Coded Focal Stack Photography

Xing Lin[1,2], Jinli Suo[1], Gordon Wetzstein[2], Qionghai Dai[1], Ramesh Raskar[2]
[1]Department of Automation, Tsinghua University, [2]MIT Media Lab

Abstract

We present coded focal stack photography as a computational photography paradigm that combines a focal sweep and a coded sensor readout with novel computational algorithms. We demonstrate various applications of coded focal stacks, including photography with programmable non-planar focal surfaces and multiplexed focal stack acquisition. By leveraging sparse coding techniques, coded focal stacks can also be used to recover a full-resolution depth and all-in-focus (AIF) image from a single photograph. Coded focal stack photography is a significant step towards a computational camera architecture that facilitates high-resolution post-capture refocusing, flexible depth of field, and 3D imaging.

1. Introduction

Traditional photography requires a user to carefully choose the camera parameters before taking an image. Moreover, in almost every commercial camera the focal surface is approximately planar over the sensor's area. Nowadays, some of the most important applications of computational cameras are post-capture refocusing, flexible control and extension of depth of field, and synthetic aperture imaging. For any camera designs supporting these applications, high-dimensional visual information, such as an all-in-focus (AIF) image and depth or the light field, has to be optically encoded in a recorded sensor image and computationally decoded. Currently available consumer products that follow this paradigm include time-of-flight (ToF) and light field cameras. Unfortunately, ToF cameras require extensive modification of the sensor circuitry and on-board processing, while light field cameras currently achieve only low image resolution.

We define a focal stack as a 3D volume composed of a sequence of images captured under different focal settings. With a focal sweep, we refer to a single exposure that captures a focal stack integrated over its depth.

In this paper, we explore coded focal stack photography– a family of computational photography techniques that combines a single-shot focal sweep and a coded sensor pixel readout. Previously proposed depth and light field cameras aim at capturing high-dimensional visual information in a scene; this facilitates a wide range of post-processing algorithms. In contrast, we propose flexible capture modes that are tailored to a variety of different applications and allow for the capture of higher-resolution images or reducing the requirements on computational resources.

We achieve these benefits by sweeping the focal plane over a range of depths within the exposure time of a single photograph while selectively modulating the readout pat-

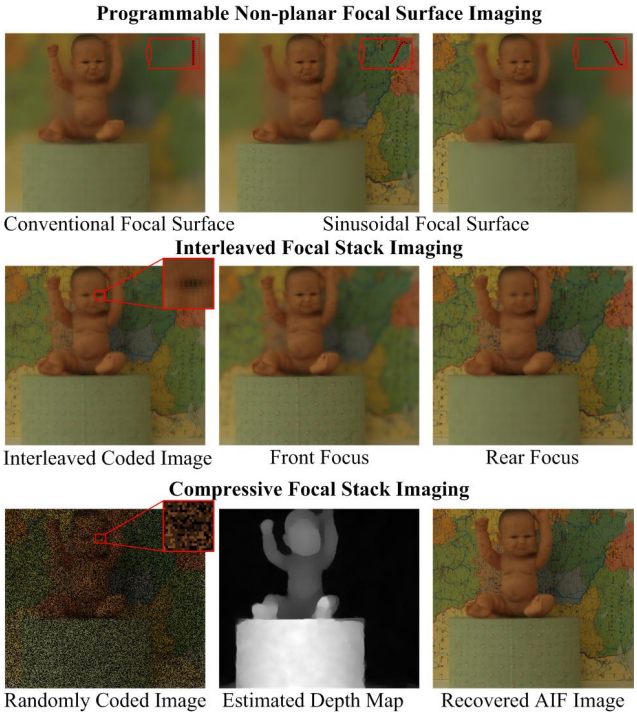


Figure 1: Coded focal stack photography explores the combination of focal sweep and per-pixel coded sensor readout. We demonstrate how to create programmable non-planar focal surfaces in photographs (top row) and how to multiplex a focal stack into a single sensor image (center row). Furthermore, we explore compressive single-shot focal stack acquisition and reconstruction which allows for full-resolution depth and AIF image to be computed (bottom row).

terns of each pixel. This approach is shown to allow for interesting all-optical effects, such as programmable unconventional and non-planar focal surfaces (Fig. 1, top row). Alternatively, the focal stack can be multiplexed into a sensor image and reconstructed using sampling and interpolation techniques inspired by the color de-mosaicing method (Fig. 1, center row). We also introduce compressive focal stack photography, where a coded projection of the focal stack is acquired and used to reconstruct the depth and all-in-focus image through sparse coding (Fig. 1, bottom row). An overview of our coded focal stack photography approach is shown in Fig. 2.

In particular, we make the following contributions:

- We introduce coded focal stack photography as a combination of a focal sweep and a coded sensor readout. We show how this approach facilitates new applications, including programmable non-planar focal surface imaging and interleaved focal stack acquisition.
- We propose compressive focal stack photography, an approach that exploits sparsity in focal stack for recovering a full-resolution depth and AIF image from a single sensor image.
- We evaluate our computational camera architecture in simulation and build a prototype implementation.

2. Related Work

Depth from Focus/Defocus (DFF/DFD). DFF/DFD is a class of techniques that use defocus cues to infer scene depth. These methods have been studied extensively over past decades (e.g, [7]). Usually, DFF determines the local depth according to the focus setting of highest sharpness in the focal stack [21]; a large number of defocused images is required to ensure sufficient reconstruction accuracy. DFD infers scene depth from the amount of blur using a single [19] or multiple [6] defocused images. Accuracy of depth estimation from a single image is limited due to the intrinsic ambiguity in depth inference. Hence, most DFD algorithms require two or more images. In Section 4, we introduce a mathematical framework to robustly reconstruct image and depth from the single coded image.

