Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2006-058 September 2, 2006

# Random Lens Imaging

Rob Fergus, Antonio Torralba, and William T. Freeman

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

# Random Lens Imaging

**Rob Fergus***
CSAIL, MIT
Cambridge, MA 02139
fergus@csail.mit.edu

**Antonio Torralba***
CSAIL, MIT
Cambridge, MA 02139
torralba@csail.mit.edu

**William T. Freeman**
CSAIL, MIT
Cambridge, MA 02139
billf@csail.mit.edu

## Abstract

We call a random lens one for which the function relating the input light ray to the output sensor location is pseudo-random. Imaging systems with random lenses can expand the space of possible camera designs, allowing new trade-offs in optical design and potentially adding new imaging capabilities. Machine learning methods are critical for both camera calibration and image reconstruction from the sensor data. We develop the theory and compare two different methods for calibration and reconstruction: an MAP approach, and basis pursuit from compressive sensing [5]. We show proof-of-concept experimental results from a random lens made from a multi-faceted mirror, showing successful calibration and image reconstruction. We illustrate the potential for super-resolution and 3D imaging.

## 1 Introduction

An ordinary lens provides a particular relationship between an incoming light ray and the refracted ray: when the imaging system is in focus, all the rays emanating from a point in space are mapped onto the same location of the sensor array. This makes estimating the image from sensor data particularly simple: the sensor data directly gives the estimated image. On the other hand, this condition imposes tight constraints on the optical design, and restricts the form of the responses to out-of-focus scene elements. Taking advantage of the opportunity for computer processing after image acquisition, other functions relating the input and output light rays have begun to be explored. Such extensions include: methods to achieve all-focus cameras, called wavefront imaging [3, 6]; generalizations of the pinhole camera called coded aperture imaging [7] and reference aperture tomography [1]; and methods to surpass detector aperture resolution limitations using small perturbations to the point spread function [9].

In this paper, we explore further in computational imaging and ask: what if the input to output relationship of the light rays were randomized? How could you use such a camera? What would be the benefits and the drawbacks of such an approach? The ability to use a camera with those characteristics would open new regions of the space of possible optical designs, since in many cases it is easier to randomize light ray directions than to precisely direct them. We will show that this tradeoff could potentially allow ultra-thin designs, simplified multi-spectral and high-dynamic range imaging, as well as all-focus imaging or the three-dimensional reconstruction of nearby objects.

Mathematical techniques for constrained optimization and Bayesian inference are essential to such an optical design. In this paper, we provide a number of the tools needed to calibrate such a camera

*The contributions of the first and second authors varied randomly from day to day but were equal in total.

and to form images from it. We believe these mathematical problems will be of interest to the machine learning community, and that the community, in turn, may discover other problems of interest in this area. We describe the basics and the promise of random lens imaging in Section 2. Section 3 develops theory for both calibration and image decoding. Section 4 shows proof-of-concept experimental results, showing successful calibration, decoding and super-resolution and comparing different image reconstruction methods.

## 2 Basics of random lens imaging

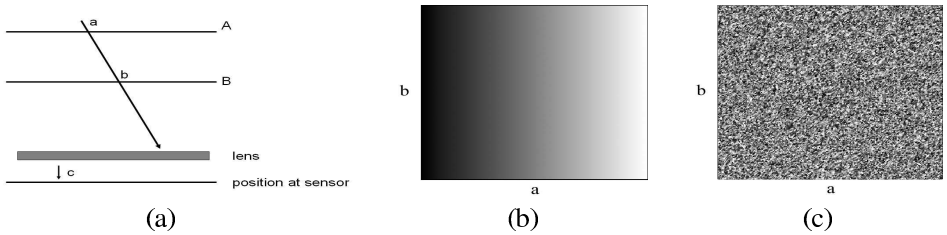


Figure 1: Explanatory figures for a 2-dimensional world. (a) Light ray parameterization. A lens directs each ray (a,b) to some sensor position, c. (b) Output position, shown as intensity for each ray (a,b) for a conventional lens focussed at plane A. (c) Same, for a random lens system.

