Resolution Preserving Light Field Photography Using Overcomplete Dictionaries And Incoherent Projections

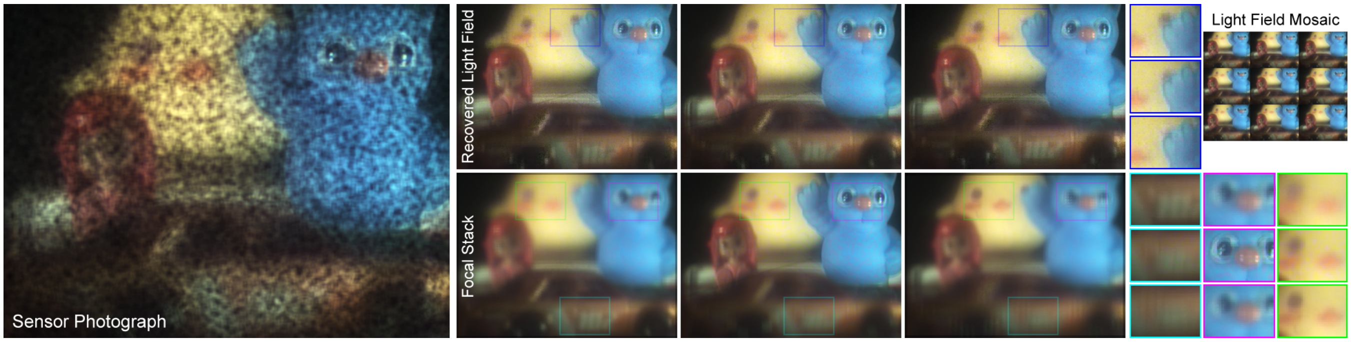


Figure 1: *Light field reconstruction from a single, coded sensor image (left). We show how to capture the essence of natural light fields in learned dictionaries, which—in combination with optical attenuation masks and compressive computational reconstruction—facilitate resolution-preserving light field recovery. Parallax is preserved both horizontally and vertically (upper right); the lower row demonstrates applications to refocusing a photograph after capture. As opposed to previous work, our dictionary-based approach to compressive light field sampling handles specularities, occlusions, and other complex effects as observed on the blue bear's eye and hand (upper row), respectively.*

Abstract

We present a computational framework and mask-based optical design for resolution-preserving light field reconstructions from a single modulated sensor image. Compressive computational reconstruction techniques are used in combination with learned overcomplete dictionaries that capture the essential building blocks of natural light fields. The mask patterns in the camera create incoherent projections of the recorded light field on the sensor image. Unlike traditional methods for light field super-resolution, our technique can recover fine image details, occlusions, specularities, translucencies, and other challenging illumination effects. With a prototype camera, we demonstrate the practicality of the proposed framework and show reconstructed light fields with applications in changing viewpoint and focus after an image is captured.

1 Introduction

Conventional cameras capture a two-dimensional photograph—the projection of the four dimensional radiance function incident on the sensor. Affordable light field cameras, capturing the full 4D radiance function, are emerging on the consumer market [Lytro 2012]. The main functional advantage offered by these cameras is the ability to change viewpoint and focus in post-processing; a feature that will be commonplace in next-generation cameras. This flexibility is facilitated by the joint design of camera optics and computational processing of the recorded data, a concept that has the potential to transform both photography and imaging science.