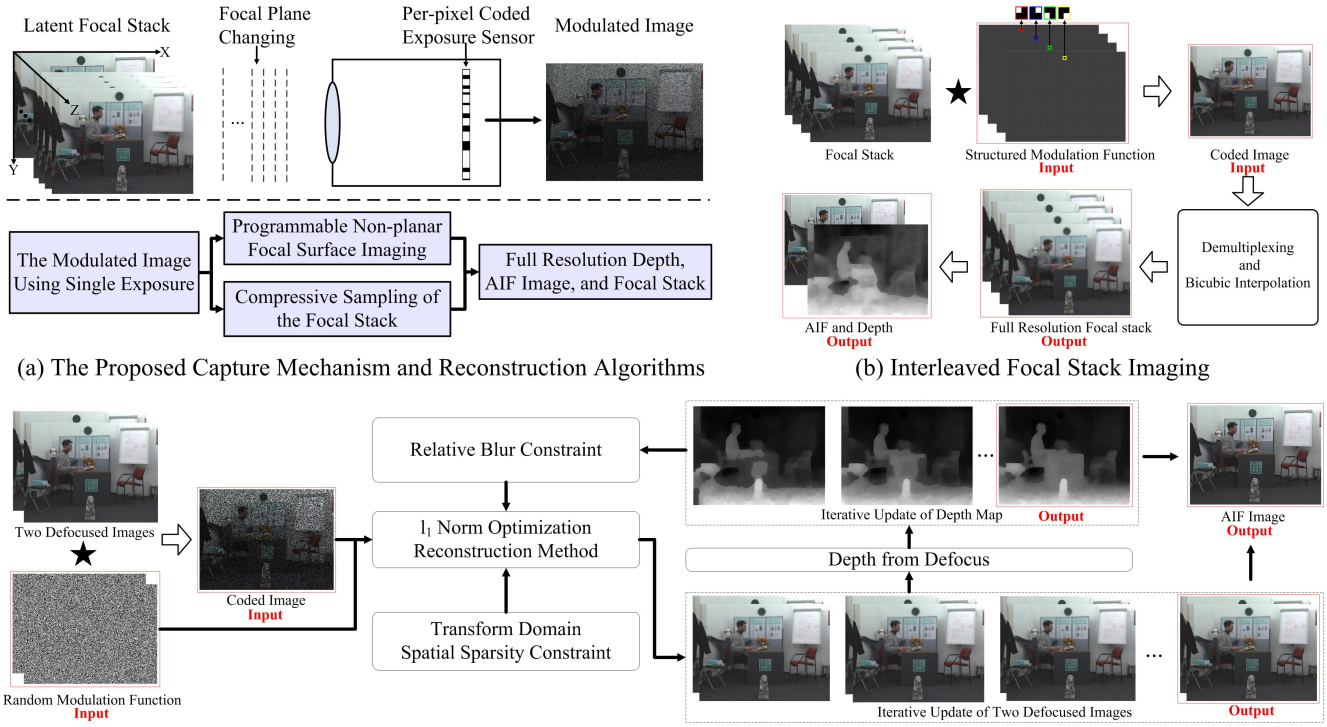
Light Field Imaging. Light field imaging has been actively investigated for decades [2]. Ng et al. [22], for instance, built a hand-held plenoptic camera by attaching microlenses to the sensor. Levoy et al. [17] extended that system to microscopic imaging. Alternatives to microlens-based cameras include mask-based architectures [25] and coded apertures [18]. All of these approaches either reduce image resolution to multiplex the 4D light field onto the 2D sensor or require multiple images. The main application of light fields is flexible depth of field control. We demonstrate that one application of coded focal stack photography achieves high-resolution, single-shot image refocus, as well as the direct acquisition of programmable unconventional focal surfaces.

Image and Depth from a Single Exposure. A variety of computational photography methods have been proposed to obtain depth and AIF image with single exposure. Usually, the point spread function of an optical system is modified to become depth-independent. This can be achieved using focal sweep [14], apertures coded with diffusion [5] or chromatic lens aberration [4]. In a post-processing step, AIF image is computed by inverting the depth-independent blur. Lattice-focal lens [16] and coded apertures [15, 25, 3] have also been explored for the purpose of recovering a depth map with an AIF image. With coded focal stack photography, we combine focal sweep and coded sensor readout for novel applications such as programmable unconventional focal surface imaging, multiplexed focal stack acquisition, and high-resolution AIF image and depth estimation.

Focal Stack Photography. Focal stack is a common tool in medical and scientific imaging, such has microscopy. Usually, the slices of a focal stack are combined to create a single AIF image. Hasinoff and Kutulakos [11] analyzed the optimal settings for focus and aperture to minimize the number of required photographs for a given depth range. The combination of focal stack capture and varying aperture sizes has also been shown to allow for the reconstruction of depth information [10]. In this paper, we explore single-shot coded focal stack approaches that have a variety of different applications.

Per-pixel Coded Exposure. Per-pixel coded exposure photography has been utilized in many applications, including high dynamic range (HDR) imaging, feature detection, object recognition [20], high-speed imaging, HDR imaging, and image deblurring [8], as well as for flexible spatio-temporal resolution tradeoffs [9]. Our application utilizes a coded per-pixel readout or modulation throughout the exposure time of a single photograph.

Compressive Computational Photography. Exploiting the intrinsic redundancy of visual information to build next-generation computational imaging systems is an active area of research in computational photography. Recently proposed methods include the reconstruction of short video clips from a single coded image [12, 24, 23, 13] using compressive sensing paradigms. Usually, these include randomly coded image acquisition followed by sparse reconstruction. In one of our applications, we show that the pixels within a focal stack are highly redundant and propose a novel method to acquire a single coded projection from which the high-resolution focal stack is reconstructed using sparse coding.



(a) The Proposed Capture Mechanism and Reconstruction Algorithms

(b) Interleaved Focal Stack Imaging

(c) Compressive Sampling of Focal Stack

Figure 2: Overview of coded focal stack photography. We explore optical camera configurations that combine focal sweep and per-pixel coded sensor readout (top left). We show a variety of applications, including programmable non-planar focal surfaces, interleaved focal stack acquisition (top right), and compressive focal stack photography for high-resolution image and depth reconstruction (bottom).

