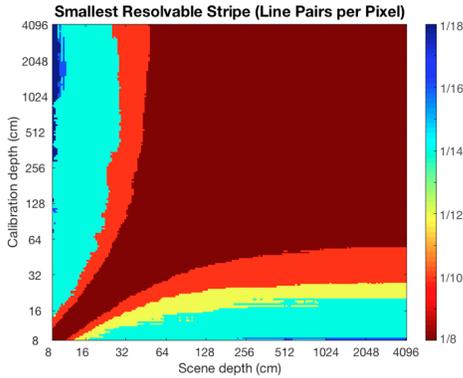
While ray tracing provides an intuitive understanding of the Φ matrix, in reality, it is more practical to obtain Φ using a depth-dependent calibration procedure explained in [3, 7]. The scene depth used in obtaining the Φ matrix is the *calibration depth*. Calibrating for different calibration depths yields different Φ matrices, with each Φ being optimized for reconstructing images captured at its calibration depth. Reconstructing an image captured at a depth other than the calibration depth may yield a blurred image, according to reasons discussed above.

3. DEPTH OF FIELD ANALYSIS

Typically, one captures images of objects at multiple distances, and it is rare that the imaged object is at a precisely known depth. The choice of the calibration depth used to obtain the Φ matrix for reconstruction may then become crucial for image quality. In this section, we evaluate the resolution quality of the reconstructed image as a function of the calibration depth and the actual scene depth (the object's



(a)



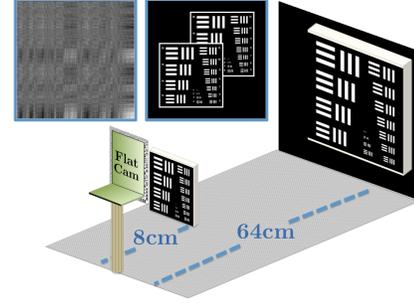
(b)

Fig. 3: (a) Simulated reconstructed images for various calibration and scene depths. The image is clearest when the two depths are equal. The quality difference for larger depths is very small. (b) A heatmap of the smallest stripe pattern with $MC \geq 0.15$. The calibration depth (y -axis) is the depth the Φ matrix is modeled for reconstruction while the scene depth (x -axis) is the depth used to simulate image capture.

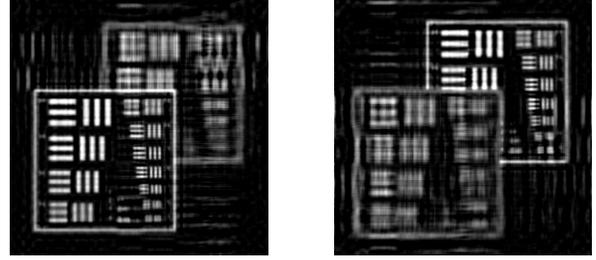
depth during capture) using simulation based on ray-tracing methods explained above. We simulate our sensor to have 512×512 pixels, a pixel pitch of $5.3 \mu\text{m}$, 12-bit precision with 1-bit noise, and a half-width half-maximum acceptance angle of 25° . The scene is discretized to have 512×512 pixels. We simulate our mask to be 1 mm away from the sensor with apertures $25 \mu\text{m} \times 25 \mu\text{m}$ in size in the pattern of the outer product of two maximum length sequences (MLS) [8]. A prototype with these parameters was reported in [7].

