

Compressive Epsilon Photography for Post-Capture Control in Digital Imaging

Atsushi Ito¹, Salil Tambe², Kaushik Mitra², Aswin C. Sankaranarayanan³, and Ashok Veeraraghavan²

¹SONY Corporation

²Rice University

³Carnegie Mellon University

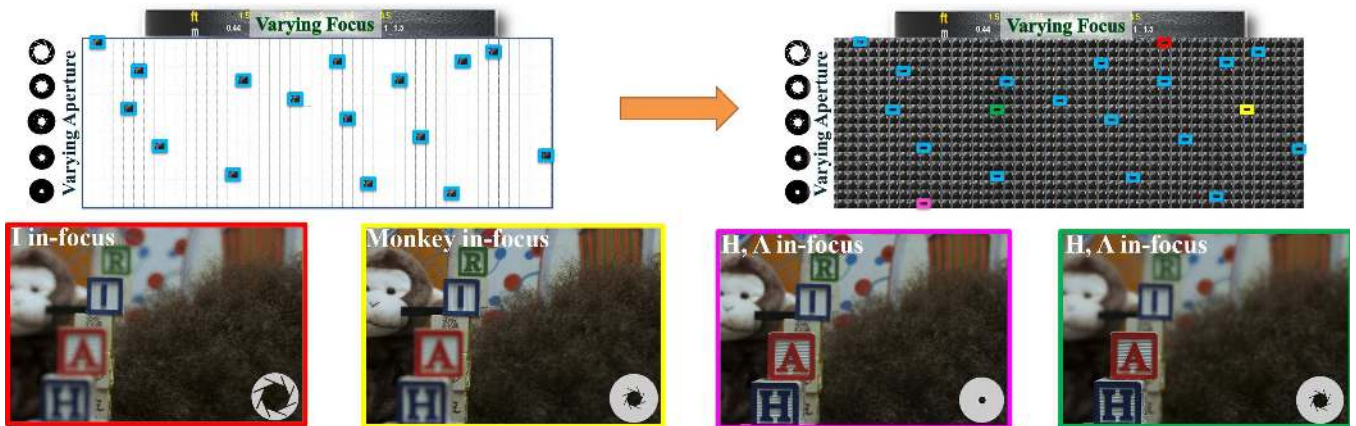


Figure 1: Compressive epsilon photography enables post-capture freedom from a few carefully selected photographs of a scene. *Lurking among the gamut of photographs that can be obtained by varying camera parameters are a few photographs that can truly capture the aesthetics of a scene. In this paper, we enable complete post-capture freedom by estimating the entire collection of photographs corresponding to varying camera parameters from a few select photographs at pre-determined camera parameters (marked in blue).*

Abstract

A traditional camera requires the photographer to select the many parameters at capture time. While advances in light field photography have enabled post-capture control of focus and perspective, they suffer from several limitations including lower spatial resolution, need for hardware modifications, and restrictive choice of aperture and focus setting. In this paper, we propose “compressive epsilon photography,” a technique for achieving complete post-capture control of focus and aperture in a traditional camera by acquiring a carefully selected set of 8 to 16 images and computationally reconstructing images corresponding to all other focus-aperture settings. We make the following contributions: first, we learn the statistical redundancies in focal-aperture stacks using a Gaussian Mixture Model; second, we derive a greedy sampling strategy for selecting the best focus-aperture settings; and third, we develop an algorithm for reconstructing the entire focal-aperture stack from a few captured images. As a consequence, only a burst of images with carefully selected camera settings are acquired. Post-capture, the user can then select any focal-aperture setting of choice and the corresponding image can be rendered using our algorithm. We show extensive results on several real data sets.

CR Categories: I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—3D/stereo scene analysis I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—

Sampling I.4.10 [Image Processing and Computer Vision]: Image Representation—Multidimensional, Statistical

Keywords: Post-capture processing, Refocusing, Focus-aperture stacks, Compressive sensing, Gaussian mixture models

Links: [DL](#) [PDF](#) [WEB](#) [VIDEO](#)

1 Introduction

The transition from film to digital was largely about convenience. While there has been remarkable technological breakthroughs in optical flexibility and computational capabilities of digital cameras, photography still mimics film in some unfortunate ways: the photographer is still required to set all camera parameters such as focus, aperture, exposure, and ISO at capture-time and has limited flexibility in changing these settings post-capture. While professional photographers have mastered the art of making the correct choices during capture-time, the need to get all the camera parameters correct in the heat of the moment impedes casual photographers from acquiring breathtaking photographs. Given the enormous strides in resolution, low-light performance, user interface and physical size that have been made in digital cameras, pre-capture settings remain stubbornly as one of the last frontiers in digital photography severely limiting the ability of casual photographers. *Thus, there is an immediate need for methods that enable near-complete post-capture control in digital photography.*

Digital photography has slowly started making a move in this direction with the advent of consumer light-field cameras [Gortler et al. 1996; Levoy and Hanrahan 1996]. Light-field cameras such as the Lytro [Lytro] and the Raytrix [Raytrix] use a micro-lens array to capture multiple viewpoints of the scene simultaneously. This allows for post-capture control of focus albeit at the loss of spatial resolution. Green et al. [Green et al. 2007] proposed a method for multi-aperture photography that allows for near-complete post-capture control of aperture size by exploiting a special modification to the lens that captures the rays from concentric disks around the aperture in 4 separate fields on the image plane. In these tech-

niques, careful hardware modifications to an existing camera with the use of appropriate reconstruction algorithms allow for limited post-capture control. Hardware modifications to existing digital cameras pose a practical challenge and as a consequence there is a need for techniques that can be applied on current digital cameras.

Epsilon photography [Raskar 2009] is an alternative technique where a large set of images acquired with varying parameter settings on the camera allows for limited post-capture flexibility. Several examples of such techniques have been popular including focal stack photography [Kutulakos and Hasinoff 2009], high dynamic range (HDR) photography [Debevec and Malik 1997], confocal stereo [Hasinoff and Kutulakos 2006], focal sweep [Hasinoff and Kutulakos 2006], multi-image panorama stitching [Brown and Lowe 2007], and lucky imaging [Law et al. 2005; Joshi and Cohen 2010]. In almost all these techniques, the goal is to capture multiple images with varying parameter settings on the camera and then fuse these images together to generate a composite image, or to make some meaningful inference (e.g., depth) about the scene. While such techniques are becoming popular, they require a large number of images to be acquired.

For one dimensional parameter variations such as HDR imaging and focus stack photography, epsilon photography is becoming the de-facto method especially for static or slow-moving scenes. Nevertheless, when two-dimensional (or more) parameter variations are needed, epsilon photography becomes exceedingly cumbersome, needing the capture of several thousand images. Consider confocal stereo as an illustrative example; capturing 61 focus settings and 13 aperture settings as in [Hasinoff and Kutulakos 2006] requires the capture of 793 images requiring several hours of capture time and necessarily eliminating any hope of such techniques extending to account for any scene motion.

Key idea. *The central theme of this paper is to study, characterize, model and exploit the redundancy in the intensity variations at a pixel when the camera parameters change, and develop a technique for recovering the entire epsilon photography stack from a small subset of observed images with varying parameter settings.*

Let us motivate the intuition for this redundancy by looking at intensity variations as a single camera parameter is varied. When only the ISO is varied, the observed intensity changes linearly until saturation (barring sensor noise). When the aperture is varied while keeping focus and ISO constant (exposure duration is changed to keep light level constant), the observed intensity changes smoothly due to spatial blur variations. Similarly, when the focus setting is varied while keeping ISO, aperture, and exposure constant, the observed intensity changes smoothly due to the smooth changes in depth-dependent blur. In both these cases, the blur-induced intensity variations are low frequency phenomena and can be modeled as having low-dimensional support. Thus, the intensity variations observed at an individual pixel are inherently low-dimensional and can be accurately modeled using appropriate signal processing techniques, thereby allowing for their reconstruction from just a few observed images (with carefully selected camera settings that span the entire low-dimensional subspace).

Contributions: The key technical contributions of this paper are

- We carefully study and characterize the intensity variations observed at a pixel as camera parameters vary, and show that this ‘per-pixel intensity profile’ is inherently low-dimensional.
- We show that approximating the per-pixel intensity profile using a Gaussian Mixture Model captures 99.6% of the energy resulting in a very accurate approximation.
- We propose ‘compressive epsilon photography’, where a few images acquired with carefully selected camera settings enables

complete recovery of the per-pixel intensity profile resulting in post-capture flexibility and control hitherto unachievable.

- We propose a greedy algorithm to select camera settings for maximize reconstruction performance.
- We show several examples of compressive epsilon photography ranging including one-dimensional focal stacks and two-dimensional focal-aperture stacks (confocal stereo).