3. Coded Focal Stack Acquisition and Reconstruction using Single Exposure

Let $F(\mathbf{y}, z)$ denote the 3D focal stack with $\mathbf{y} = \{y_1, y_2\}$ being the 2D spatial coordinates and z being the depth dimension of the focal stack. A coded sensor image $I(\mathbf{y})$ is measured by sweeping a lens over the focal stack depth within a single exposure time and modulating each pixel with the attenuation code $M(\mathbf{y}, z)$. The recorded sensor image is then a coded projection of the focal stack:

$$I(\mathbf{y}) = \int_{\Omega_z} F(\mathbf{y}, z) M(\mathbf{y}, z) dz, \tag{1}$$

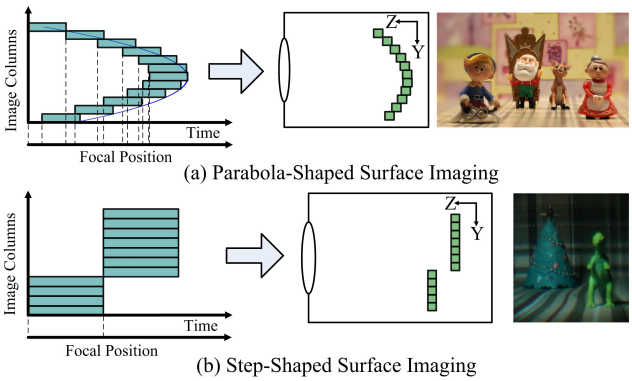
where Ω_z is the range of z. In the following sections, various specific modulation functions are presented to achieve different effects, such as programmable non-planar focal surfaces, and recovery of the focal stack from the single exposure coded image with known modulation function and camera parameters.

3.1. Programmable Non-planar Focal Surfaces Imaging

Achieving programmable non-planar focal surfaces imaging is an intuitive solution for flexible control of the focal surface. Actually, our spatio-depth modulation can provide an equivalent implementation, in which a virtual sensor shape is determined by the modulation function. Eq. (1) gives that the intensity at position $\mathbf{y}$ in the coded image is determined by the pixels of a focal stack at the same position with the modulation function being 1. Formally, the focal shape is represented as $\varphi(\mathbf{y}) = z$, and we define the modulation function here as

$$M(\mathbf{y}, z) = \delta(z - \varphi(\mathbf{y})). \tag{2}$$

Differently shaped non-planar focal surfaces can be obtained by designing different modulation functions according to Eq. 2. Fig. 3 shows two examples of non-planar focal surface results: Fig. 3(a) shows the result of a parabola-shaped surface with $\varphi(\mathbf{y}) = y_2^2$, all of the three toys are in focus here. Fig. 3(b) is the step-shaped surface imaging result captured by our prototype system with $\varphi(\mathbf{y}) = aH(y_2)$, where a is a constant coefficient and $H(\cdot)$ is a step function, notice that the dinosaur and Christmas tree are both in focus. We also compare conventional imaging with planar a focal surface and sinusoidal focal surface $\varphi(\mathbf{y}) = \sin(y_2)$ (Fig. 1, top row); notice how the effect of complex focal surfaces can be achieved in single exposure.



(a) Parabola-Shaped Surface Imaging

(b) Step-Shaped Surface Imaging

Figure 3: Programmable non-planar focal surfaces imaging. Different focal surfaces can be achieved by designing different modulation functions.

3.2. Interleaved Focal Stack Imaging

With interleaved focal stack imaging, a low resolution focal stack $F(\mathbf{y}_z, z), \mathbf{y}_z \subset \mathbf{y}$ can be obtained by de-modulating the sensor image $I(\mathbf{y})$: $F(\mathbf{y}_z, z) = I(\mathbf{y}_z) \cdot M(\mathbf{y}_z, z)$, with $M(\mathbf{y}, z)$ being the modulation function implementing interleave-shaped focal surface. Specifically, $M(\mathbf{y}, z)$ is defined as

$$M(\mathbf{y}, z) = \delta(z - h(\mathbf{y})), \tag{3}$$

where $h(\mathbf{y}) = \sum_m g(\mathbf{y}) \otimes \delta(\mathbf{y} - m \cdot T), m \in \mathbf{Z}$ is the periodic function with the cycle being T, the domain of $g(\mathbf{y})$ is defined in $[0, T]$, and $\otimes$ denotes the convolution operator.

Assuming a discretized latent band-limited focal stack with n depth layers($z = 1, ..., n, T = n$), then $g(\mathbf{y})$ can be defined as

$$g(\mathbf{y}) = \sum_{i=1}^{n} a_i \cdot rect(\mathbf{y} - i), \tag{4}$$

where $rect$ is the rectangle function, and

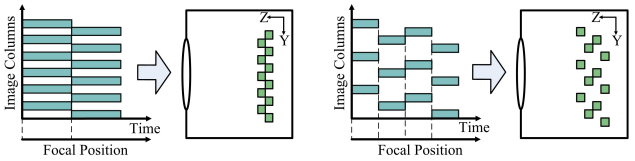
$$\{a_1, ...a_i...a_j..., a_n\} = \{1, ..., n\}, a_i \neq a_j, n \in \Delta\mathbf{y}, \tag{5}$$

with $\Delta\mathbf{y}$ being a neighbourhood of $\mathbf{y}$ representing the local sensor region contains all the layers we intend to recover.

Fig. 4(a) and (b) illustrate interleaved focal stack imaging at $n = 2$ and $n = 4$, and the de-modulation method is illustrated in Fig. 2(b). In order to compensate for the reduced spatial resolution, we adopt bicubic interpolation for each defocused image to upsample to the full-resolution in the reconstruction.

3.3. Compressive Focal Stack Sampling

This section uses the random modulation function as the sensing matrix to get the single coded focal stack and then recover the full-resolution focal stack by using its intrinsic sparsity. Fig. 5 shows the sparsity essence of the focal stack.



(a) Two Layers Interleaved-Shape Surface

(b) Four Layers Interleaved-Shape Surface

Figure 4: Interleaved focal stack imaging. We use the interleaved non-planar focal surface to capture focal stack in single exposure, different numbers of layers with different interleaved-shapes can be achieved with different interleaved modulation functions.