Figure 1 shows, for a two-dimensional world, the functional characteristics of a conventional lens and a random lens. A lightfield of rays [8], parameterized by intersection points a and b with lines A and B, impinges on a lens. The lens redirects each ray to a sensor position, c. For a conventional lens focussed at line A, the function c(a, b) is independent of the intersection point b, giving the function c(a, b) shown in Fig. 1 (b). A random lens may have a function c(a, b) as shown in Fig. 1 (c).

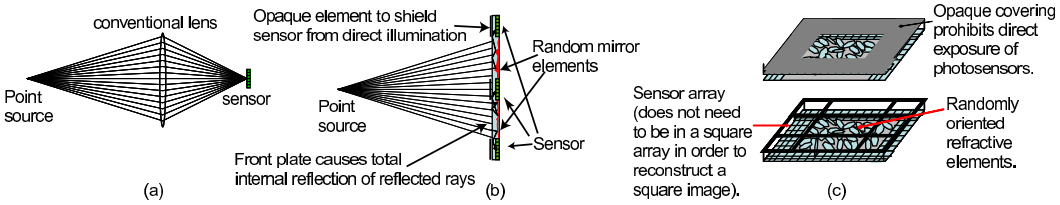


Figure 2: Candidate physical designs. (a) Conventional lens. (b) Random lens using reflective elements, (c) Random lens using refractive elements.

Accepting ray transfer functions of this form can allow for unconventional optical designs. Figure 2 shows a conventional lens design, (a), and two possible random lens devices, using (b) reflective and (c) refractive optical elements. For a conventional lens, a sufficient distance between the sensor and the lens is needed to allow small angular differences in light rays to translate to sufficiently different positions at the sensor. In both random lens designs, the distance allowing the translation of small angular variations into different sensor positions is folded into the camera aperture itself, enabling a camera that only needs to be large in two dimensions, not three. (Others are also pursuing such folded light path cameras [9], although a non-randomized design imposes much stricter physical tolerances.)

Donoho and others have recently pointed out advantages of measuring random projections from structured signals, which they term compressive sensing [2, 5]. Building on this, researchers [13] have developed an optical bench system that images onto a micro-mechanical mirror array, then onto a single sensing element, allowing measurement of random projections of an image in a time-multiplexed manner, one projection at a time. Images from a random lens system are also random projections of a structured signal, although, in contrast with [13], the desired measurements are recorded simultaneously at all the sensor pixels and we exploit the randomization to simplify the optical design. In Section 4, we compare the performance of basis pursuit projection proposed for compressive sensing [5] with a conjugate-gradient ascent solution of the corresponding MAP estimation problem.

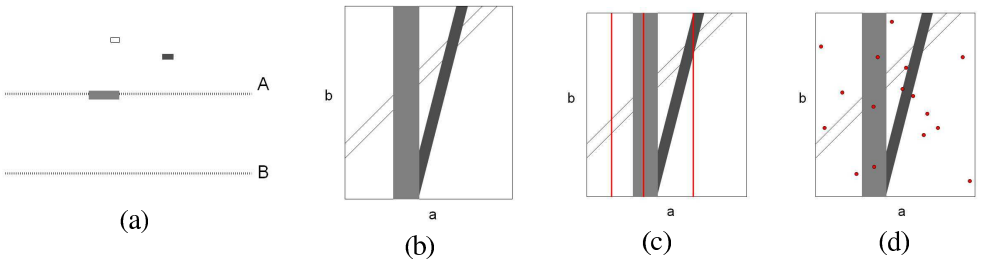


Figure 3: (a) Consider these 3 Lambertian objects in our 2-d world. (b) The resulting lightfield, or intensity of each ray (a,b). Under most conditions, the lightfield exhibits extraordinary structure and redundancy. (c) Conventional lens, focussed at A, integrates at each sensor position along vertical slices of this lightfield, like the 3 integral lines shown in red. (d) A random lens sensor element integrates over a pseudo-random set of lightfield points.