Limitations. A key limitation is that we do not account for motion during the acquisition (or motion blur) and hence, the results presented here are only applicable for static or slow-moving scenes. We show that in practice 8 – 16 images are sufficient for most scenes. By using the burst mode capability of cameras, these 8 – 16 images can be captured within a short duration; this significantly expands the operational regime of conventional epsilon photography. Per-pixel modeling, the basis of the proposed approach, assumes that the same scene point is observed at the same pixel when the camera settings are varied. When this is not the case, the calibration procedure described in Section 5 should be used to pre-process the data before applying the reconstruction algorithms. Finally, the training data used for learning the model parameters needs to be rich and diverse enough in order to capture all the natural variations observed in the test datasets. In particular, scenarios such as sharp specularities, spectral separation due to diffraction, lens flares and glares, etc. may not lead to accurate reconstructions due to limited occurrence in training data. Per-pixel modeling also fails when a very bright light source is adjacent to a pixel; in such cases, the contribution of the high intensity point to the blur may be underestimated resulting in intensity errors. Finally, the algorithm is only trained to generate the natural blur of traditional cameras and if the rendering task is to simulate a specific structured bokeh, then additional processing may be necessary.

2 Prior work

2.1 Epsilon photography

While image-based rendering and inference techniques that rely on multiple images acquired by varying camera parameters such as exposure, field-of-view, ISO, aperture, viewpoint and focus have been in use for over a decade, recently they have been collectively referred to as “epsilon photography” [Raskar 2009] since they rely on acquiring a set of images with incremental (or epsilon) changes to the camera settings.

Multi-image denoising/deblurring. Multiple images can be used to improve image denoising performance [Buades et al. 2009], and image deblurring [Sroubek and Milanfar 2012; Yuan et al. 2007].

HDR imaging: One widely used technique for HDR is exposure bracketing where a series of images are captured with exponentially increasing exposure durations [Mann and Picard 1994; Debevec and Malik 1997]. In addition to varying exposure, varying the ISO parameters can further enhance performance of HDR imaging in noisy and low light areas [Hasinoff et al. 2010]. There have also been some attempts to acquire HDR images by varying the aperture in a camera [Hasinoff and Kutulakos 2007], but in such cases the varying depth of field in the images needs to be taken into account.

Focal stack photography. The collection of images obtained by sweeping the focus position through the scene is called focal stack [Agarwala et al. 2004]. This enables depth estimation as well as acquisition of an extended depth of field (EDOF) image [Kuthirummal et al. 2011]. In microscopy, this enables 3D reconstruction and compensates for the shallow depth of field due to the large numerical aperture of the microscope [McNally et al. 1999; Sibarita 2005]. In both these cases, the number of images acquired increases linearly with the aperture size and the volume being imaged, since every scene point needs to be sharp in at least one of the acquired

images. Compressive epsilon photography learns the redundancies in intensity variations caused due to varying focus and therefore can reconstruct images with intermediate focus settings accurately, reducing the number of images required for focal stack methods.

Aperture stack photography. Hasinoff and Kutulakos [Hasinoff and Kutulakos 2007] showed that acquiring images with varying aperture sizes enables interesting and useful rendering effects, including depth of field manipulation and HDR acquisition. Our compressive reconstruction algorithm enable a significant reduction in the number of aperture images that need to be acquired by effectively interpolating images with unseen apertures.

Confocal stereo. Most techniques for epsilon photography have been restricted to varying one camera parameter. In confocal stereo [Hasinoff and Kutulakos 2009], both the aperture size and the focus position are simultaneously varied and a large collection of images is acquired. Analyzing the intensity of a single pixel, as both aperture and focus settings vary, shows the property of “confocal constancy”, i.e., the normalized intensity (to account for aperture area) does not change with aperture size when that particular pixel is in focus. This results in an algorithm for estimating per-pixel depth without resorting to any neighborhood based methods. All previous techniques such as depth from focus, depth-from-defocus and stereo require neighborhood smoothness assumptions [Krotkov 1988; Grossmann 1987] and therefore cannot handle complex scenes with thin single-pixel subjects such as hair. Confocal stereo was the first method to be able to produce independent, per-pixel estimates of depth. Unfortunately, this method requires an inordinately large number of input images; for example for many scenes in this paper, traditional confocal stereo would require $61 \times 13 = 793$ images. This makes the technique impractical except for static scenes. Exploiting the redundancy in per-pixel intensity profiles enables us to reconstruct the entire confocal data from a subset of images (typically 8 – 16); thus making confocal stereo methods practical even for non-static scenes.

Light-field [Gortler et al. 1996; Levoy and Hanrahan 1996] has become a popular and useful representation for image-based rendering. Light-fields offers immense flexibility in post-capture control including focus, aperture, and perspective changes. However, the direct methods of capturing light-fields [Levoy and Hanrahan 1996; Lytro ; Raytrix] either trade off spatial resolution or require significant hardware changes/resources [Boominathan et al. 2014] for the same sensor resolution and provide much limited aperture/focus control as compared to our framework. Several techniques have also been developed to reconstruct light-fields from focal or aperture stacks [Levin and Durand 2010; Green et al. 2007]; as we show in our comparisons, we outperform all of these significantly in providing a finer control over depth-of-field (aperture) and focus.

2.2 Compressive sensing

Compressive sensing enables reconstruction of signals from undersampled data by exploiting the redundancy in the signal structure [Baraniuk 2007; Candès and Wakin 2008]. This is done by exploiting signal sparsity and utilizing algorithms that minimize the ℓ_1 -norm of the signal. Such methods have resulted in several compressive imaging systems for capturing light-fields [Marwah et al. 2013; Tambe et al. 2013], light-transport matrices [Peers et al. 2009], videos [Park and Wakin 2009] and hyper-spectral data [Wagadarikar et al. 2008]. This paper is similar in spirit to compressive sensing; in that, we reconstruct the entire epsilon photography stack from just a few carefully selected images.

Gaussian Mixture Models. Recently, it has been shown that Gaussian Mixture Models (GMMs) are an effective prior for compressive sensing applications [Baron et al. 2010; Yang et al. ; Bourrier et al. 2013]. Following this, methods for compressive sensing of images

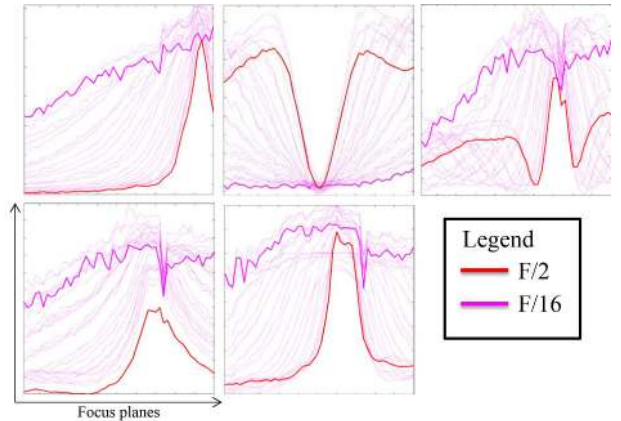


Figure 2: Dependence on focus stack intensity profiles for varying aperture. Intensity variations observed at a few pixels for the “animals” scene in Figure 3. Each figure shows how intensity varies at a pixel with focus plane variations for different aperture setting. Note how the intensity profiles gracefully degenerate from the larger-variation profile observed at large apertures (F/2) to the fairly shallow profile at small apertures (F/16).

[Yu et al. 2012], videos [Yang et al. 2013], and light-fields [Mitra and Veeraraghavan 2012] have been proposed using GMMs. GMM has also been used to study compressive computational imaging systems [Mitra et al. 2014]. In this paper, we utilize a GMM-based compressive sensing framework to reconstruct the complete epsilon photography stack from undersampled set of images. We also utilize the GMM model to greedily select a set of camera parameters to acquire image with.

3 Redundancy in epsilon photography

Epsilon photography refers to the space of images that can be captured using a camera by varying exposure time, aperture size, focus plane, and ISO. While the space of such images is large, there are key redundancies that we can identify which we eventually use to obtain compressive epsilon photography. In this section, we look at different slices of the epsilon photography stack and infer geometric properties for each. Specifically, we look at the *focus stack* — the space of images obtained by varying only the focus position — and *focus-aperture stack* — the space of images obtained by varying both the focus and aperture.

3.1 Per-pixel intensity profiles

Our goal is photographic-quality rendering of the space of images that can be captured by a camera. Inspired by work in confocal stereo [Hasinoff and Kutulakos 2009], we adopt to model the per-pixel intensity variations observed when the camera parameters are varied. The key promise of per-pixel modeling is that the spatial resolution of the photographs is easily preserved and no smoothness assumptions need to be made, resulting in the ability to extract sharp information even on single-pixel edges such as hair.