Existing approaches to light field capture can be divided into four categories: (a) camera arrays [Wilburn et al. 2005; Georgiev et al. 2008; Taguchi et al. 2010] (b) micro-lens arrays on the sensor [Adelson and Wang 1992; Ng et al. 2005; Lytro 2012] (c) attenuation masks in front of the sensor [Ives 1903; Lippmann 1908; Veeraraghavan et al. 2007; Lanman et al. 2008; Ihrke et al. 2010], and (d) CMOS integrated angle-sensitive pixels [Wang et al. 2011; Sivaramakrishnan et al. 2011]. While the technologies used for capturing light fields varies significantly between the four categories, they all share a common limitation that significantly hampers their widespread adoption: spatial resolution is sacrificed for a gain in extra angular resolution. This resolution tradeoff is fixed in the optical design and represents one of the main limitations of all existing light field camera designs. In practice, angular resolution required for typical applications such as synthetic refocus varies between 7×7 to 14×14; the image resolution is reduced by a factor of $49 - 196$, turning even a modern 9 megapixel (MP) (e.g., 3000×3000 px) sensor image into a measly 430×430 photograph or a 215×215 thumbnail. This clearly is a huge handicap and has resulted in widespread interest in light field super resolution techniques [Bishop et al. 2009; Lumsdaine and Georgiev 2009; Georgiev and Lumsdaine 2010] for hallucinating the lost plenoptic resolution by employing prior knowledge about the structure of the light field.

In this paper, we address the problem of designing a resolution-preserving light field camera that overcomes conventional limits through incoherent random projections using optical attenuation masks combined compressive computational reconstruction.

1.1 Contributions

We explore joint optical light attenuation via incoherent projections of the light field using attenuation masks and compressive computational reconstructions. The latter are demonstrated to benefit from learned dictionaries that capture the essential building blocks of natural light fields. The proposed approach overcomes traditional resolution tradeoffs. Specifically, our contributions include:

- We introduce a new approach to capturing compressive sensing of light fields through attenuation masks that are mounted at a slight offset to a sensor image. The measurements are incoherent projections of the incident light field on the sensor image.
- We propose a resolution-preserving light field reconstruction approach. Using sparse reconstruction routines, we show how to overcome traditional resolution tradeoffs in plenoptic cameras.
- We explore the space of high-dimensional basis functions and demonstrate learned, overcomplete dictionaries to best repre-

Figure 2: *Reconstructed and refocus scene showing two books.*

Approach	Image Resolution	Moving Scenes	Optical Complexity	Computational Cost	Light Transmission
Integral Imaging	medium*	yes	medium	low	high
Mask-based	low	yes	low	medium	medium
Time-sequential	high	no	low	low	high
Camera Array	high	yes	high	medium	high
Compressive (ours)	high	yes	low	high	medium

Figure 3: *Comparing benefits of a variety of light field acquisition approaches. Existing technologies either reduce the image resolution or the optical complexity of the system to capture a dynamic light field. We propose a new resolution-preserving light field camera architecture that overcomes many of the current technological limitations. The asterisk denotes previous attempts to light field super-resolution.*

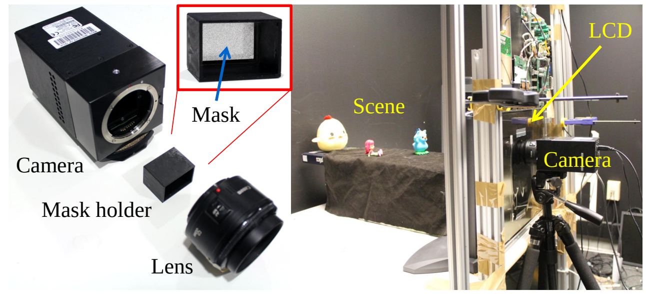
sent light fields in a sparse manner. These dictionaries capture the essential building blocks of natural light fields and allow for robust reconstruction routines.

- We derive theoretical bounds of several aspects the proposed camera design, including depth of field and depth-dependent reconstruction quality.
- We build a compressive light field camera prototype. The proposed reconstruction approach is demonstrated to successfully recover light fields from the captured data; we detail calibration routines and validate the data using synthetic refocus of the reconstructed light fields.