Various characteristics of random lenses can be exploited in cameras which use them. Any attenuation at the lens surface is automatically incorporated into the camera calibration. Thus, three-color, multi-band, or high dynamic range imaging can all be enabled by painting patterns of spectral or intensity attenuation over the random lens surface. (During camera calibration, the appropriate generalizations of the greyscale calibration patterns described later would be used. For example, color calibration patterns would be required for color image reconstructions). A color image, painted onto the lens surface, could suffice to measure full-color image data with that lens, allowing advertising, favorite photographs, or even images designed to conceal the camera, to be painted on, while enabling color photography. High-dynamic range attenuation patterns would allow high-dynamic range photography.

The sensor pixels of a random lens camera do not need to be in a square array, or even contiguous. A random lens camera could use solid state sensors with damaged or missing pixels as easily as perfect ones, allowing the use of sensors that would otherwise need to be rejected for conventional lens cameras.

Because the light ray direction is encoded in the sensor response, there is the potential to measure 3-d position as well as image intensity. This application, even more than 2-d imaging, will require the use of strong signal priors, since, even with high resolution image sensors, there won't be enough sensor data to describe all possible configurations of 3-d position and intensity. However, the light fields associated with conventional images and surfaces are extremely structured and redundant [8], creating the potential for recovering 3-d information, if the appropriate priors and reconstruction methods are used. Figure 3 shows the analog of 3-d imaging for our pedagogical 2-d visual world. We illustrate 3-d sensitivity with our experimental apparatus in Section 4.

We want to point out some drawbacks of random lens imaging. First is the need for both initial camera calibration and a reconstruction procedure before any image can be viewed. Sensor signal-to-noise will limit image reconstruction quality, although strong image priors will help in that regard. Finally, there is, in principle, the need for one matrix the size of the number of sensor elements to specify how to reconstruct the output image *at each output image pixel*. A million pixel output image could require 1 million matrices, each of a million elements! However, in practise these matrices are *very* sparse: the underlying physical surfaces generating the random lens ray deviations can be described with many fewer numbers. The camera methods we develop in the next section encourage sparse calibration matrices. Such sparsity is required for random lens imaging.

## 3 Calibration and image decoding

Before use, the camera must be calibrated to determine the mapping between the real world and the sensor output. Then, images of natural scenes may be decoded from the jumbled images recorded at the sensor. While calibration and decoding are two distinct operations from a practical perspective, the underlying theory is similar for the two problems and we propose a common solution to both problems.

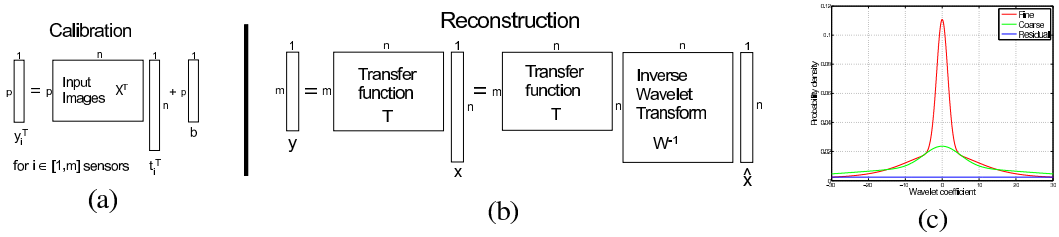


Figure 4: The two stages of operation of a random camera. (a) In calibration we solve for the transfer function of the camera $T$, row-by-row using a pictures $X$ and the corresponding sensor responses $Y$. (b) In reconstruction, the roles of the calibration matrix and images are reversed. The input image $\mathbf{x}$ is now inferred from the sensor readings, via the transfer function. The decoding is regularized by requiring $\mathbf{x}$ to be sparse in some wavelet basis $W$. (c) shows the mixture-of-Gaussians prior used in the MAP decoding method to enforce sparsity in the image's wavelet coefficients $\hat{\mathbf{x}}$. The different priors used at each wavelet scale level are shown in different colors.