Let $I_{u,v}(f, a, i, s)$ be the intensity profiles observed at a pixel (u, v) for varying focus setting f , aperture size a , ISO values i , and exposure times s . For static scenes, the parameters governing exposure becomes largely irrelevant except for governing overall light-levels. To this end, we fix the exposure setting such that the overall light-level is held constant. This implies that $s = Ca^2$, where C is an arbitrary constant. For the rest of the paper, we consider only variations in focus and/or aperture and hence, only consider $I_{u,v}(f, a) = I_{u,v}(f, a, i, Ca^2)$ implicitly fixing exposure time and holding ISO constant at some pre-determined value to achieve constant light-levels. Note that, if we fix the aperture a , $I_{u,v}(f)$ purely

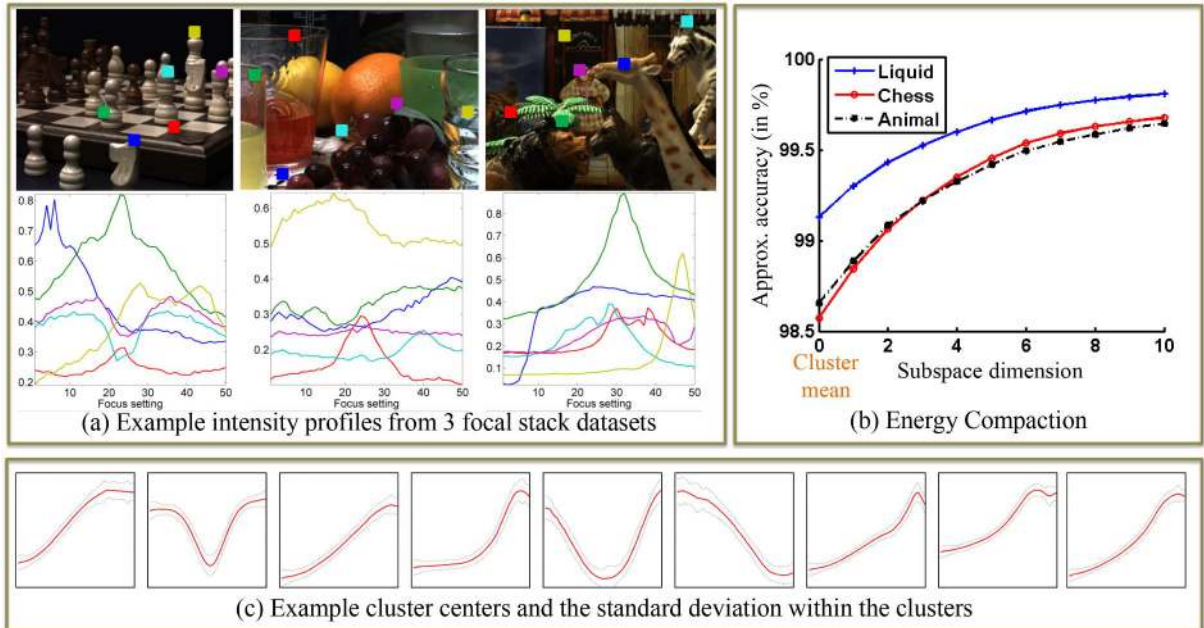


Figure 3: Redundancies in focus stack datasets. (a) Images from three different focus stack datasets and color-coded intensity variations for varying focus observed at different pixels. Note how the intensity profiles are largely unimodal with the peak/valley corresponding to the depth of the pixel. (b) We apply k -means clustering on the intensity profiles observed at individual pixels for varying focus settings. Plot of energy compaction with 100 clusters shows that more than 98.5% of the energy is captured by the 100 cluster centers. A 10 dimensional subspace around each cluster center allows us to capture more than 99.5% of the energy. (c) Shown are the intensity profiles corresponding to the top-9 cluster means (in red) and the standard deviation around the mean (in black, dotted). The ensuing set of intensity profiles form tight clusters that can be used to build strong statistical models. These statistical models enable compressive epsilon photography.

models intensity variations with varying focus setting — referred to as the focus stack [Agarwala et al. 2004][Kutulakos and Hasinoff 2009]. Similarly, $I_{u,v}(f, a)$ captures intensity variations with aperture and focus — also referred to as aperture-focus image (AFI), a key construct used in many prior papers including the seminal work of Hasinoff and Kutulakos [Hasinoff and Kutulakos 2009]. We study both these cases, individually, next.

3.2 Focus stack intensity profiles

Figure 3 shows intensity variations for varying focus settings as observed in three different datasets captured with a Canon SLR using a 50mm lens. Since we are interested solely in focus variations, we obtained images at ISO100 and aperture F/1.4; the aperture chosen corresponds to the largest aperture setting in our camera. Typically, most intensity variations are observed at the largest aperture setting since for smaller apertures produce larger depth of field and smaller defocus blurs (see Figure 2). At this setting, we varied the focus settings electronically to obtain 45 unique focus planes.

From Figure 3 it is clear that a large percentage of the intensity profiles are *unimodal* with a well-defined peak/valley. The geometric explanation to this is that, over the small perspective of the aperture, scene points are nearly Lambertian and hence, there is a specific focus setting where the point is in focus (an intensity peak or valley is reached at this focus setting). For other focus setting, defocus blur averages this point with intensities from the neighborhood of the pixel. As seen in the figure, while there are subtle variations, the predominant trend is a gradual shift of the intensity from the pixel’s own intensity value when it is in focus to the average intensity value of its neighbors when defocus blur is very large.

3.3 Focus-aperture stack intensity profiles

Figure 4 shows per-pixel AFIs of size 45×18 corresponding to 45 focus settings and 18 F-stops. Similar to the focus stack in-

tensity profiles, they exhibit predictable structures. First, when a scene point is in focus, changing the aperture does not change its intensity — hence, we obtain a vertical equi-intensity line at the correct focus. This property is the *confocal* property used in [Hasinoff and Kutulakos 2009]. Second, at the smallest aperture setting, the depth-of-field is very large and hence, there is little variation in intensities as we vary the focus settings. This leads to an approximately equi-intensity horizontal line. Finally, the specifics of the neighborhood around a point leads to various equi-intensity profiles linking the intensity profile at the largest aperture and the smallest aperture. These are harder to predict since they depend on local neighborhood structures; however, they are smooth variations which suggests associated redundancies.

3.4 Clustering and predictability

For both focus-stack and aperture-focus stack datasets, we performed k -means clustering on the per-pixel intensity variations after normalizing the mean and standard deviation of each profile. Figure 3 shows the top 9 recovered mean intensity profiles and the standard deviation within each cluster. We recovered 100 clusters — however, the top 50 clusters accounted for over 80% of the total intensity profiles. We observe distinct cluster means that have well-defined peaks/valleys and tight clustering around each mean profile. The small deviation around the mean profiles indicate that intensity variations around the peak — once suitably normalized — are largely predictable once we associate a cluster label to it.

The intensity variations in the focal stack are due to blur which is a low frequency phenomenon. This suggests that low-dimensional models may be sufficient to capture these variations. Figure 3 shows the energy compaction achieved by modeling these variations using 100 clusters and a PCA basis around each cluster. The 100 cluster centers alone capture more than 98.5% of the energy and even a 10 dimensional PCA basis account nearly 99.5% of the

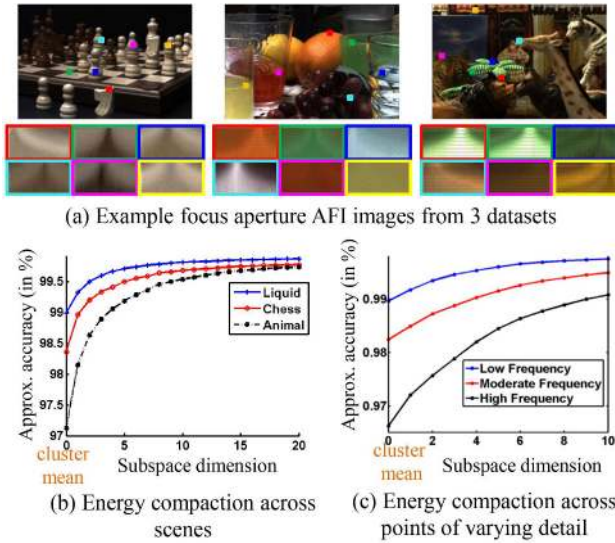


Figure 4: Redundancies in focus-aperture datasets. (a) Shown are images from three datasets and the AFI images observed at different pixels. (b) We apply k -means clustering on the AFI images. Plot of energy compaction with 450 clusters shows that more than 97% of the energy is captured by the 450 cluster centers. A 10 dimensional subspace around each cluster center allows us to capture more than 99% of the energy. (c) We look at energy compaction plot, but now for AFI images separated into three types corresponding to varying levels of high-frequency textures at patches surrounding the point. Again, we see very high redundancies even at textured regions.

energy indicating that these profiles are highly ‘compressible’.

Similar conclusions can be drawn from clustering per-pixel AFIs. Figure 4 also shows the energy compaction that can be achieved by modeling these intensity variations using 450 cluster centers and low dimensional PCA basis around these cluster centers. The 450 cluster centers alone capture more than 97% of the energy and even a 10 dimensional PCA basis around these clusters account for more than 99% of the energy. The energy compaction results largely remain the same when we look at AFI images clustered according to level of texture. Specifically, for each point, we look at the occurrence of strong gradients on the smallest aperture image (equivalently, the all-in focus image) and separate the pixels into three groups according to this measure. Even for highly-textured regions where we can expect our per-pixel modeling to fail, we obtain a compaction of 96% with just the cluster center and 99% with a 10-dimensional subspace approximation about the cluster center. In all, this indicates that the AFIs are very ‘compressible’.