The random modulation function

$$M(\mathbf{y}, z) = rand(\mathbf{y}, z), \tag{6}$$

is used for encoding the focal stack. Specifically, we use a binary random modulation function $rand(\mathbf{y}, z) \in \{0, 1\}$ in our paper. Then the estimated focal stack $\tilde{F}(\mathbf{y}, z)$ can be recovered by optimizing following objective

$$\tilde{F}(\mathbf{y}, z) = \arg\min_{F(\mathbf{y},z)} (E_d(F(\mathbf{y}, z)) + \alpha E_m^s(F(\mathbf{y}, z))). \tag{7}$$

Here the data term is derived from Eq. (1)

$$E_d(F(\mathbf{y}, z)) = || \int_{\Omega_z} M(\mathbf{y}, z) F(\mathbf{y}, z) dz - I(\mathbf{y}) ||_2^2, \tag{8}$$

with α being the weighting factor, $\Omega_\mathbf{z}$ is the range of $\mathbf{z}$. For the regularization term we enforce sparsity of the focal stack in the transform domain by minimizing the l_1 norm of coefficients

$$E_m^s(F(\mathbf{y}, z)) = ||\Psi F(\mathbf{y}, z)||_1, \tag{9}$$

where Ψ is the sparsity transform basis. We use the discrete cosine basis (DCT). As shown in Fig. 5, focal stack representation coefficients are very sparse in the DCT basis. Other bases, such as Wavelets, can be used alternatively.

The performance of focal stack recovery from compressively coded measurement is shown in Fig. 6(a)(using the "Art" example of Fig. 7). The performance degenerates as the number of defocused images increases, so in the next section we show that in order to further exploit the relative blur constraint in the focal stack with depth and AIF image estimation, we adopt only two defocused images for compressive modulation.

4. Depth and AIF Image from a Coded Image

Depth and AIF recovery from a focal stack has been proposed in the past [7]. In Sections 4.1–4.3 we briefly review this formulation and show in Section 4.4 how it can be incorporated into a novel iterative feedback loop for compressive focal stack sampling.

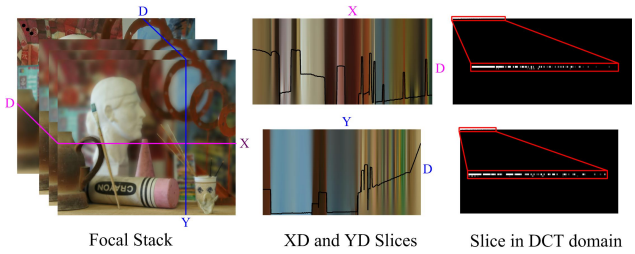


Figure 5: Focal stacks are redundant. A synthetic focal stack is shown on the left, with a horizontal and a vertical slice illustrated in the center. The image information along the depth-dimension of the focal stack barely changes, which results in sparsely populated coefficients in the discrete cosine transform (DCT) domain (right).

4.1. Relative Blur Constraint in the Focal Stack

Let $R(\mathbf{y})$ denote the AIF image, the defocused image $F(\mathbf{y})$ focusing on a certain depth can be represented as [7]:

$$F(\mathbf{y}) = \int_{\mathbf{x}\in N(\mathbf{y})} h_\sigma(\mathbf{y},\mathbf{x})R(\mathbf{x})d\mathbf{x}. \tag{10}$$

Here $N(\mathbf{y})$ is the neighborhood of $\mathbf{y}$ and the blur kernel $h_\sigma(\mathbf{y},\mathbf{x})$ can be approximated by Gaussian convolution model, and $\sigma(\mathbf{y}) = \kappa b(\mathbf{y})$ is the amount of depth-related blurring, with the calibration parameter κ and the blur radius $b(\mathbf{y}) = Dv/2 \cdot |1/F - 1/v - 1/s(\mathbf{y})|$. Here D denotes the aperture diameter, s denotes the depth map, v denote focus setting and the focal length is F.

Given two registered defocused images $F(\mathbf{y}, z_1)$, $F(\mathbf{y}, z_2)$, where z_i denotes the distance from the object to the lens, with focal setting v_{z_1}, v_{z_2} and keeping the other camera parameters consistent, the relative blur based convolution model [6] is defined as:

$$\begin{cases} F(\mathbf{y}, z_2) \approx F_{\Delta\sigma}(\mathbf{y}, z_1) \\ \qquad = \int_{\Omega_\mathbf{x}} h_{\Delta\sigma}(\mathbf{y},\mathbf{x})F(\mathbf{x}, z_1)d\mathbf{x}, \sigma^2_{z_2}(\mathbf{y}) > \sigma^2_{z_1}(\mathbf{y}), \\ F(\mathbf{y}, z_1) \approx F_{\Delta\sigma}(\mathbf{y}, z_2) \\ \qquad = \int_{\Omega_\mathbf{x}} h_{\Delta\sigma}(\mathbf{y},\mathbf{x})F(\mathbf{x}, z_2)d\mathbf{x}, \sigma^2_{z_2}(\mathbf{y}) < \sigma^2_{z_1}(\mathbf{y}), \end{cases} \tag{11}$$

where $\Delta\sigma(\mathbf{y}) = \sqrt{\sigma^2_{z_2}(\mathbf{y}) - \sigma^2_{z_1}(\mathbf{y})}$ is the depth-related relative blurring.

So, the relative blur operator Φ between $F(\mathbf{y}, z_1)$ and $F(\mathbf{y}, z_2)$ can be derived by using Eqs. 11 when depth and camera parameters are provided, and the relative blur constraint for optimization in Eq. 7 can be formulated as

$$E^d_m(F(\mathbf{y})) = ||\Phi_z F(\mathbf{y}, z)||^2_2. \tag{12}$$

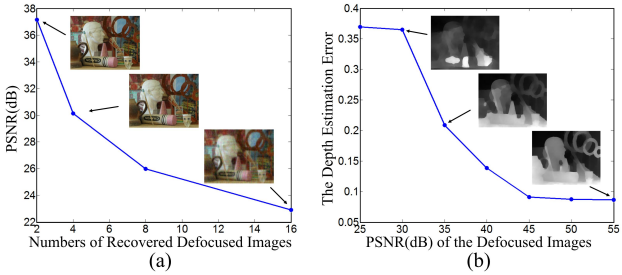


Figure 6: The performance of compressive focal stack imaging without relative blur constraint and the depth estimation algorithm. In order to ensure the good initial depth estimation, we modulate two defocused images to obtain initial defocused images with high PSNR.