To quantify the resolution of a reconstructed image, we calculate the Modulation Transfer Function (MTF) with the technique used by [10]. The FlatCam is used to image a regular pattern of 3 white stripes alternating with 2 black stripes, a setup that mimics a USAF 1951 resolution target. We then define the modulation contrast (MC) to be

$$MC = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}},$$



(a) Simulation setup



(b) 8 cm reconstruction

(c) 64 cm reconstruction

Fig. 4: (a) FlatCam is simulated to capture two objects: one at 8 cm away and one at 64 cm away. The insets show the FlatCam sensor measurements and the image that would be captured by a standard camera. (b-c) Images are reconstructed from the FlatCam measurements with two Φ matrices: one calibrated at 8 cm and one calibrated at 64 cm. The reconstructed images show that one can select which object to focus on by choosing the appropriate Φ matrix for reconstruction.

where I_{\max} is the mean of the 3 maxima of the white stripes and I_{\min} is the mean of the 2 minima of the black stripes of the resulting reconstructed image. In the ideal case, the white stripes would have a value of 1, and the black stripes would have a value of 0, yielding $MC = 1$. The worst case scenario is when the stripes blur and bleed into each other, leaving a uniform intensity throughout the image and yielding $MC = 0$. We define $MC = 0.15$ to be the minimum quantity such that the stripes can be visually resolved.

The target image the FlatCam is simulated to capture is an image containing patterns of stripes with different widths, as seen in Figure 3. Since it contains maximum contrast, this test image provides the worst case scenario for the resolution quality of the reconstruction of any scene. The target image is simulated to cover the entire field of view of FlatCam, irrespective of scene depth, discretized to 512×512 . The field of view is determined by the sensor pixel's acceptance angle, in this case -25° to 25° . This results in the two center scene points having an angular distance of approximately 0.109° from the sensor's perspective regardless of depth.

We simulate the FlatCam capturing this image at a number of depths (the scene depths) using a method based on ray tracing. We simulate calibration by obtaining Φ matrices for dif-

ferent calibration depths using ray tracing methods. For each of the captured images, we solve for x in (1) using Tikhonov regularization for the different Φ matrices (different calibration depths). Figure 3a shows some samples of reconstructed images for a few calibration and scene depths. The images are clearest when the calibration depths equal the scene depths, and images at larger depths have very similar quality.

For each of these reconstructed images, we then calculate the MC value for each stripe width. Larger stripes would be easier to resolve, thus yielding larger MC values. Figure 3b shows a heatmap of the stripe width such that $MC \geq 0.15$ for all horizontal stripes of that size and larger, that is, the smallest resolvable stripe.

The heatmap shows that the narrowest stripes resolvable by this simulated system are those with $\frac{1}{8}$ line pairs per pixel. For a given calibration depth, we define the imaging depth of field to be the range of scene depths such that the reconstructed image can resolve the stripes of $\frac{1}{8}$ line pairs per pixel. The heatmap shows that the resolution improves as the scene approaches the calibration depth. Note that for calibration distances under around 50 cm, the depth of field is very narrow. This allows for selective refocusing because reconstructing at a particular calibration depth leaves everything outside of this narrow field out of focus. For distances larger than 50 cm, the depth of field is very large. This allows imaging for a large range of scene depths with only one calibration. Both properties are desirable. The former allows post-capture refocusing for small distances, while the latter allows a wide range of imaging with minimal calibrations.

4. SELECTIVE REFOCUSING

The narrow depth of field for small depths enables selective refocusing. Given two objects at different depths and one FlatCam image capture, one can choose which object to focus on by selecting the calibration depth with which to perform image reconstruction, provided the objects are not both at a large depth. This is in contrast to lens-based systems where the distance of the lens and the sensor determines the depth of focus and cannot be decided post-capture. Figure 4 shows a simulation wherein FlatCam captures a scene with two objects at different depths. With only one image capture, one can reconstruct two different images focusing on each of the objects. [7] shows this experimentally with a FlatCam prototype for depths of 7 cm and 27 cm, as shown in Figure 5.