Both of these clustering results are a direct consequence of the unimodality which provides a certain level of *alignment* across intensity profiles. It is well known that any form of registration/alignment enhances linear correlations that can be captured well using PCA. This observation suggests that focus stack intensity variations observed at pixels are highly redundant — once we associate a specific focus plane to a scene point, we can expect to predict its entire intensity profile using just a few samples. *This forms the basis for compressive epsilon photography.*

4 Compressive epsilon photography

Our goal is to perform per-pixel reconstructions of intensity variations as we change camera parameters; the promise of such methods is in their ability to preserve the true resolution of the input

images. For this purpose, we learn per-pixel priors for different tasks such as focal stack and focus-aperture stack. For focus stack, we learn a prior for $I_{u,v}(f)$ and for focus-aperture stack we learn a prior for $I_{u,v}(f, a)$ or AFIs. From Figures 3 and 4, it is clear that the intensity profiles are tightly clustered around the cluster centers. A natural way for modeling such signals is via Gaussian Mixture Models (GMMs). Note that we could also model such signals using dictionaries [Rubinstein et al. 2010]; however, we choose GMM because it is analytically tractable [Mitra and Veeraraghavan 2012] and can be used for selecting camera parameters.

4.1 GMM prior

We collected 5 focal stack datasets, extracted per-pixel intensity profiles, and then used these as training data to learn the GMM priors. Similarly, for AFI in focus-aperture stack, we collect 9 datasets, with varying scene complexity, extract per-pixel AFI and then learn GMM priors from them. To learn the GMM prior, we use an EM algorithm to iterate through learning cluster membership and learning mean and covariance for each cluster. At the end of the EM algorithm, we obtain all the GMM parameters which include the number of clusters K , cluster weights p_k (fraction of data that belongs to a cluster), cluster means $u^{(k)}$, and covariances $C^{(k)}$. The prior model is given as

$$f(x) = \sum_{k=1}^K p_k \mathcal{N}(x; u^{(k)}, C^{(k)}).$$

4.2 Reconstruction using GMM prior

During reconstruction, our goal is to estimate the per-pixel intensity profile (in case of focal stack) or AFI (in case of focus-aperture stack) from observed sub-samples. The relation between the observed sub-samples and the complete profile is of-course linear and can be expressed as $y = Hx + n$, where y represents the observed sub-sampled in vectorized form, x represents the complete intensity profile or AFI in vectorized form, H is the transformation matrix, and n is observation noise. We assume noise to have Gaussian distribution $\mathcal{N}(0, C_n)$ with zero mean and covariance C_n .

For reconstruction we use the Minimum Mean Square Error (MMSE) estimator. Given an observation y , the posterior distribution $f(x|y)$ of x is again a GMM with new cluster weights $\alpha^{(k)}(y)$ and new (posterior) cluster Gaussian distributions $f^{(k)}(x|y)$:

$$f(x|y) = \sum_{k=1}^K \alpha^{(k)}(y) f^{(k)}(x|y),$$

where $f^{(k)}(x|y)$ is the posterior distribution of the k^{th} Gaussian

$$f^{(k)}(x|y) = \mathcal{N}(x; u_{x|y}^{(k)}(y), C_{x|y}^{(k)})$$

with mean $u_{x|y}^{(k)}(y)$ and covariance $C_{x|y}^{(k)}$, given by

$$\begin{aligned} u_{x|y}^{(k)}(y) &= u^{(k)} + C^{(k)} H^T (H C^{(k)} H^T + C_n)^{-1} (y - H u^{(k)}), \\ C_{x|y}^{(k)} &= C^{(k)} - C^{(k)} H^T (H C^{(k)} H^T + C_n)^{-1} H C^{(k)}. \end{aligned} \quad (1)$$

The new weights, $\alpha^{(k)}(y)$, given as

$$\alpha^{(k)}(y) = \frac{p_k f^{(k)}(y)}{\sum_{i=1}^K p_i f^{(i)}(y)},$$

are the old weights p_k modified by $f^{(k)}(y)$, the probability of y belonging to the k^{th} mixture component

$$f^{(k)}(y) = \mathcal{N}(y; H u^{(k)}, H C^{(k)} H^T + C_n).$$

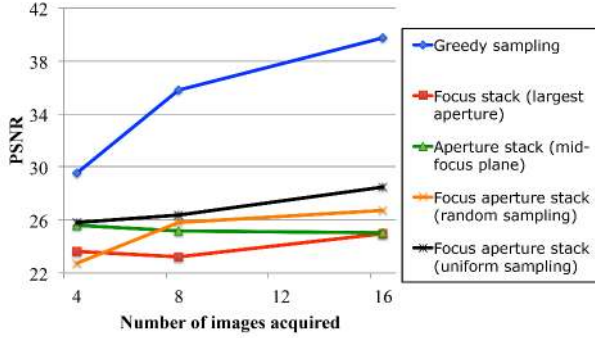


Figure 5: Greedy image selection. Comparison of PSNR as a function of the number of acquired images for various methods of sampling clearly shows the efficacy of our greedy sampling scheme. Note that the proposed scheme provides almost 10dB improvement over traditional sampling methods.

The MMSE estimator $\hat{x}(y)$ is the mean of the posterior $f(x|y)$, i.e.,

$$\hat{x}(y) = \sum_{k=1}^K \alpha^{(k)}(y) u_{x|y}^{(k)}(y).$$

The corresponding MMSE is given by

$$\text{MMSE}(H) = E\|x - \hat{x}(y)\|^2 \quad (2)$$

4.3 Greedy algorithm for image selection

Our goal is to reconstruct the focal and focus-aperture stacks by observing a minimal number of images corresponding to certain choices of camera focus and aperture settings. Thus, choice of camera parameters to observe is an important consideration¹. We propose a greedy algorithm based on minimizing the MMSE (2). A tight analytic lower bound for the MMSE has been derived by [Flam et al. 2012][Flam et al. 2011] and we exploit this lower bound to derive a greedy sampling strategy. The lower bound of MMSE is given by:

$$\text{MMSE}(H) = \sum_{k=1}^K p_k \text{Tr} \left(C_{x|y}^{(k)} \right), \quad (3)$$

where $C_{x|y}^{(k)}$ is the posterior cluster Gaussian covariance (1).

Finding optimal camera parameters is a combinatorial problem since we need to choose parameters from a pre-defined set of focus-aperture values. We instead rely on a greedy strategy that selects one camera setting at a time that best minimizes the MMSE given previously selected camera parameters. Suppose that there are a total of N_f focus settings and a total of N_a aperture settings in a given camera. If we are interested in capturing only m images corresponding to m camera settings, then the brute-force version of the algorithm will require evaluating the MMSE (3) for $\binom{N_f N_a}{m}$ times, which becomes practically impossible. Hence we devise a greedy algorithm. We first find the optimal pair of camera parameters, i.e., $m = 2$ by evaluating the MMSE $\binom{N_f N_a}{2}$ times. Given this pair, we then update the posterior covariance matrices $C_{x|y}$ to take into account the effect of the current selected camera parameters. Each choice of camera parameter correspond to a row h_i in the H matrix. After the i^{th} iteration, the posterior covariance is updated as follows:

$$C_{x|y,i}^{(k)} = C_{x|y,i-1}^{(k)} - C_{x|y,i-1}^{(k)} \hat{h}_i^T (\hat{h}_i C_{x|y,i-1}^{(k)} \hat{h}_i^T + C_n)^{-1} \hat{h}_i C_{x|y,i-1}^{(k)}$$

¹For focal stack, uniform sub-sampling of the focus axis is a good choice as this corresponds to uniform sampling in scene depth. But for focus-aperture stack, the choice of optimal camera parameters is not so obvious.

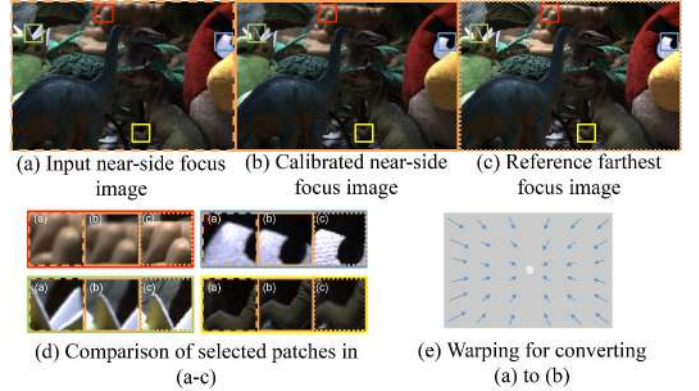


Figure 6: Geometric calibration for precise alignment. An input near-side focus image (a) is calibrated (b) so that the corresponding objects appear at the same locations in images captured with a different focus-aperture setting. Calibration is performed using a farthest focus image (c) as reference. The figures in (d) clearly show that enlarged patches appearing at distinct depths are accurately aligned against the reference image after calibration. In (e), the trajectories of warping directions are shown which suggest that the images corresponding to near-side focus (a) are magnified in comparison with to far-side images (c) as can be seen clearly.

with the initial posterior covariances $C_{x|y,0}^{(k)}$ being the same as the prior GMM covariances, $C_x^{(k)}$. After this covariance update step, we find the next camera setting by evaluating the MMSE expression with updated covariances. Note that from the second iteration, we need to evaluate the MMSE expression just $N_f N_a$ times, which provides a significant reduction in computations. Figure 5 shows how the MMSE decreases as a function of the number of chosen camera parameters m . We can conclude that by capturing 8 images with this prescribed camera settings, we should be able to reconstruct the focus-aperture stack with $N_f = 45$ and $N_a = 20$.