4.2. Depth Estimation from the Relative Blur

Depth can be estimated from two defocused images $F(\mathbf{y}, z_1)$ and $F(\mathbf{y}, z_2)$ by optimizing

$$\tilde{s} = \arg\min_s (E_d(s) + \gamma E_m(s)), \tag{13}$$

where $E_d(s)$ and $E_m(s)$ are the data term and regularization term respectively, with γ being a weighting factor. Specifically, the data term can be written as:

$$\begin{aligned} E_d(s) = &\int_{\Omega_\mathbf{y}} H(\Delta\sigma(\mathbf{y})) \cdot |F(\mathbf{y}, z_2) - F_{\Delta\sigma}(\mathbf{y}, z_1)|^2_2 d\mathbf{y} \\ &+ \int_{\Omega_\mathbf{y}} (1 - H(\Delta\sigma(\mathbf{y}))) \cdot |F(\mathbf{y}, z_1) - F_{\Delta\sigma}(\mathbf{y}, z_2)|^2_2 d\mathbf{y}, \end{aligned} \tag{14}$$

where H denotes the Heaviside function, and regularization term $E_m(s)$ is defined as the total variation to favor piecewise smooth scene depth: $E_m(s) = ||\nabla s(\mathbf{y})||_1$.

The performance of depth estimation from relative blur is shown in Fig. 6(b)(using the "Art" example of Fig. 7). One can see that the depth estimation error, measured by the absolute difference [7], increases with the decreasing of PSNR of the defocused images.

4.3. All-in-focus Image Estimation

According to Eq. 10 and initial depth estimation, we can calculate two spatially varying blur kernel maps $h_{\sigma_{z_1}}, h_{\sigma_{z_2}}$ corresponding to defocused images $F(\mathbf{y}, z_1), F(\mathbf{y}, z_2)$ respectively. Then the optimization of AIF image can be formulated as a spatially varying deblurring:

$$\tilde{R} = \arg\min_R (E_d(R) + \lambda E_m(R)), \tag{15}$$

where λ is the weighting factor. The data term is defined as

$$\begin{aligned} E_d(R) = &|| \int_{\mathbf{x}\in N(\mathbf{y})} h_{\sigma_{z_1}}(\mathbf{y},\mathbf{x})R(\mathbf{x})d\mathbf{x} - F(\mathbf{y}, z_1)||^2_2 \\ &+ || \int_{\mathbf{x}\in N(\mathbf{y})} h_{\sigma_{z_2}}(\mathbf{y},\mathbf{x})R(\mathbf{x})d\mathbf{x} - F(\mathbf{y}, z_2)||^2_2, \end{aligned} \tag{16}$$

and for regularization we incorporate the sparse prior of natural images: $E_m(R) = ||\nabla R(\mathbf{y})||_{0.8}$.

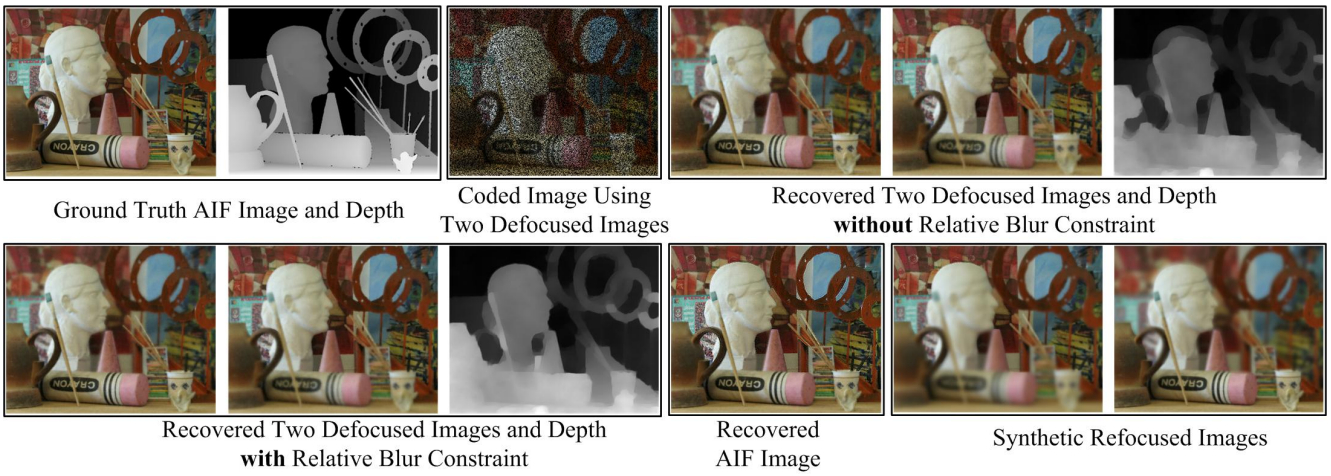


Figure 7: Simulated reconstruction of the "Art" dataset. We synthesize two defocused images from ground truth AIF image and depth, then the coded image can be obtained with defocused images and corresponding two random modulation masks. The recovered two defocused images and depth with and without relative blur constraint are shown above, with which the AIF image is also recovered and the refocused images can be synthesized.

4.4. Focal Stack, Depth and AIF Image Recovery

We iteratively estimate depth and defocused images until convergence, then the AIF image can be obtained by optimization in Eq. 15.

For interleaved focal stack imaging, we can directly perform depth estimation using the algorithm in Section 4.2 after obtaining the interpolated focal stack. For compressive sampling of the focal stack, because of the difficulty in demultiplexing compressively modulated data, we encode only two defocused images $B(\mathbf{y}, z_1)$ and $B(\mathbf{y}, z_2)$ at different focal settings for high PSNR and also propose a feedback scheme to increase robustness.