The camera is characterized by an $m$ by $n$ sparse transfer function matrix $T$ which maps real world images $\mathbf{x}$ to sensor values $\mathbf{y}$. The resolution $n$ of image $\mathbf{x}$ is typically greater than the $m$ pixels in the sensor. In calibration the goal is to determine $T$ using a set of calibration images $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ and associated sensor readings $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_p\}$. In decoding the goal is to recover the real world picture $\mathbf{x}$ from the recorded sensor values $\mathbf{y}$ using $T$. These procedures are summarized in Figure 4(a) & (b). Note that structure of both problems differs only in the roles of $X$ and $T$.

In calibration, we have $mn$ unknowns so require $n$ independent images, each containing $m$ values, to completely determine $T$. Given that $m$ is $O(10^6)$ in a modern CMOS sensor, it is not practical to gather $n$ images so we are forced to estimate $T$ from $<< n$ images. In decoding, we record $m$ values at the sensor and wish to reconstruct a higher dimensional signal of length $n$. Hence, in both cases the number of unknowns is greater than the number of observations, making them under-determined problems. To make the problems tractable we assume that the unknowns are *sparse* in some basis (i.e. the majority have negligible magnitude). An additional constraint is the additive nature of light, meaning that $\mathbf{x}$ and $T$ are confined to $\mathcal{R}^+$.

### 3.1 Sparsity

We adopt two different approaches to solving the under-determined systems of linear equations encountered in calibration and decoding. The first is Basis Pursuit De-Noising (BPDN) [4, 11], as used by Donoho [5]. This method imposes sparsity on the $l_1$ form of the signal to be recovered. In the case of the decoding problem, this signal is $\hat{\mathbf{x}}$, sparse in some wavelet basis $W$ and BPDN solves:

$$\min_{\hat{\mathbf{x}}} \frac{1}{2}||\mathbf{y} - TW^{-1}\hat{\mathbf{x}}||_2^2 + \lambda||\hat{\mathbf{x}}||_1 \quad (1)$$

BPDN reformulates this minimization as a perturbed linear program guaranteed to find the exact solution in polynomial time.

In the second approach, we pose the problem as MAP estimation. A likelihood term ensures that the reconstructed signal is close to the observed one. The prior term, modeling the sparsity in the signal, guides the estimation procedure towards solutions of high posterior probability. The details of the likelihood and prior function used are given in Sections 3.2 and 3.3. For comparison purposes, we also use Matlab's `pinv` function which in under-determined problems provides the minimum $l_2$ norm solution.

### 3.2 Calibration

In calibrating the camera we need to determine the transfer function matrix $T$, a procedure that only needs to be performed once. While $T$ has $mn$ elements it is very sparse, the number of non-zero elements $e$ being determined by the level of randomness in the lens. A theoretical result from Donoho [5] predicts that the number of random projections (calibration images in our case) required to determine T is $O(e \log n)$, a drastic reduction from $O(n)$ in the case of a full $T$.

To find $T$ we present a set of $p$ calibration images (stored in matrix $X$) to the camera, recording the sensor values in another matrix $Y$. $X$ is $n$ by $p$ while $Y$ is $m$ by $p$. Then we solve for each row of $T$ independently, using $X$ and the corresponding row from $Y$. See Figure 4(a) for an overview.

In the MAP formulation, we seek the most probable $\mathbf{t}_i$ given $X$ and $\mathbf{y}_i$, where $\mathbf{t}_i$ is the $i$-th row of $T$. We choose the likelihood term to be Gaussian, with variance $\sigma^2$. To account for DC offsets due to background lighting, we include a mean offset $\mathbf{b}$ in the estimation. For simplicity, we assume the elements in $\mathbf{t}_i$ to be independent and model them with an exponential prior, reflecting their heavy-tailed nature and positivity. The posterior we optimize is thus:

$$p(\mathbf{t}_i|X,\mathbf{y}_i) = k\, p(\mathbf{y}_i|X,\mathbf{t}_i)\, p(\mathbf{t}_i) = k\, \mathcal{N}(\mathbf{y}_i^T|X^T\mathbf{t}_i^T + \mathbf{b}, \sigma^2) \prod_j \mathcal{E}(t_{ij}|\kappa) \tag{2}$$