Figure 5 shows a comparison of PSNR as a function of the number of acquired images for (a) proposed sampling strategy, (b) focus stack with large aperture (c) aperture stack with focus position at mid point, (d) random sampling of focus-aperture pairs and (e) uniform sampling of focus-aperture pairs. Note that the proposed scheme provides almost 10 dB improvement over all traditional sampling methods. This result indicates, that even when the goal is not complete post-capture control, but rather traditional focal or aperture stacking (say for depth estimation), our optimized sampling strategy is significantly better.

5 Geometric and photometric calibration

Since we employ a per-pixel based model for learning and reconstruction, precise geometric and photometric calibration with sub-pixel level accuracy is essential. For this purpose, we adopt the procedure in [Hasinoff and Kutulakos 2009].

Geometric Calibration. It is a well known fact that changes in focal settings result in a non-linear warp of the objects in the scene. In [Hasinoff and Kutulakos 2009], it was shown that this warp can be accurately modeled by considering parameters for image magnification, lens distortion and translation. For estimating these parameters, we collected images of a calibration chart containing black dots on a grid at the largest aperture (F/1.4). Registration among images in the dataset was realized by unwarping the images according to the estimated parameters. Figure 6 shows an example of geometric calibration achieving precise alignment.

Photometric Calibration. Modifying the aperture causes a change not only in depth of field but also results in vignetting. This vi-

gnetting is corrected for by collecting and normalizing against a set of reference white images.

6 Experimental results

To validate our approach, we collected 5 test datasets and 6 training datasets for focus-aperture stack reconstruction. For focus-stack reconstruction, we use a subset of the samples from the 5 focus-aperture stack test datasets. All datasets were captured using Canon EOS-40D camera with a Canon EF 50mm fixed focal length lens. The cameras were controlled from a computer. We use a combination of 18 aperture and 45 focus settings for capturing the focus-aperture stack. The aperture was varied from F/2.2 to F/16 while the 45 focus settings covered a depth range of 0.45m to 1m.

Raw-images were captured at a resolution of 1988×1296 . However, we down-sample the images to a lower resolution of 600×400 since the geometric calibration process is slow and is directly proportional to the number of pixels being processed. However, given the per-pixel processing algorithms used, the quality of our results are independent of the actual resolution of the images — although processing times do scale linearly with number of pixels. In all experiments described in the paper, the training and testing datasets were completely different and there was no overlap between the training and testing datasets.

6.1 Focus stack reconstructions

We reconstruct the entire focus stack from only a few input images. All focus stacks were collected with the largest aperture of $F/1.4$. Intuitively, the stack corresponding to the larger apertures exhibit greater variability as compared to the smaller apertures that have large depth-of-fields. Given a few input images — corresponding to a few select focus settings, we reconstruct per-pixel intensity profiles for the entire focus range using our GMM algorithm.

A key test for our algorithms is to verify whether scene points that are never obtained in sharp focus can be focused. Figures 7 and 8 show multiple examples of this; in both cases, we reconstruct a focal stack of 45 images from just 8 input images. We obtain reconstruction SNRs of 25.8 dB for the “Animals” dataset (Figure 7) and 32 dB for the “Liquid” dataset (Figure 8). In both Figures, we also show reconstructed per-pixel intensity profiles for select points — these match the ground truth profiles accurately. Figure 7 (c) and (d) show reconstructed images at intermediate focus planes to input images. The input images in Figure 7(c) has “leaf 1” is focus in the first image (at focal setting 31) and “leaf 3” in focus in the second image, whereas “leaf 2” is not in focus in either. We can clearly see that the blur in leaf 1 increases linearly, whereas that in leaf 3 decreases linearly. Also leaf 2, which was blurred in both the input images, is sharper in the intermediate images.

Note that our reconstructions can handle a wide range of complex materials quite gracefully including transparent objects (“glass” in Figure 8(c)), specularities (“grape 1” and “grape 2” in Figure 8(d)), sub-surface scatterers (“orange” in Figure 8(c)).

A key application of focus stack photography is its ability to provide high-quality all-focus images. We validate our ability to recover such images from reconstructed focus stacks. As a comparison, we also show results from focal stack reconstructed using cubic spline interpolation. We compute the depth map and all focus image using a commercially available software HeliconSoft. Figure 9 shows that the depth map and all focus image obtained using our reconstructed stack is very close to the depth map and all focus image computed from the ground truth 45 image focal stack. The depth maps obtained using spline interpolated focal stack are relatively poor. Moreover, the all focus image obtained from spline interpolation shows lots of artifacts.

6.2 Focus-aperture stack reconstructions

We reconstructed focus-aperture stacks for 3 datasets using 4, 8, 16 and 32 input images. The reconstruction performance as a function of number of input images is shown in Figure 10. There is not much improvement in the “Chess” dataset as the number of input images is increased from 4 to 32; however for more complex datasets such as the “Animal” dataset or the “Glassball” dataset owing to the increased texture and refractive elements, there is a significant improvement as one moves from 4 to 16 images and a slight improvement as we go to 32 images. Hence, irrespective of the complexity of the scene, we can reliably reconstruct the focus-aperture stack using 16 input images. Part (c) of the same figure shows that we are able to generate AFI images which are very close to ground truth from the reconstructed focus-aperture stack. As a result, the estimated depth map obtained by applying the confocal stereo algorithm on our reconstructed focus-aperture stack is very close to the one obtained by applying the confocal stereo algorithm on the ground truth focus-aperture stack directly.

In Figure 11, we show that we are able to accurately reconstruct patches from the input datasets even for a combination of focus and aperture settings that are far away from the focus-aperture combinations at which the 16 input images were captured. A close-up of results on the “chess” and “glassball” dataset is shown in Figure 12.

In Figure 13, we visualize the reconstructed focus-aperture stacks by looking at two subsets. In Figure 13(a), we look at a focus stack where we keep the aperture fixed at F/2.0 — the largest setting — and vary the focus. Observe the clear transition of focus from the Glassball to Pooh to Tigger to the Angry Bird as we sweep through the focus planes. In Figure 13(b), we look at an aperture stack where we keep the focus fixed and vary the aperture. Observe the increase in depth of field as we increase the aperture from F/2.2 to F/11. In summary, we are able to reliably reconstruct objects placed at unobserved focal planes even for large aperture settings.

Figure 14 shows 24 samples from the reconstructed focus-aperture stack of a tennis racquet kept outdoors. The entire stack containing 810 images was reconstructed from just 16 captured images. All the essential characteristics of a focus-aperture stack such as blur, depth of field and focus are exhibited naturally.

Subjective evaluation. Quantitative comparisons in terms of PSNR are often not the best indicators of visual quality. To further validate our claims, we performed subjective evaluations with the goal of determining if the reconstructions were distinguishable from the original images. We performed a visual perception study using a group of 13 test-subjects. We would show the subjects a pair of images — one original and reconstructed — and note the subjects preference of the higher quality image or if there was no preference at all (see Figure 15(a)). The subjects had no restrictions on the amount of time required to make their choice as well as had the ability to zoom-in over different regions for a closer look. Each subject had to complete 30 evaluations and on an average, the amount of time taken to complete them was 5 minutes. Further, to establish reliable controls, out of the total of $13 \times 30 = 390$ image-pairs instances, we randomly placed 69 control evaluations where both images were the same. Out of these, in 60/69 instances the users marked no preference — thereby lending remarkable significance to our evaluations. Figure 15(b) shows the histogram of answers obtained. In a majority of instances (51%), the subjects had no preference between the two images. In about a third of the instances, the subjects preferred the original image and preferred the reconstructed image in a sixth of the instances. Overall this indicates that in a significant percentage of instances (close to 66%) the subjects either preferred our reconstruction or had no particular preference for the original and even when they had a preference the preference was marginal.

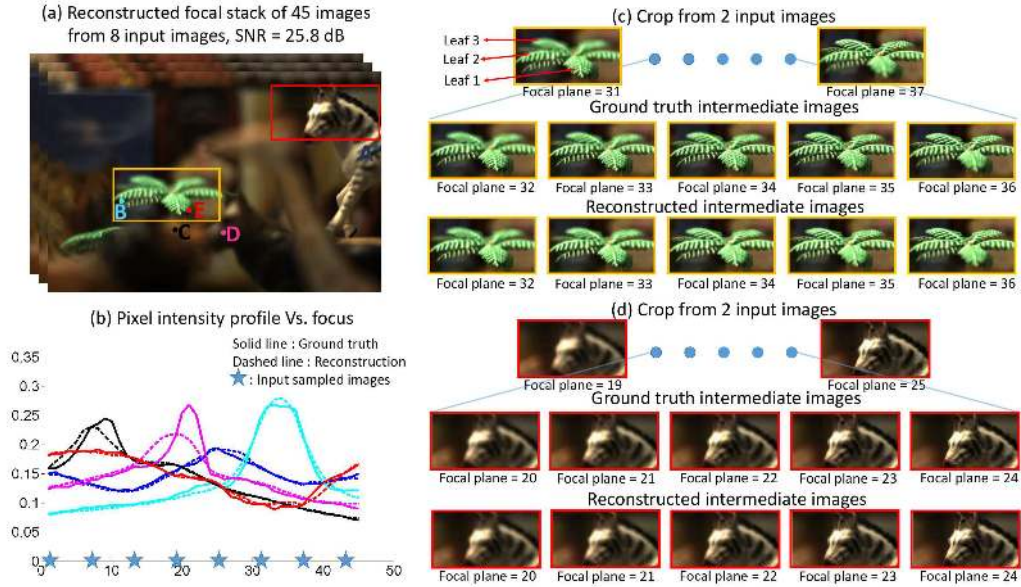


Figure 7: Focal stack reconstruction of the “Animal” dataset. We reconstruct a focal stack of 45 images from just 8 images. (a) An image from the reconstructed focus stack. We obtain a reconstruction SNR of 25.8 dB. (b) Reconstructed intensity profiles for points marked in (a). (c, d) Focus stack reconstructions for the insets in (a). In each case, we show two input images at the top and show recovered intermediate focus planes. The two regions were selected so as to showcase the ability of our algorithm to hypothesize the correct focus plane as well as the bokeh of the camera accurately. (c) “leaf 1” is in focus in the first image and “leaf 3” in focus in the second image (at focal setting 37), whereas “leaf 2” is not in focus in either. In the reconstructed intermediate images, we can clearly see that the blur in “leaf 1” increases linearly, whereas that in “leaf 3” decreases linearly. Also “leaf 2”, which was blurred in both the input images, is sharper in the intermediate images. (d) Blur on “zebra” decreases linearly.