Firstly, to recover the two defocused images, the demodulation algorithm in Section 3.3 utilizes only spatial sparsity, so we further improve the performance by using relative blur constraint in Section 4.1 (illustrated in Fig. 2(c)). Then, we obtain the initial depth from Eq. 13 using the two defocused images and then feedback the depth constraint regularization into optimization Eq. 7 to get a better demodulation:

$$\tilde{F}(\mathbf{y}, z) = \arg \min_{F(\mathbf{y},z)} (E_d(F) + \alpha E_m^s(F) + \beta E_m^d(F)), \tag{17}$$

where α and β are weighting factors, $z = \{z_1, z_2\}$, and with three terms defined in Eqs. 8, 9 and 12 respectively.

In data capturing, there is a scaling between the defocused images when we change the focal plane of the camera, we approximate it with affine transformation [7]: $\mathbf{y}_2 = \mathbf{a}\mathbf{y}_1 + \mathbf{b}$. For calibration, we keep a small aperture and use four markers in the scene to calibrate the parameters $\mathbf{a}, \mathbf{b}$. Thus, after we obtain the initial two defocused images by using the optimization Eq. 7, we apply registration to them before depth estimation. Correspondingly, we add an affine transformation operator $\mathbf{T}$ to Eq. 17 by changing relative blur constraint $E_m^d(B(\mathbf{y}))$ into

$$E_m^d(F(\mathbf{y})) = ||\Phi_z \mathbf{T}_z F(\mathbf{y}, z)||_2^2. \tag{18}$$

We use the Alternating Direction Method(ADM) numerical solution [27] to solve both depth estimation and focal stack recovery. For depth estimation, good initial defocused images have been provided and numerical solution is robust to noise. For defocused images recovery, depth errors usually occur in the textless areas and spatial sparsity constraint in transform domain helps correction. Therefore, the iterative tends to converge in spite lacking theoretical proof. And for the AIF image estimation, we extend the Iterative Re-weighted Least Squares process (IRLS) [15] to deal with spatially varying convolution and incorporate information from multiple input images.

5. Experimental Results

5.1. Synthetic Data

This experiment tests the effectiveness of the proposed demultiplexing algorithm from a compressively modulated focal stack by utilizing the spatial and relative blur redundancy. We use the "Art" data from Middlebury [1] with its spatial resolution down sampled to 695×555 pixels, as shown in two top left subfigures of Fig. 7. Empirically, we assume the depth range to be 2m ∼5m, the focal length be 50mm, the $f/\#$ be 8 and the calibration parameter $\kappa = 3e4$.

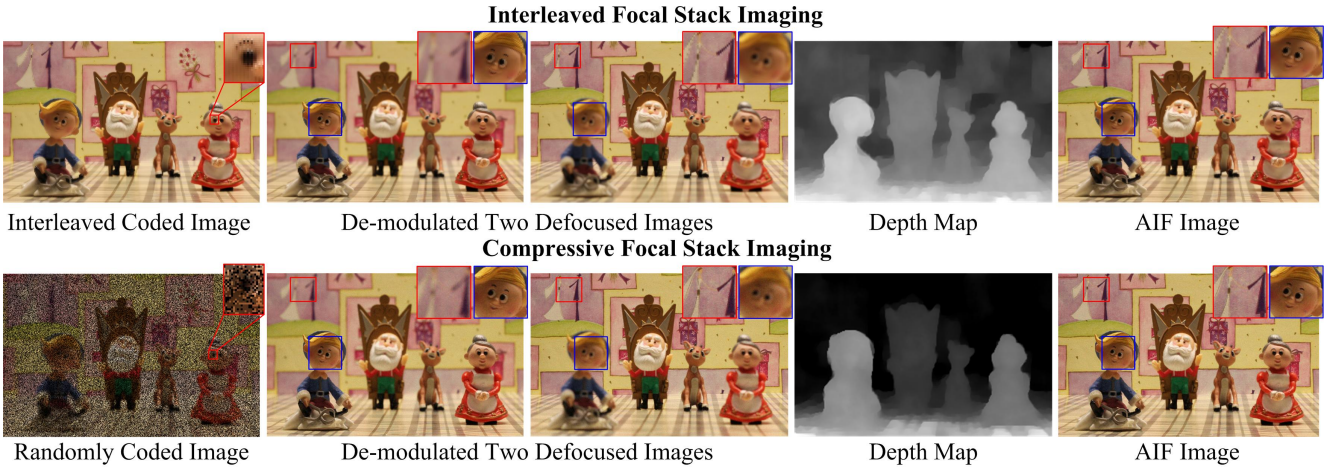


Figure 8: Experiment on synthetic data. The top row displays the result on interleaved focal stack imaging, while the bottom row gives the results on compressive sampling of the focal stack. The close-up images show that compressive focal stack imaging can provide higher spatial resolution with more accurate depth and AIF image estimation compared to the interleaved focal stack imaging.

We synthesize two defocused images focusing at 2m and 5m according to Eq. 10, and then use two random masks to modulate them within a single image, as shown in the 3rd image of the top row in Fig. 7. Next, we use the method proposed in Section 3.2 to recover the two defocused images, whose average PSNR is 37.17dB and the depth map, as shown the right-most three images in top row of Fig. 7. In order to get better defocused images, we introduce relative blur constraint in de-modulation algorithm and the iterative optimization framework (the maximum iteration number is set to be 5). With an un-optimized MATLAB implementation on a PC with an Intel(R) 3.4GHz Core(TM) i7-2600K CPU, 8.0GB, RAM 64-bit operating system, the total processing time is approximately 25 minutes. The reconstructed defocused images and depth map are largely improved, as shown in the left-most three images in the bottom row of Fig. 7, the average PSNR of the defocused images is 41.53dB. The large improvement can mainly attributed to the relative blur constraint. We also display the recovered AIF, with PSNR being 33.26dB.

Comparing with the ground truth depth map, as shown in the second image of top row in Fig. 7, one can see that we obtain a comparable depth map and this further validate the proposed coded focal photography. In the last two images in Fig. 7, we give a synthetic defocused images using the recovered AIF image and depth map.