We also apply BPDN in the manner described in Eqn. 1 but with $\mathbf{t}_i$ being the unknown rather than $\hat{\mathbf{x}}$. Since we expect the matrix to be sparse, we need not project $\mathbf{t}_i$ into any basis, thus we use $W = I$. The choice of images for calibration is important to ensure reliable estimation of $T$. We used random binary images (e.g. Fig. 6(b)). This choice was motivated by the investigations of Tsiag and Donoho [12], who found that the quality of reconstruction obtained in their compressed sensing scheme was dependent only on the independence of the projections; the exact nature of the random distribution being unimportant.

### 3.3 Decoding

When trying to recover $\mathbf{x}$ (the image) there are two problems: (a) typically $m < n$ giving an under-determined system, and (b) both $\mathbf{y}$ and $T$ are corrupted by noise. To regularize the problem, we use the sparsity property of wavelet decompositions of natural images. Representing the wavelet transformation by the matrix $W$, we seek sparsity in the wavelet coefficients $\hat{\mathbf{x}} = W\mathbf{x}$.

In the MAP framework, we employ a multi-scale prior, capturing the difference in coefficient magnitudes across scale (see Figure 4(c)). At each scale $s$ the distribution is modeled by a mixture-of-Gaussians with $K = 4$ components. The parameters of the prior $\pi_k^s, v_k^s$ are estimated with EM from a collection of natural images. Following a similar formulation to the calibration, we have:

$$\mathbf{x} = \arg\min_{\mathbf{x}} p(\mathbf{y}|T,W,\mathbf{x})\, p(W\mathbf{x}) = \mathcal{N}(\mathbf{y}|T\mathbf{x},\sigma^2) \prod_s \prod_{i^s} \sum_{k=1}^{K} \pi_k^s\, \mathcal{N}(\hat{\mathbf{x}}_i^s|0, v_k^s) \tag{3}$$

BPDN is employed in the manner described in Eqn. 1. It differs from the MAP reconstruction in that the sparsity in $\hat{\mathbf{x}}$ is enforced uniformly over scales.

## 4 Experimental results

We show results obtained with a prototype random lens camera, designed to show the feasibility of such a device and to validate some of the theoretical results derived above. Figure 5 shows the hardware setup used for all experiments. We modified a Pentax stereo adapter so that one of the mirrors has a random reflective surface (the other aperture was blocked off) and mounted it on a Canon DSLR camera. The random surface was constructed by gluing many small pieces of reflective Mylar onto a piece of black card. Note that the setup has no conventional lens: only one planar mirror and one random reflective surface guides the light to the sensor.

Experiments were performed by projecting images onto a wall in a darkened room, and taking pictures of the scene with the random lens camera. Figure 6 shows examples of pairs of projected images and the corresponding output of the sensor. The goal is to first compute the transfer function of the camera and then use it to decode the random outputs obtained by projecting test images of natural scenes.

### 4.1 Calibration

The calibration was performed by projecting binary random images at different resolutions (8x8, 16x16, 32x32 and 64x64 pixels). These patterns form $X$ and the images captured by the camera

Figure 5: A closeup of the random reflective surface and camera setup used in our experiments. The schematic diagram on the right shows the light path to the sensor.

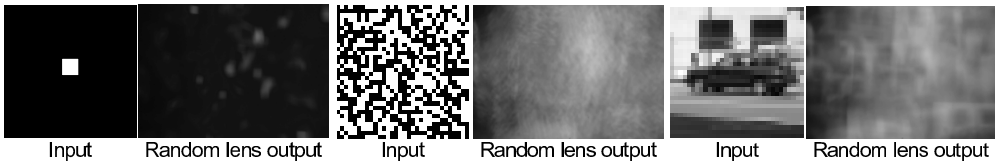


Figure 6: Examples of pictures taken with our random lens camera. Each pair shows an image projected on the wall, and the output of the camera.

form $Y$. We then computed $T$ using the MAP and BPDN methods described above. For a baseline comparison, a least squares reconstruction was also performed. Fig. 7 shows a small portion of $T$ estimated using the different methods.