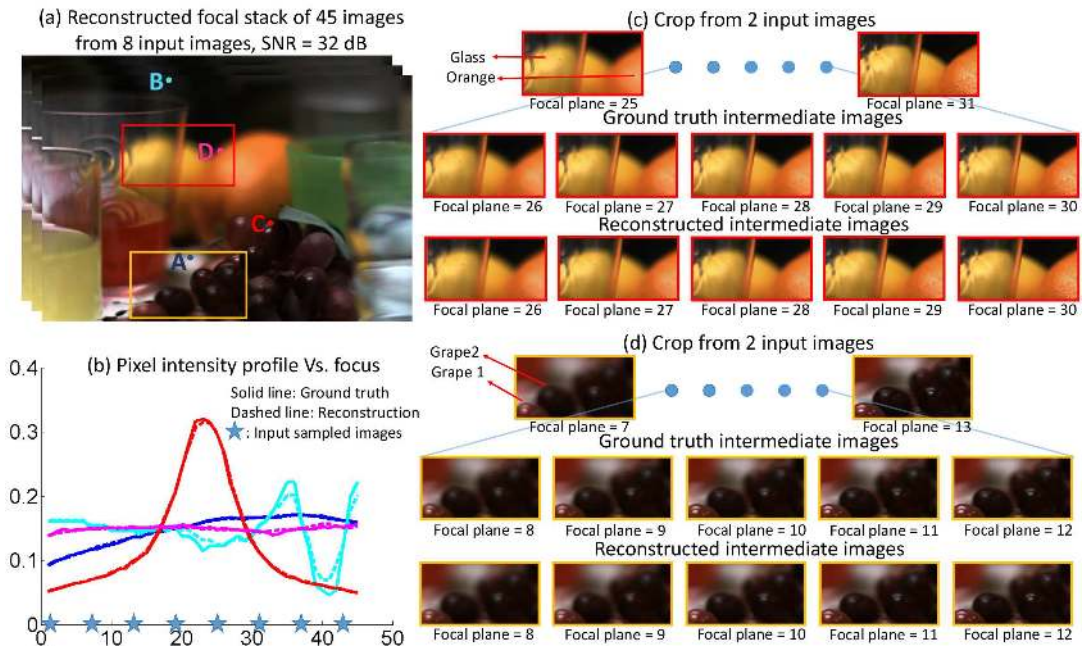


Figure 8: Focal stack reconstruction of “Liquid” dataset. We reconstruct a focal stack of 45 images from 8 images using our algorithm. (a) An image from the reconstructed focal stack. We obtain a reconstruction SNR of 32 dB. (b) Intensity profile (pixel value vs. focus) for various scene points. (c, d) Reconstructed images between selected focal setting. (c) Note that the blur reduces linearly in the orange. Also, note that our algorithm handles transparent object such as glass quite well. (d) We are able to handle specularities in “grape 1” and the blur reduces linearly in “grape 2”.

6.3 Confocal stereo

Confocal stereo is a powerful per-pixel depth estimation algorithm that uses properties of the AFI images observed at a pixel. However, this requires capturing the entire focus-aperture stack that typ-

ically has several hundreds to thousands of images. A key benefit of our compressive epsilon photography framework is that we can obtain this space of images from very few images. Figure 16 shows depth estimates obtained from 16 input images. Our depth estimates are comparable to those obtained by applying the confocal

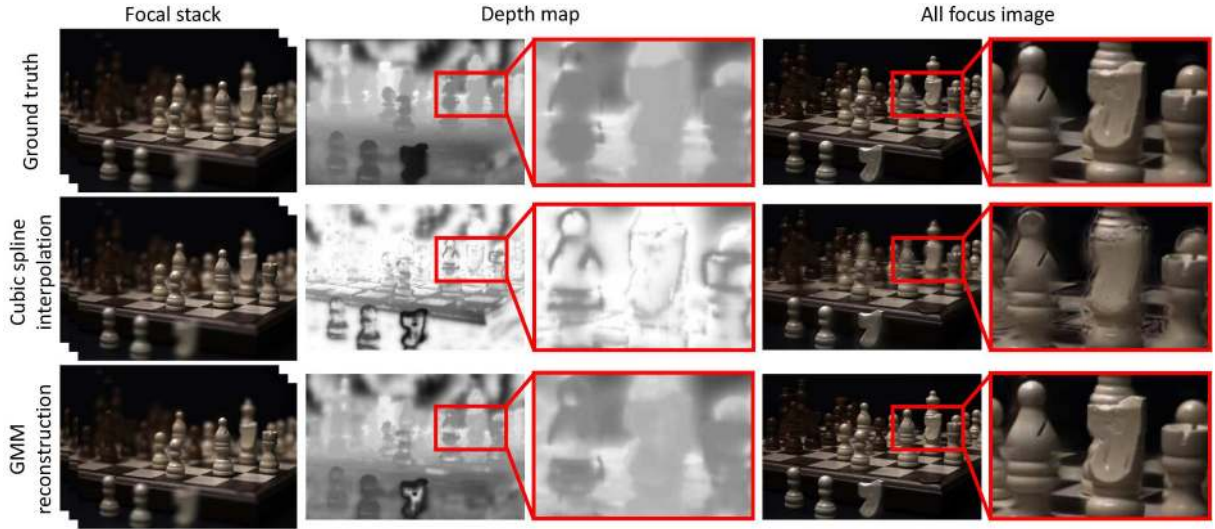


Figure 9: Depth map and all focus image from reconstructed “Chess” dataset. We reconstruct a focal stack of 45 images from 8 images using our algorithm and cubic spline interpolation. We then compute the depth maps and all focus images from these reconstructed stacks using a commercially available software Helicon soft. The depth map and all focus image obtained using our reconstructed stack are very close to the depth map and all focus image computed from the ground truth 45 images focal stack. The depth map obtained using spline interpolated focal stack is relatively poor. The all focus image obtained from spline interpolation show lots of artifacts.

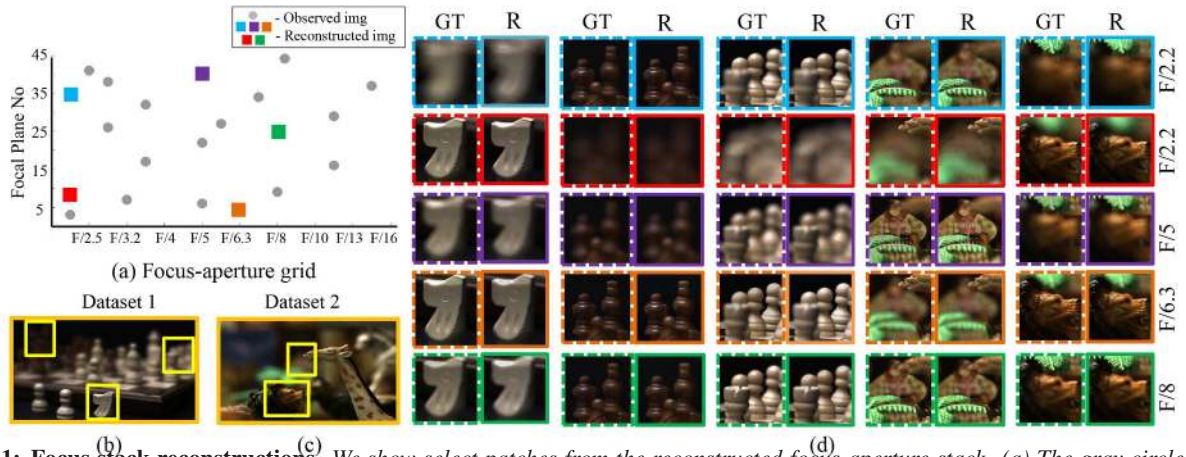


Figure 11: Focus stack reconstructions. We show select patches from the reconstructed focus-aperture stack. (a) The gray circles correspond to the focus-aperture combinations at which the 16 input images were obtained. (b, c) Sample images from reconstructed dataset. (d) We show the reconstructed patches (R) at locations which are far away from the sampled focus-aperture input images (see colored squares in (a)). A comparison with patches obtained from the ground truth (GT) data reveals that our reconstruction quality is high.

stereo algorithm to the entire focus-aperture stack. Further, looking at the histogram of depth errors over three different datasets, we note that our compressive epsilon photography approach starts producing competitive depth estimates from as little as 8 input images — a dramatic improvement over the original algorithm which required 793 images.