The above experiment results in the Fig. 7 show that the accuracy of both AIF image and depth map improve greatly after iteration. And compared to the defocused images demodulated directly from the coded measurement with higher compression ratio, using the synthetic version generated from estimated depth and AIF image can result in higher PSNR.

5.2. Captured Data with Synthetic Modulation

In this section we capture two images at different focal settings using our Canon EOS REBEL T3 and generate coded images according to the two above mentioned modulation functions. The top row in Fig. 8 displays the modulated measurement of two-layer interleaved imaging, the demodulated defocused images, estimated depth map and AIF image. Similarly, the bottom row in Fig. 8 shows the result of compressive focal stack sampling. The experiments in Fig. 8 validate the effectiveness of the two proposed reconstruction algorithms. The PSNR of AIF images are 29.01dB and 32.21dB for interleaved focal stack imaging and compressive focal stack imaging respectively; and visually, the depth recovered from interleaved focal stack imaging has more artifacts than the compressive method because of the misalignment caused by lower spatial resolution, while has lower computational complexity.

5.3. Prototype Results

Fig. 10 shows the prototype we implemented. We hack the Canon Lens EF 50mm/1.8 as a programmable moving lens which is controlled by ARDUINO to change the focal plane of the system, and use a Canon EOS REBEL T3 camera body with a resolution of 4272 $\times$2848 pixels. For modulation, we take the EBY701-NP/C/T Touch Screen VGA LED Monitor apart and insert the LCD between two orthogonal polarizers. During capture, we use 209$\times$219 pixels out of the native resolution 800$\times$400 pixels. The camera focuses on both the LCD and the scene by using an optical relay system so that the LCD acts as a per-pixel modulator, with the focal lengths of the two relay lenses being 380mm and 900mm respectively. We use a telephoto lens with focal

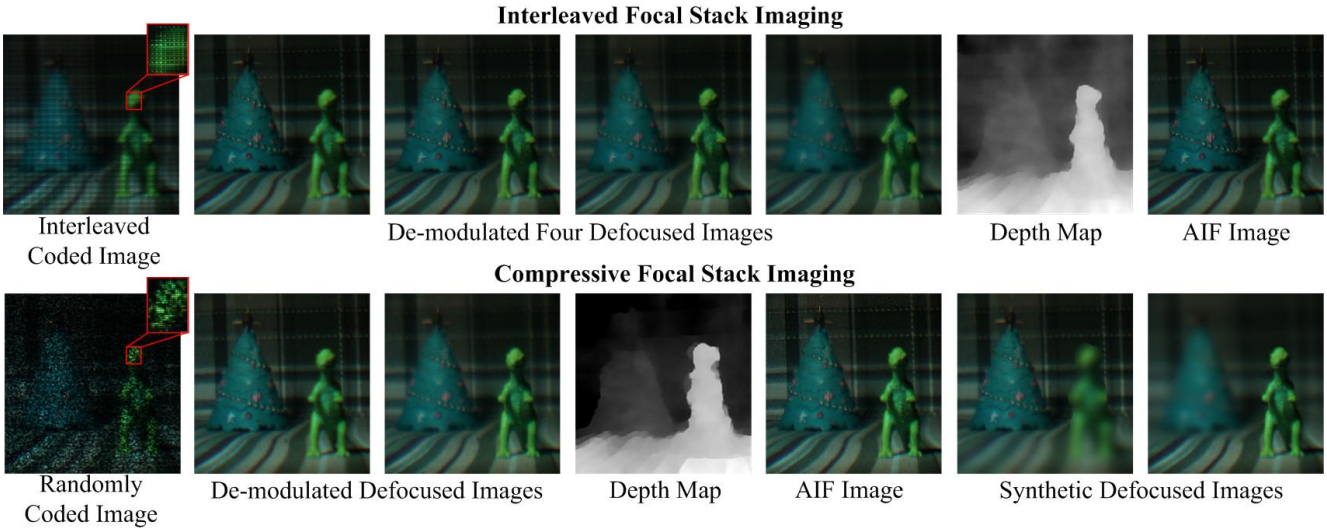


Figure 9: Experiment on data captured by our prototype. The top row shows results of interleaved focal stack imaging, while the bottom row shows results of compressive focal stack sampling.

length 300mm to utilize as many camera pixels as possible. The moving lens, LCD and exposure are synchronized with software.

We implement both interleaved focal stack imaging and compressive focal stack sampling using the prototype. For the experiments in Fig. 9, it takes 0.7 seconds to capture the interleaved coded image that modulates four defocused images with our prototype, as shown in the top left of Fig. 9. The top row of Fig. 9 displays the results of interleaved coded imaging, including the coded image of images focused at four settings, the four de-modulated defocused images after bicubic interpolation and estimated depth map and AIF image. The bottom row in Fig. 9 shows the results on compressive sampling of a focal stack, with a similar arrangement of interleaved imaging. In addition, we also show two results of flexible depth of field control, refocused at 0.9m, 1.2m respectively.

The experiments with our prototype system show the promising results for programmable non-planar focal surfaces imaging, single exposure focal stack, as well as depth and AIF image recovery.

6. Discussion

Whereas our prototype camera requires relay optics and an LCD modulator, most consumer products already include varifocal lenses and a partially controllable sensor readout, such as rolling shutters. Emerging computational cameras such as [26] can fully control the sensor readout and [28] can change the focal position at very high-speed. In the near future, it will be feasible to build portable cameras enhanced with coded focal stack photography.

By combining optically-coded focal stack acquisitions

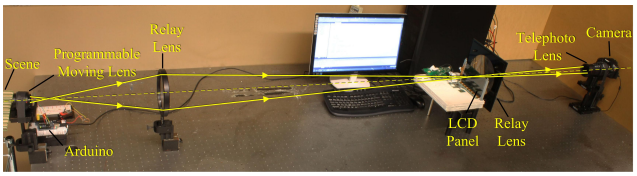


Figure 10: Our prototype and its light path. We use the moving lens and LCD panel for focal stack modulation, and the relay lens and telephoto lens are used to match the field-of-views.

with sparse coding techniques, we robustly recover high-resolution depth maps and AIF images. However, we also propose an adaptive, all-optical approach to capturing photographs with programmable unconventional focal surfaces. Coded focal stack photography, as explored in this paper, provides a significant improvement—at least for today's hardware—in readout time compared to successively capturing defocused images within a focal stack (e.g. [28]). Our optical designs pave the road for a new class of computational photography techniques.