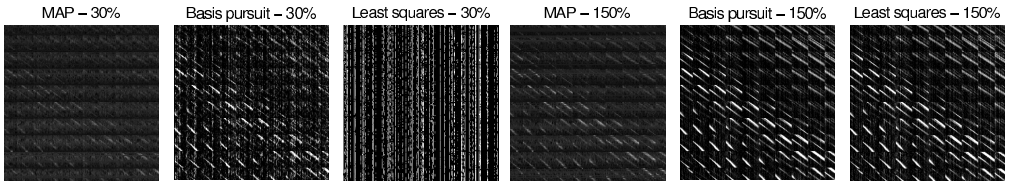


Figure 7: Cropped regions of the calibration matrices for 32x32 images estimated using three methods: MAP, BPDN and least squares (Matlab \). Note the sparse structure. On the left are matrices estimated from 30% of the 1024 required for a full rank estimation. On the right are matrices estimated using 1500 images, $\approx$ 150% of the required number. The lack of prior in the least squares reconstruction results in a degenerate solution containing unrealisable negative values if fewer than 1024 images are used.

One important observation is the experimental verification of the logarithmic-style dependency between the number of calibration images needed to achieve a given reconstruction error and the resolution. Figs. 8(a)&(c) shows the estimation error of $T$, of MAP and BPDN methods respectively, for different resolutions as a function of the number of calibration images relatively to the number of sensors. Fig. 8(b) shows that the number of calibration images needed to achieve a fixed error (0.1) grows sub-linearly with the number of sensors.

### 4.2 Decoding

We tested our camera on a set of 50 32x32 grayscale patches extracted from natural scenes. The wavelet basis $W$ chosen was a 9-tap symmetric quadrature mirror filter [10]. The decoding performance of six different transfer functions was evaluated. These were produced by three different calibration methods (MAP, BPDN and Matlab's `pinv`) using two different sizes of calibration im-

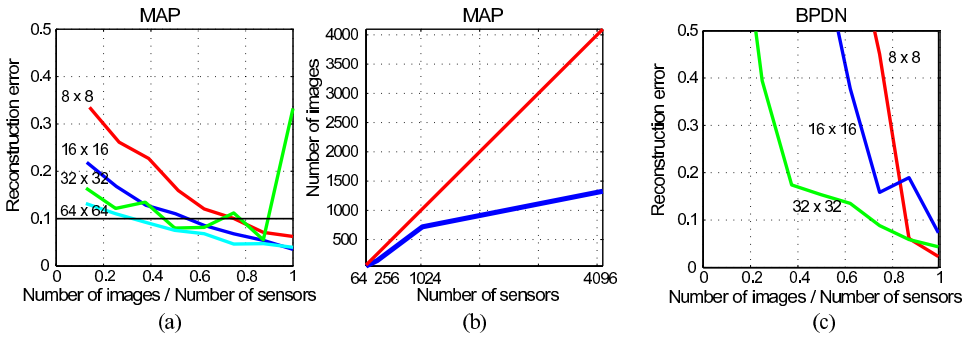


Figure 8: a) MAP calibration error. b) Logarithmic growth of number of calibration images with sensor resolution (blue), compared to linear relationship (red). c) BPDN calibration error.

age sets - 30% and 150% respectively of the 1024 needed for full rank estimation. These 6 matrices were then used to decode the test images with the MAP and BPDN methods. The reconstruction errors are shown in Table 1 and examples of reconstructed images shown in Figure 9.

| Decoding method | # images/ # sensors | Transfer function estimation method MAP | BPDN | `pinv` |
|---|---|---|---|---|
| MAP | 30% | $0.58 \pm 0.13$ | $\mathbf{0.70} \pm 0.19$ | $0.67 \pm 0.1$ |
| MAP | 150% | $0.71 \pm 0.14$ | $\mathbf{0.78} \pm 0.19$ | $0.76 \pm 0.2$ |
| BPDN | 30% | $0.15 \pm 0.08$ | $0.47 \pm 0.07$ | $0.42 \pm 0.1$ |
| BPDN | 150% | $0.22 \pm 0.11$ | $0.68 \pm 0.14$ | $0.67 \pm 0.5$ |