7 Conclusion and discussions

In this paper, we envision a framework for reconstructing the entire space of photographs that can be captured by a camera from just a few carefully selected images. Our framework enables an unprecedented level of freedom and flexibility in post-capture processing without the resolution limitations of light-field cameras. The two key ideas underlying our approach are the use of per-pixel modeling of intensity profiles and the use of Gaussian Mixture Models for capturing redundancies observed in epsilon photography stacks. Exploiting both of these, we show that we can reconstruct a pho-

tograph stack of thousands of images from just a few images. Further, applications that rely on focus-aperture stacks such as confocal stereo are enabled by collecting just a few images as opposed to hundreds to thousands of images.

Limitations. A key limitation of our algorithm, and an avenue for future work, is to account for scene motion and dynamic range. The results presented in this paper are applicable only for static or slow-moving scenes. However, our compressive framework enables reconstruction of the focus-aperture stack from as little as 8 – 16 images — which can be captured in rapid succession using the burst mode of the camera. By accounting for scene motion using optical flow-based registration, one could extend our techniques to dynamic scenes as well. Tackling dynamic range is another key avenue for future work. Our method is general enough to encompass both noise models inherent to HDR imaging as well as accounting for changes in ISO. In addition to these, clever strategies for sampling a large depth range without significantly increasing the

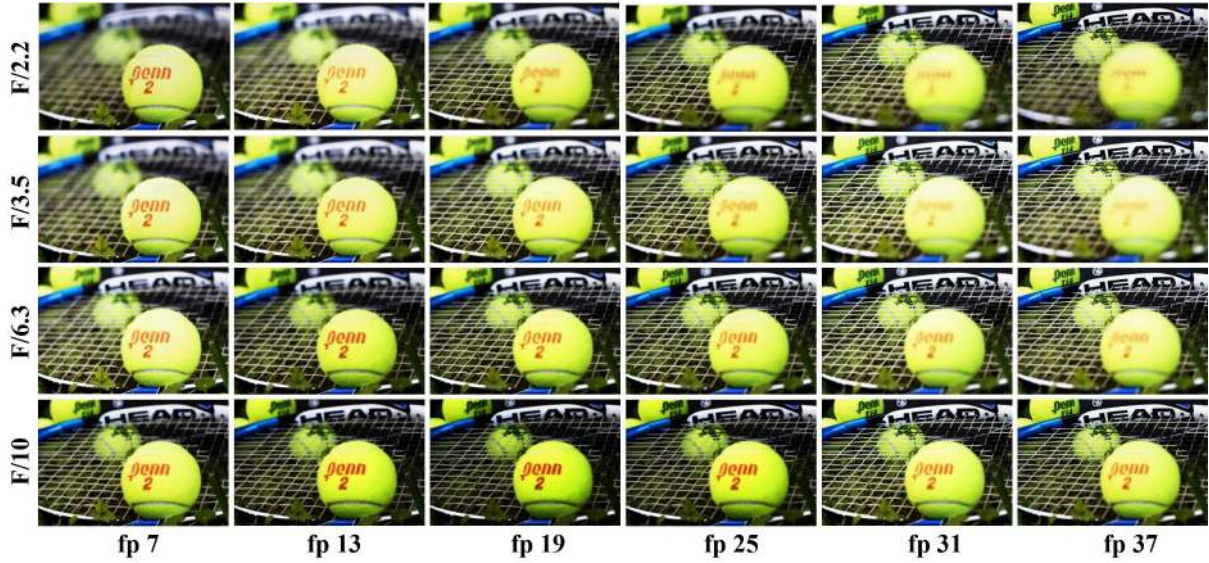


Figure 14: Focus-aperture stack reconstruction for outdoor "tennis" dataset. This figure shows the reconstruction results of a tennis racquet placed outdoors at 24 uniformly sampled focus-aperture combinations. 16 input images were used for reconstruction. Notice how the depth of field increases as one moves from F/2.2 to F/10. The blur too is modeled well, as can be seen for the images reconstructed at the largest aperture (F/2.2). Also note the change in focus as the focus plane (fp) varies from near-end to far-end with respect to the camera as one moves from left (fp 7) to right (fp 37).

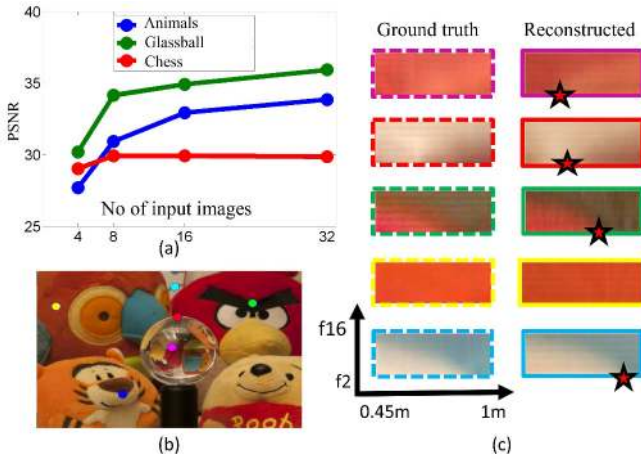


Figure 10: Reconstructed AFI images. (a) We plot the reconstruction accuracy of the focus-aperture stacks measured with PSNR as a function of number of input images. The reconstruction performance improves significantly as one moves from 4 input images to 16 input images for complex scenes with high texture content and intricate occlusions. For less complex scenes such as the chess dataset, we can reliably reconstruct the focus-aperture stack using just 8 images. (c) We show a set of AFI images obtained from the reconstructed focus-aperture stack for interesting points such as depth edges, texture edges, and flat surfaces corresponding to the points shown in (b). Note that the stars denote the computed depth/focal plane for the point under consideration.

per-pixel feature dimensionality and the associated training data requirements is an interesting avenue for future work.

Acknowledgements

This research was supported in parts by Sony Corp. AV, KM and ST were supported in part by NSF Grants IIS:1116718 and CCF:1117939, and the ONR sub-contract from Lockheed Mar-



Figure 12: Focal Aperture Stack Reconstruction of an unobserved focal-aperture setting. 16 images with varying focus and aperture settings given by our optimal sampling scheme are captured and the entire epsilon photography stack is reconstructed. (middle row) An example of a reconstructed image with zoomed in insets for two datasets. (top, bottom) Two observed images that are closest in focus and aperture settings to the one in the middle row. The reconstruction shows ability to transfer textural detail and sharpness to unobserved focal-aperture settings.

tin PO4100936535. ACS was supported in part by NSF Grant CCF:1117939.

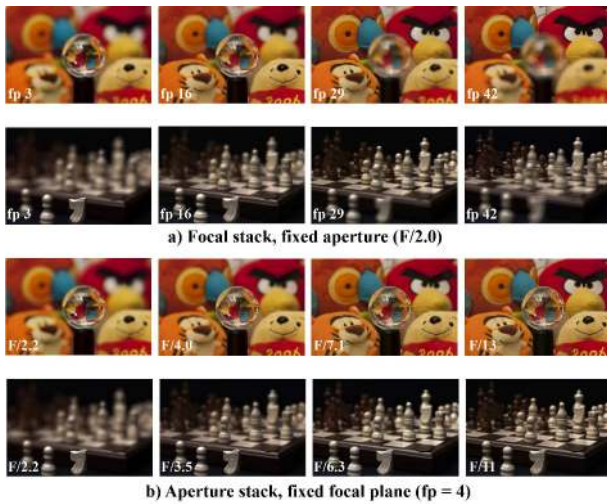
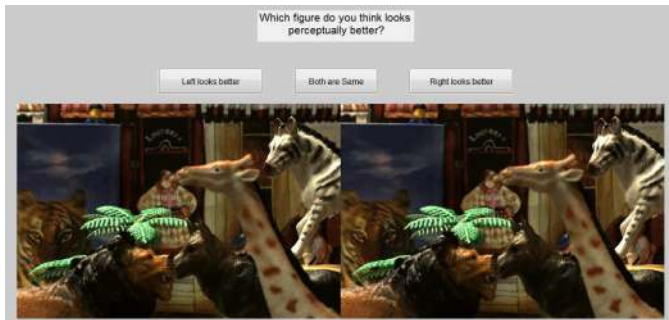
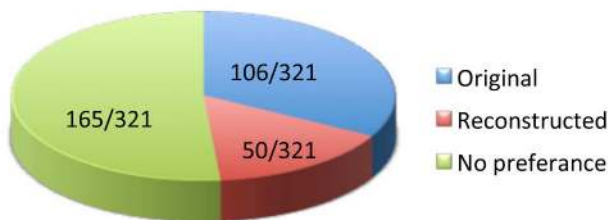


Figure 13: Focus-aperture stack reconstruction for “glassball” and “chess” datasets. *a)* shows 2 reconstructed focal-stacks for a fixed small aperture of $F/2.2$. One can clearly see the accurately modeled blurring effect as one moves from focal plane 3 (fp3) to focal plane 42 (fp42). *b)* Reconstructed aperture stacks for a fixed focal plane. As the aperture size increases from left to right, the increase in depth of field becomes apparent.