Currently, we are mostly limited by the fact that we do not have access to the per-pixel sensor readout. Our prototype camera uses relay optics combined with a transmissive spatial light modulator to optically simulate a controlled per-pixel readout. Unfortunately, the employed LCD has a low contrast, limited resolution, and a pixel-precise alignment with the sensor is also challenging. We also rely on the lens motor to refocus, which limits the proposed method to static scenes. Liquid lenses or other high-speed refocusable optics may eliminate this limitation. Furthermore, our reconstructions assume that the photographed scenes are mostly static during the exposure time, thus we do not consider motion blur.

7. Conclusion and Future Work

In summary, we introduce coded focal stack photography and demonstrate various applications such as programmable unconventional focal surfaces, interleaved focal stack acquisition, and compressive focal stack imaging for recovering full-resolution AIF images and depths.

In the future, we would like to extend our computational camera design to dynamic scenes that include object and camera motion. We believe that the key challenge for this is the exploration of advanced sparse focal stack representations that take their dynamic nature into account. Our programmable non-planar focal surface design has the potential to correct the distortion of the employed lenses. Obtaining programmable non-planar focal surface projectors and programmable non-planar sensor surface cameras without moving parts are the other two interesting avenues of future work. Finally, we would like to combine more sophisticated focal sweeping motions with new kinds of optical codes.

8. Acknowledgement

We thank the reviewers for their feedback. We thank Matthew O'Toole, Yosuke Bando and Matthew Hirsch for insightful discussion. Gordon Wetzstein was supported by the DARPA SCENICC program and an NSERC PDF. Ramesh Raskar was supported by an Alfred P. Sloan Research Fellowship and a DARPA Young Faculty Award. Tsinghua University affliated coauthors were supported by the China National Basic Research Project (No. 2010CB731800) and the Key Project of NSFC (No. 61120106003, 61035002, 61171119 and U0935001).

References

[1] http://vision.middlebury.edu/stereo/data/.

[2] E. H. Adelson and J. Y. A. Wang. Single lens stereo with plenoptic camera. In *IEEE Trans. PAMI*, 14(2):99–106, 1992.

[3] Y. Bando, B.-Y. Chen, and T. Nishita. Extracting depth and matte using a color-filtered aperture. In *ACM SIGGRAPH Asia*, 2008.

[4] O. Cossairt and S. Nayar. Spectral focal sweep: Extended depth of field from chromatic aberrations. In *ICCP*, 2010.

[5] O. Cossairt, C. Zhou, and S. Nayar. Diffusion coded photography for extended depth of field. In *ACM SIGGRAPH*, 2010.

[6] P. Favaro. Recovering thin structures via nonlocal-means regularization with application to depth from defocus. In *CVPR*, 2010.

[7] P. Favaro and S. Soatto. *3D shape reconstruction and image restoration: exploiting defocus and motion blur*. Springer Verlag, 2006.

[8] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar. Coded rolling shutter photography: Flexible space-time sampling. In *ICCP*, 2010.

[9] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan. Flexible voxels for motion-aware videography. In *ECCV*, 2010.

[10] S. W. Hasinoff and K. N. Kutulakos. Confocal stereo. In *IJCV*, 81(1):82–104, 2009.

[11] S. W. Hasinoff and K. N. Kutulakos. Light-efficient photography. In *IEEE Trans. PAMI*, 33(11):2203–2214, 2011.

[12] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar. Video from a single coded exposure photograph using a learned over-complete dictionary. In *ICCV*, pages 287–294, 2011.

[13] J. Holloway, A. C. Sankaranarayanan, A. Veeraraghavan, and S. Tambe. Flutter shutter video camera for compressive sensing of videos. In *ICCP*, 2012.

[14] S. Kuthirummal, H. Nagahara, C. Zhou, and S. K. Nayar. Flexible depth of field photography. In *IEEE Trans. PAMI*, 33(1):58–71, 2011.

[15] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *ACM SIGGRAPH*, 2007.

[16] A. Levin, S. W. Hasinoff, P. Green, F. Durand, and W. T. Freeman. 4D frequency analysis of computational cameras for depth of field extension. In *ACM SIGGRAPH*, 2009.

[17] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz. Light field microscopy. In *ACM SIGGRAPH*, 2006.

[18] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen. Programmable aperture photography: Multiplexed light field acquisition. In *ACM SIGGRAPH*, 2008.

[19] V. P. Namboodiri and S. Chaudhuri. Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera. In *CVPR*, 2008.

[20] S. K. Nayar, V. Branzoi, and T. E. Boult. Programmable imaging: Towards a flexible camera. In *IJCV*, 70:7–22, 2006.

[21] S. K. Nayar and Y. Nakagawa. Shape from focus. In *IEEE Trans. PAMI*, 16(8):824–831, 1994.

[22] R. Ng. Fourier slice photography. In *ACM SIGGRAPH*, 2005.

[23] D. Reddy, A. Veeraraghavan, and R. Chellappa. P2C2: Programmable pixel compressive camera for high speed imaging. In *CVPR*, pages 329–336, 2011.

[24] A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk. Cs-muvi: Video compressive sensing for spatial-multiplexing cameras. In *ICCP*, 2012.

[25] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *ACM SIGGRAPH*, 2007.

[26] G. Wan, M. Horowitz, and M. Levoy. Applications of multi-bucket sensors to computational photography. *Stanford Computer Graphics Laboratory Technical Report*, 2012.

[27] J. Yang and Y. Zhang. Alternating direction algorithms for L1-problems in compressive sensing. In *SIAM Journal on Scientific Computing*, 33:250–278, 2011.

[28] C. Zhou, D. Miau, and S. K. Nayar. Focal sweep camera for space-time refocusing. *Technical Report, Department of Computer Science, Columbia University*, 2012.