Table 1: Comparison of decoding methods. 6 different calibration matrices produced using 3 different estimation with 2 different sized calibration sets. Scores are normalized correlation to ground truth (i.e. higher is better)

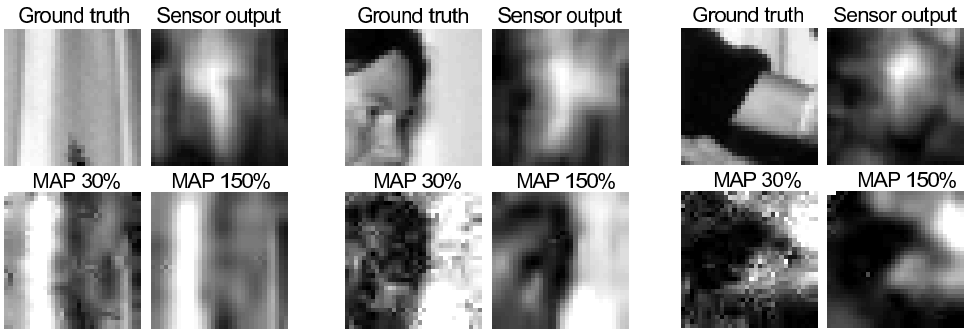


Figure 9: Reconstruction results using MAP for 3 images for two different transfer function matrices.

### 4.3 Enhanced capabilities of random lenses

**Super-resolution.** Fig. 10 illustrates (with a simulated camera) using BPDN how the signal can be recovered at a higher resolution than the number of sensors of the camera. The top row shows the original image and the images recovered when only a subset of the pixels are measured. The bottom row shows the images recovered when subsampling the output of the random lens camera. Using the image projection provided by a random lens camera allows recovering the original signal even when using fewer sensors (10%) than the final image resolution.

**Depth estimation.** Fig. 11 shows how the impulse response of the camera changes with distance. The impulse response is obtained by illuminating a black surface with a red laser pointer. The bottom row shows the corresponding pictures taken with a pinhole camera. In a pinhole camera, as the light source gets closer to the camera, it results in a blurrier picture. However, in a random lens camera (top row), depth variations produce a more complex pattern of variation which provides potential information for recovering depth. There are multiple copies of the light point and their relative positions vary with depth (the trails show the trajectory of a selected set of points to show that the

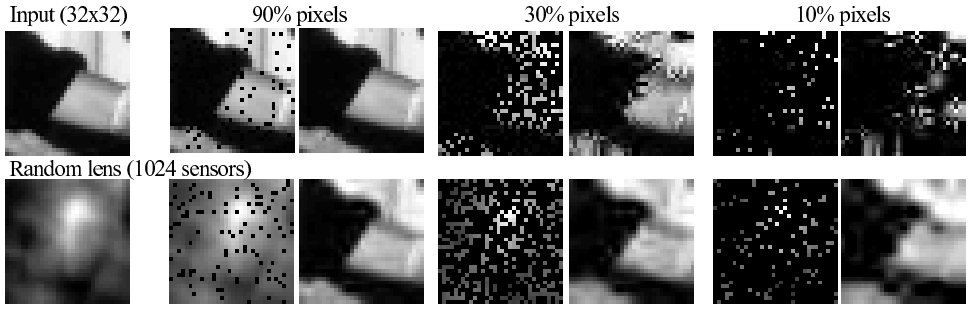


Figure 10: Super-resolution and compressive sensing with a random lens camera.

variation of configurations is more complex than a translation or a scaling). Depth estimation using a random lens remains an open issue.