1.2 Overview of Benefits and Limitations

Inherently, a mask-based design offers several advantages over refractive optical elements placed on the sensor. Attenuating masks are less costly than microlenses, more robust to misalignment, and avoid refractive errors such as spherical and chromatic aberrations. Furthermore, the optical parameters of lenslets have to match the main lens aperture [Ng et al. 2005], whereas our mask-based approach is more flexible in supporting varying main camera lenses. The proposed compressive camera design allows for a significant increase in image resolution as compared to both lenslet-based systems and previously proposed mask cameras for in-focus image regions as well as refocused parts of the scene. The key advantage of our approach is the use of natural light field statistics learned from datasets as overcomplete dictionaries. While some previous work has followed similar ideas (e.g., [Bishop et al. 2009]), the employed lenslet arrays optically filter out the visual information that is essential for a successful compressive light field reconstruction.

As most light field cameras, our system requires modifications of conventional camera hardware. Although attenuation masks preserve more visual information than lenslet arrays in the captured data, the overall light transmission is reduced by about 50%. The proposed reconstruction requires an overcomplete dictionary that



(a) (b)

Figure 4: *Photographs showing the prototype setup. (a) Exploded view of our mask-based light field camera. The inset shows a printed random mask pattern attached to the mask holder. (b) Experimental setup where we placed an LCD in front of the camera to sample incoming light fields. We moved a pinhole on the LCD to calibrate mask modulation and also to capture light fields for dictionaries. We reconstructed new light fields from a single-shot with the LCD showing a square aperture.*

captures the essence of natural light fields; this is a one-time preprocessing step and we expect improvements of our current dictionaries with an increasing amount of available lights fields, for instance captured with Lytro cameras. Finally, the increase in image resolution comes at the cost of increased computational demands. Though theoretically polynomial in time, sparse reconstructions practically require computing times ranging from a few minutes to hours for a single full-resolution sensor image on a desktop PC.

2 Related Work

Light Field Cameras: Light field acquisition has been an active area of research; more than a century ago, Frederic Ives [Ives 1903] and Gabriel Lippmann [Lippmann 1908] realized that the light field inside a camera can be captured by placing pinhole or lenslet arrays at a slight offset in front of the sensor. Within the last few years, lenslet-based systems have been integrated into digital cameras [Adelson and Wang 1992; Ng et al. 2005] and are now commercially available [Lytro 2012]. The light-attenuating codes used in mask-based systems have become much more light efficient as compared to pinhole arrays [Veeraraghavan et al. 2007; Lanman et al. 2008; Ihrke et al. 2010]. All of these approaches require modifications of the camera hardware; a popular alternative is time-sequential image capture using a moving camera [Levoy and Hanrahan 1996; Davis et al. 2012] or programmable camera apertures [Liang et al. 2008]. To allow for the acquisition of dynamic scenes, camera arrays have been employed as well [Wilburn et al. 2005]. We propose a novel, compressive approach to light field acquisition; our technique is similar in spirit to single camera, mask-based approaches but significantly increases image resolution by using compressive sensing reconstructions in combination with optimized mask patterns.

Traditional Nyquist Sampling: Traditional sampling theory is based heavily on the Shannon-Nyquist sampling theorem which states that a signal x that is band-limited to W Hz is determined completely by uniform discrete samples of the signal provided that the sampling rate is greater than $2W$. Modern sensors whether they are audio, or image sensors and more recently light field imagers are all attempting to capture discrete samples of the underlying signal. In order to satisfy the Shannon-Nyqusit theorem, these sensor architectures typically have prefiltering (or anti-aliasing) that ensures that the incoming signal bandwidth is less than half the sampling rate of the sensors. There is unfortunately a price that we pay because of this anti-aliasing: it ensures that high frequency detail (that is larger than half the sampling rate) is irreversibly lost. In the context of traditional image sensors, the finite area of the pixels in the detector array act as optical anti-aliasing filters. In the case of the various light field camera architectures, the finite sized aperture of the microlens array [Adelson and Wang 1992; Ng et al. 2005; Lytro 2012] and/or the finite size of the pixels in the detector act as anti-aliasing filters irrevocably reducing the bandwidth of these systems. Recently, light field super resolution techniques [Bishop et al. 2009; Lumsdaine and Georgiev 2009; Georgiev and Lumsdaine 2010] have proposed methods for hallucinating the lost plenoptic resolution by employing prior knowledge about the structure of the light field. In this paper, we take a radically different approach and draw inspiration from recent advances in sampling theory to explicitly recover light fields from a single modulated captured image. Since, there is no angular anti-aliasing in our camera, subsequently the high resolution information is never suppressed and this allows us to recover details both in texture and in specular and non-lambertian parts of the light field.