5. MINIMAL CALIBRATIONS FOR LARGE DEPTH OF FIELD

Another observation from Figure 3b is that larger calibration depths have a large imaging depth of field. For example, a calibration depth of 55.9 cm has a depth of field ranging from 25.5 cm onwards. If one is limited to only one calibration, calibrating at 55.9 cm still allows focused imaging of objects

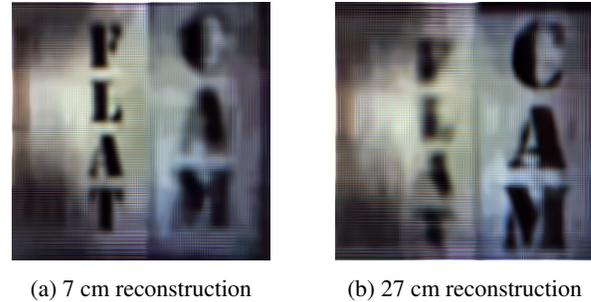


Fig. 5: A FlatCam prototype is used to take a photo of two words. The word “FLAT” is 7 cm from the system, while the word “CAM” is 27 cm away. The images are reconstructed with the two calibration depths, resulting in photos focused on the different words. This figure is modified from [7].

from 25.5 cm onwards. By choosing additional calibration depths wisely, one can expand the depth of field to include scene depths below 25.5 cm with minimal calibrations.

Table 1: Calibration Depths for Complete Depth of Field.

Calibration Depth (cm)	Imaging Depth of Field (cm)
8.0	8.0 – 9.4
10.0	8.3 – 12.4
17.0	12.4 – 25.5
55.9	25.5 onwards

Table 1 shows for this particular simulation the calibration depths that would yield the largest depth of field. Only four calibrations are required to have a depth of field from 8cm onwards. While these numbers may not hold exactly for FlatCam systems with different physical specifications, the trend is expected to be similar, such that a few calibrations would be sufficient to enable imaging at a large range of depths.

6. CONCLUSION

Lensless cameras, such as the FlatCam, provide new imaging architectures with numerous advantages over lens-based systems. By characterizing the FlatCam’s depth of field properties, we show that it is able to selectively refocus at different depths after image capture and to image at a large depth of field with only one calibration. Such features would be beneficial for using FlatCam in real applications.

Acknowledgments

This work was supported in part by NSF grants CCF-1527501, CCF-1502875, and CNS-1338099, by NGA grant HM04761510007, and by the Ken Kennedy Institute Computer Science & Engineering Enhancement Fellowship, funded by the Rice Oil & Gas HPC Conference.

7. REFERENCES

- [1] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *2006 IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognition*, vol. 1, June 2006, pp. 339–346.
- [2] M. J. DeWeert and B. P. Farm, "Lensless coded-aperture imaging with separable Doubly-Toeplitz masks," *Optical Eng.*, vol. 54, no. 2, p. 023102, Feb. 2015.
- [3] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Trans. Comput. Imag.*, July 2016, In press.
- [4] G. Huang, H. Jiang, K. Matthews, and P. Wilford, "Lensless imaging by compressive sensing," in *2013 IEEE Int. Conf. on Image Process.*, Sept 2013, pp. 2101–2105.
- [5] X. Cui, L. M. Lee, X. Heng, W. Zhong, P. W. Sternberg, D. Psaltis, and C. Yang, "Lensless high-resolution on-chip optofluidic microscopes for caenorhabditis elegans and cell imaging," *Proc. Nat. Academy of Sci.*, vol. 105, no. 31, pp. 10 670–10 675, Aug. 2008.
- [6] H. M. L. Faulkner and J. M. Rodenburg, "Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm," *Phys. Rev. Lett.*, vol. 93, p. 023903, July 2004.
- [7] 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 Process. Mag.*, vol. 33, no. 5, pp. 23–35, Sept. 2016.
- [8] F. J. MacWilliams and N. J. A. Sloane, "Pseudo-random sequences and arrays," *Proc. IEEE*, vol. 64, no. 12, pp. 1715–1729, Dec. 1976.
- [9] P. Xu, "Truncated SVD methods for discrete linear ill-posed problems," *Geophysical J. Int.*, vol. 135, no. 2, pp. 505–514, June 1998.
- [10] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3-d deconvolution for the light field microscope," *Optics Exp.*, vol. 21, no. 21, pp. 25 418–25 439, Oct. 2013.