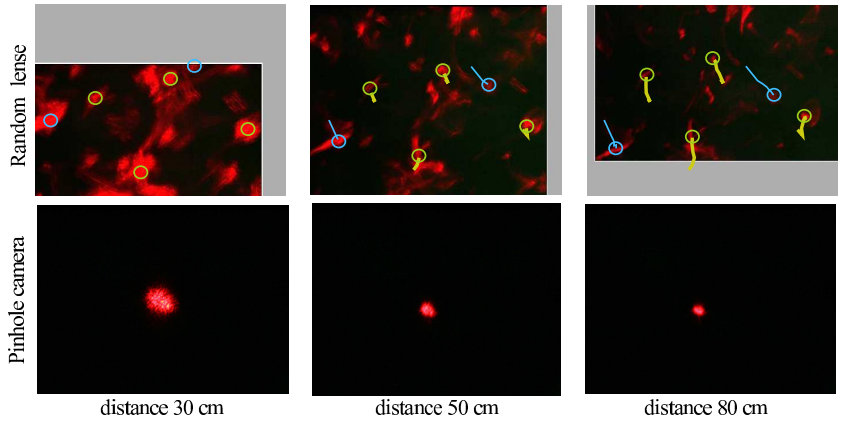


Figure 11: Impulse response of a random lens (top) and a pinhole camera (bottom) for light sources at three different distances. The images are aligned so that there is one stable feature at the center (this results in some image cropping).

## 5 Discussion

These experiments test and compare the algorithms developed for calibration and reconstruction, and validate the promise for super-resolution performance for random lens imagers. This research area leverages the power of machine learning methods to enable radical engineering designs.

### Acknowledgments

The authors would like to thank Fredo Durand, Ted Adelson, George Barbastathis and Stephen Fantone for their insights and suggestions. We would also like to thank David Donoho and his group for making their compressed sensing code available online. We are indebted to Fredo Durand for letting us borrow much of the equipment used in our experiments.

## References

[1] D. J. Brady, N. P. Pitsiansis, and X. Sun, "Reference structure tomography", *J. Opt. Soc. Am. A*, Vol. 21, No. 7, July 2004

[2] E. Candes, J. Romberg and T. Tao. "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", 2004.

[3] W. T. Cathey, R. Dowski, "New paradigm for imaging systems", *Applied optics* Vol. 41, pp. 1859–1866, 1995.

[4] S. Chen, D. Donoho, and M. Saunders. " Atomic Decomposition by Basis Pursuit." *SIAM J. Sci Comp.* Vol. 20(1), pp. 33–61, 1999.

[5] D. Donoho. "Compressed Sensing." `http://wwwstat.stanford.edu/donoho/Reports/2004/CompressedSensing091604.pdf`, 2004.

[6] E. Dowski and G. Johnson. "Wavefront coding: A modern method of achieving high performance and/or low cost imaging systems", *Current Developments in Optical Design and Optical Engineering VIII, Proc. SPIE* Vol. 3779, pages 137–145, Oct. 1999.

[7] E. Fenimore, T. Cannon. "Coded aperture imaging with uniformly redundant rays." *Applied Optics*, 1978.

[8] M. Levoy and P. Hanrahan "Light Field Rendering", *Proc. SIGGRAPH*, 1996.

[9] M. Neifeld and A. Askok"Imaging Using Alternate Point Spread Functions: Lenslets with Pseudo-Random Phase Diversity", `https://montage.ece.arizona.edu/public/Alternate_PSFs.pdf`, 2006.

[10] E. Simoncelli and E. Adelson,"Subband Transforms", *Subband image coding*, Ch. 4, pages 143–192, Kluwer Academic Publishers, 1990.

[11] SparseLab software. `http://sparselab.stanford.edu`

[12] Y. Tsaig and D. Donoho. "Extensions of Compressed Sensing." *EURASIP Signal Processing Journal*, 2006.

[13] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly and R. Baraniuk "A New Compressive Imaging Camera Architecture using Optical-Domain Compression" *Proc. of Computational Imaging IV at SPIE Electronic Imaging, San Jose, CA*, Jan. 2006.

[14] `https://montage.ece.arizona.edu/public/Folded_Lens_Briefing.ppt`