Compressive Sampling and Dictionary Learning: Recent advances in sampling theory have shown that if a signal $x \in R^N$ can be represented as $k-$sparse in some basis D (usually called a Dictionary), then the signal can be robustly and accurately recovered from $O(klog(\frac{N}{k}))$ samples instead of the N samples required using traditional Shannon-Nyquist techniques. Compressive sensing [Candès et al. 2006; Candès and Tao 2006; Donoho 2006a] enables reconstruction of such sparse signals from under-sampled linear measurements typically using techniques from convex optimization. The rich image processing and signal processing literature has yielded a huge number of data independent basis such as wavelets, DCT, and Fourier in which images and other such signals have been shown to be sparse. We show that learned dictionaries provide sparser representations of natural light fields than conventional bases.

Recently, it has been shown that learning and adapting dictionaries to the specific rich geometric structure of the data results in significant performance improvements over traditional data independent dictionaries. Several algorithms [Kreutz-Delgado et al. 2003; Mairal et al. 2008; Kreutz-Delgado and Rao 2000; Aharon et al. 2005] for learning such dictionaries from sample datasets have been proposed, most of them iterating between a sparse approximation and a model fitting step. We rely on the advances in dictionary learning and learn patch based dictionaries for light field data. Unlike most light field analysis and super-resolution techniques [Bishop et al. 2009; Lumsdaine and Georgiev 2009; Georgiev and Lumsdaine 2010; Levin and Durand 2010], we do not assume that the materials in the scene are lambertian. Instead, we learn a patch based dictionary for light fields from available light field data and this allows us to tackle more complex optical phenomena such as translucency and specularities.

Compressive Light Field Acquisition Broadly speaking, the idea of performing compressive light field acquisition has been attempted in the past. It could be argued that approaches to perform light field super-resolution [Bishop et al. 2009; Lumsdaine and Georgiev 2009; Georgiev and Lumsdaine 2010] are compressive light field rendering methods. Unfortunately, in these examples since the microlens arrays act as anti-aliasing filters reducing the spatial resolution of the incoming radiance function before being captured on the sensor, these approaches are inherently limited in their applicability. Recently, Babacan et al. [2009] showed that reasonable 7×7 light field reconstructions can be obtained from about 7 images acquired with random coded apertures. Similarly, Ashok et al. [2010] showed that multiple images acquired with coded apertures placed either at the aperture plane or in front of micro-lens arrays allows us to reduce the number of measurements required for acquiring full resolution light fields. Unfortunately, like all other multi-image based methods such techniques cannot handle dynamic scenes. In contrast, our technique is a single-shot, single image technique and so it has the potential to handle fast moving and dynamic scenes with appropriate short exposure imaging. Further, most of the existing results in compressive light field acquisition have been predominantly in simulations. Here, we build a working prototype of our compressive light field imager. Finally, we also perform theoretical analyses of the various designs and show that our compressive light field camera has better spatial frequency support and depth of field properties.

3 Light Field Sensing and Reconstruction

This section presents a framework for compressive light field sensing. First, we introduce a mathematical model describing how a light field is sensed, through a number of light attenuating masks, with multiple photographs. Second, we show how this general image formation represents the measurement matrix Φ in general compressive sensing formulations; we briefly review these formulations along with their properties and fundamental limitations. Third, we introduce an approach to capture the essence of natural light fields, as a mathematical prior, in a learned, overcomplete dictionary and interprete the structure of the fundamental light field elements captured in the learned dictionaries. We conclude by showing that natural four-dimensional light fields are more sparse in this adaptive basis than in generic bases often used in compressive sensing reconstructions. The mathematical formulations in this section are derived for the 2D spatio-angular flatland case with straightforward extensions to the full 4D light field space.