(a) Snapshot of our evaluation tool



(b) User preference statistics

Figure 15: Subjective evaluation. *(a)* We showed pairs of images – one original and one reconstructed – to 13 subjects and asked them to either pick a perceptually better image or to mark no preference between the two. Each subject was asked to evaluate 30 image pairs with no constraints on time for finishing the tasks. *(b)* In 51% percentage of instances, the subjects had no preference between the two images. This indicates both the high quality of our reconstructions and our ability to avoid perceptually glaring artifacts.

References

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. In *ACM Trans. Graphics*, vol. 23, 294–302.

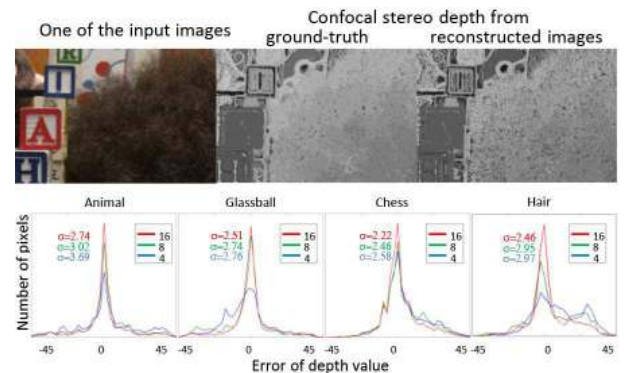


Figure 16: Compressive confocal stereo. *Confocal stereo* is a per-pixel depth estimation algorithm capable of estimating the depth for thin per-pixel structures. It uses the entire focal-aperture stack to produce these estimates. We obtain similar depth estimates in spite of acquiring only a few images. *(top-left)* An input image. *(top-center)* Depth estimate obtained from the entire focus-aperture stack of 1125 images. *(top-right)* Recovered depth map from focus-aperture stack reconstructed from 16 input images. *(bottom)* For four different scenes, we show histogram of depth errors obtained for different number of input images. Obtaining as little as 8 images is sufficient to obtain very precise depth estimates, as indicated by the sharp peak at the origin.

BARANIUK, R. G. 2007. Compressive sensing. *IEEE Signal Processing Magazine* 24, 4, 118–121.

BARON, D., SARVOTHAM, S., AND BARANIUK, R. G. 2010. Bayesian compressive sensing via belief propagation. *IEEE Trans. Signal Processing* 58, 1, 269–280.

BOOMINATHAN, V., MITRA, K., AND VEERARAGHAVAN, A. 2014. Improving resolution and depth-of-field of light field cameras using a hybrid imaging system. In *IEEE Intl. Conf. Computational Photography*.

BOURRIER, A., GRIBONVAL, R., PÉREZ, P., ET AL. 2013. Compressive gaussian mixture estimation. In *IEEE Intl. Conf. Acoustics, Speech and Signal Processing*.

BROWN, M., AND LOWE, D. G. 2007. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision* 74, 1, 59–73.

BUADES, T., LOU, Y., MOREL, J.-M., AND TANG, Z. 2009. A note on multi-image denoising. In *Intl. Workshop on Local and Non-Local Approximation in Image Processing*, 1–15.

CANDÈS, E. J., AND WAKIN, M. B. 2008. An introduction to compressive sampling. *IEEE Signal Processing Magazine* 25, 2, 21–30.

DEBEVEC, P., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, 369–378.

FLAM, J. T., CHATTERJEE, S., KANSANEN, K., AND EKMAN, T. 2011. Minimum mean square error estimation under gaussian mixture statistics. *arXiv preprint arXiv:1108.3410*.

FLAM, J. T., CHATTERJEE, S., KANSANEN, K., AND EKMAN, T. 2012. On MMSE Estimation: A linear model under gaussian mixture statistics. *IEEE Trans. Signal Processing* 60, 7, 3840–3845.

GORTLER, S., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. 1996. The lumigraph. In *SIGGRAPH*, 43–54.

- GREEN, P., SUN, W., MATUSIK, W., AND DURAND, F. 2007. Multi-aperture photography. *ACM Trans. Graphics* 26, 3, 68.
- GROSSMANN, P. 1987. Depth from focus. *Pattern Recognition Letters* 5, 1 (Jan.), 63–69.
- HASINOFF, S. W., AND KUTULAKOS, K. N. 2006. Confocal stereo. In *European Conf. Computer Vision*. Springer, 620–634.
- HASINOFF, S. W., AND KUTULAKOS, K. N. 2007. A layer-based restoration framework for variable-aperture photography. In *IEEE Intl. Conf. Computer Vision*, 1–8.
- HASINOFF, S. W., AND KUTULAKOS, K. N. 2009. Confocal stereo. *International Journal of Computer Vision* 81, 1, 82–104.
- HASINOFF, S. W., DURAND, F., AND FREEMAN, W. T. 2010. Noise-optimal capture for high dynamic range photography. In *IEEE Conf. Computer Vision and Pattern Recognition*, 553–560.
- JOSHI, N., AND COHEN, M. F. 2010. Seeing Mt. Rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal. In *IEEE Intl. Conf. Computational Photography*, 1–8.
- KROTKOV, E. 1988. Focusing. *International Journal of Computer Vision* 1, 3, 223–237.
- KUTHIRUMMAL, S., NAGAHARA, H., ZHOU, C., AND NAYAR, S. K. 2011. Flexible depth of field photography. *IEEE Trans. Pattern Analysis and Machine Intelligence* 33, 1, 58–71.
- KUTULAKOS, K., AND HASINOFF, S. W. 2009. Focal Stack Photography: High-performance photography with a conventional camera. In *Intl. Conf. Machine Vision and Applications*, 332–337.
- LAW, N. M., MACKAY, C. D., AND BALDWIN, J. E. 2005. Lucky imaging: High angular resolution imaging in the visible from the ground. *arXiv preprint astro-ph/0507299*.
- LEVIN, A., AND DURAND, F. 2010. Linear view synthesis using a dimensionality gap light field prior. In *IEEE Conf. Computer Vision and Pattern Recognition*, 1831–1838.
- LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *SIGGRAPH*, 31–42.
- LYTRO. The lytro camera. <https://www.lytro.com/>.
- MANN, S., AND PICARD, R. 1994. *Being undigital with digital cameras*. MIT Media Lab Perceptual.
- MARWAH, K., WETZSTEIN, G., BANDO, Y., AND RASKAR, R. 2013. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graphics* 32, 4, 46.
- MCNALLY, J. G., KARPOVA, T., COOPER, J., AND CONCHELLO, J. A. 1999. Three-dimensional imaging by deconvolution microscopy. *Methods* 19, 3, 373–385.
- MITRA, K., AND VEERARAGHAVAN, A. 2012. Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch prior. In *IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, 22–28.
- MITRA, K., COSSAIRT, O., AND VEERARAGHAVAN, A. 2014. Can we beat hadamard multiplexing? data driven design and analysis for computational imaging systems. In *IEEE Intl. Conf. Computational Photography*.
- PARK, J. Y., AND WAKIN, M. B. 2009. A multiscale framework for compressive sensing of video. In *Picture Coding Symposium*, 1–4.
- PEERS, P., MAHAJAN, D. K., LAMOND, B., GHOSH, A., MATUSIK, W., RAMAMOORTHY, R., AND DEBEVEC, P. 2009. Compressive light transport sensing. *ACM Trans. Graphics* 28, 1, 3.
- RASKAR, R. 2009. Computational photography: Epsilon to coded photography. In *Emerging Trends in Visual Computing*. Springer, 238–253.
- RAYTRIX. 3d light field camera technology. <http://www.raytrix.de/>.
- RUBINSTEIN, R., BRUCKSTEIN, A. M., AND ELAD, M. 2010. Dictionaries for sparse representation modeling. *Proceedings of the IEEE* 98, 6, 1045–1057.
- SIBARITA, J.-B. 2005. Deconvolution microscopy. In *Microscopy Techniques*. 201–243.
- SROUBEK, F., AND MILANFAR, P. 2012. Robust multichannel blind deconvolution via fast alternating minimization. *IEEE Trans. Image Processing* 21, 4, 1687–1700.
- TAMBE, S., VEERARAGHAVAN, A., AND AGRAWAL, A. 2013. Towards motion-aware light field video for dynamic scenes. In *IEEE Intl. Conf. Computer Vision*.
- WAGADARIKAR, A., JOHN, R., WILLETT, R., AND BRADY, D. 2008. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics* 47, 10, B44–B51.
- YANG, J., LIAO, X., YUAN, X., LLULL, P., BRADY, D. J., SAPIRO, G., AND CARIN, L. Compressive sensing by learning a gaussian mixture model from measurements.
- YANG, J., YUAN, X., LIAO, X., LLULL, P., SAPIRO, G., BRADY, D. J., AND CARIN, L. 2013. Gaussian mixture model for video compressive sensing. In *International Conference on Image Processing*.
- YU, G., SAPIRO, G., AND MALLAT, S. 2012. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE Trans. Image Processing* 21, 5.
- YUAN, L., SUN, J., QUAN, L., AND SHUM, H.-Y. 2007. Image deblurring with blurred/noisy image pairs. In *ACM Trans. Graphics*, vol. 26, 1.