3.1 Light Field Sensing

Compressive plenoptic cameras comprise a conventional camera with lens and sensor as well as a stack of light attenuating masks that optically modulate the four-dimensional light field before it reaches the two-dimensional sensor. This design is illustrated in Figure 5; for full generality, we assume that multiple photographs can be captured with dynamically changing mask patterns.

The image captured by a conventional sensor $i\left(x\right)$ is a projection from spatio-angular light field space along the angular dimension:

$$i\left(x\right) = \int_{\mathcal{V}} l\left(x, \nu\right) d\nu. \tag{1}$$

The light field is denoted as $l\left(x, \nu\right)$. We adopt a two-plane parameterization [Levoy and Hanrahan 1996], where x is the spatial dimension on the sensor plane and ν denotes the position on the aperture plane at distance d (see Fig. 5). A single attenuation mask with pattern $f(\xi)$ modulates the light field before the sensor integrates over the angular dimension as

$$i\left(x\right) = \int_{\mathcal{V}} l\left(x, \nu\right) f\left(x + \frac{d_l}{d}\nu\right) d\nu. \tag{2}$$

In this formulation, d_l is the distance between sensor and mask. Mounting a stack of N light-attenuating masks $f^{(n)}, n = 1 \ldots N$ at distances d_n from the sensor changes the optical image formation to

$$i\left(x\right) = \int_{\mathcal{V}} l\left(x, \nu\right) \prod_{n=1}^{N} f^{(n)}\left(x + \frac{d_n}{d}\nu\right) d\nu. \tag{3}$$

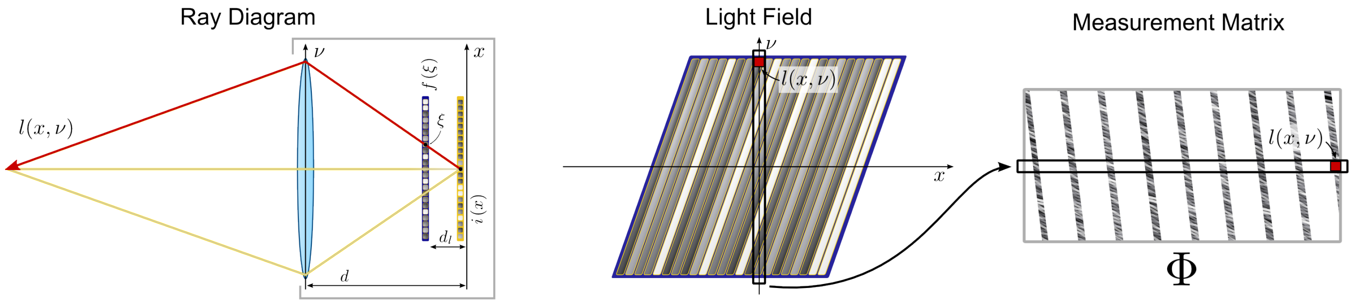


Figure 5: *Illustration of ray optics, light field modulation through coded attenuation masks, and incoherent projection matrix. Left: ray diagram illustrating the optical setup. One or more coded attenuation masks are mounted between camera sensor and aperture. Center: the mask patterns modulate a four-dimensional light field (only two dimensions shown) before the camera sensor optically integrates over the angular dimensions. Right: in a discretized form, the image formation can be expressed as a sparse, random projection matrix used in a compressive reconstruction framework.*

Again, d is the distance between sensor plane and aperture plane. For full generality, we also consider taking M photographs with mask patterns $f_m^{(n)}$ that change for each shot $m = 1 \dots M$ but stay constant throughout the exposure time of each photo:

$$i_m(x) = \int_{\mathcal{V}} l(x,\nu) \prod_{n=1}^{N} f_m^{(n)} \left(x + \frac{d_n}{d}\nu \right) d\nu. \tag{4}$$

This projection can be expressed, in a discretized form, as a matrix-vector multiplication:

$$\mathbf{i} = \mathbf{\Phi l}, \quad \Phi_{ij} = \prod_{n=1}^{N} f_{[i]_m}^{(n)} \left([i]^x + \frac{d_n}{d}[j]^\nu \right), \tag{5}$$

where all M sensor images are vectorized as $\mathbf{i}$ and the light field in its vectorized form is $\mathbf{l}$. A row in the projection matrix Φ corresponds to the contributions of all light field rays to a single sensor pixels; a column to a single light field ray and its contribution to each sensor pixel. The matrix row index $[i]_m^x$ corresponds to the order of sensor image vectorization—row or column major— and $[j]^\nu$ is the matrix column index for a particular light field ray.

A ray diagram illustrating the optical setup is shown in Figure 5 (left) with the corresponding interpretation in light field space shown in the central column of Figure 5. Assuming that each mask pattern attenuates rays incident on that plane equally for all incident directions, each of these patterns corresponds to a sheared copy of the corresponding pattern with constant values along the diagonals. The corresponding, discretized projection matrix Φ is also visualized. In the following, this notation makes it convenient to apply standard signal processing notation of the compressive light field reconstruction.

3.2 Compressive Light Field Reconstruction

We begin by providing a brief introduction to compressive sensing and then return to the problem of light field capture via compressive sensing.

A brief tour of compressive sensing: Compressive sensing [Candès et al. 2006; Candès and Tao 2006; Donoho 2006a] enables reconstruction of sparse signals from under-sampled linear measurements. A vector $\mathbf{s}$ is termed K-sparse if it has at most K non-zero components, or equivalently, if $\|\mathbf{s}\|_0 \leq K$, where $\|\cdot\|_0$ is the ℓ_0 norm or the number of non-zero components. Consider a signal (in our example the light field $\mathbf{l}$) $\mathbf{l} \in \mathbb{R}^N$, which is sparse in a (possibly overcomplete) basis Ψ (a matrix of size $N \times D$). Since the light field $\mathbf{l}$ is k-sparse in Ψ, we can write $\mathbf{l} = \Psi\mathbf{s}$, where $\mathbf{s} \in \mathbb{R}^D$, and $\|\mathbf{s}\|_0 \leq K$. Traditional examples of popular sparsifying basis Ψ for images includes DCT and wavelets. While 4D extensions of such popular basis functions may work reasonably well for light fields, here we learn a data-dependent adaptive dictionary that represents the geometric structure of light field data better. The details regarding the dictionary learning are described in Section 3.3. For now, we will assume that Ψ is known.

The main problem of interest is that of sensing the signal $\mathbf{l}$ from linear measurements $\mathbf{i} = \mathbf{\Phi l}$. With no additional knowledge about $\mathbf{l}$, N linear measurements of $\mathbf{l}$ are required to form an invertible linear system. The theory of compressive sensing shows that it is possible to reconstruct $\mathbf{l}$ from M linear measurements even when $M \ll N$ by exploiting the sparsity of $\mathbf{s}$ in the basis Ψ.

Consider the measurements obtained using the mask based light field camera design described in the previous section. The measurement vector $\mathbf{i} \in \mathbb{R}^M$ obtained using such a compressive light field camera can be represented as

$$\mathbf{i} = \Phi\mathbf{l} + e = \Phi\Psi\mathbf{s} + e = \Theta\mathbf{s} + e \tag{6}$$

where e is the measurement noise and $\Theta = \Phi\Psi$. The components of the measurement vector $\mathbf{i}$ are called the *compressive measurements* or compressive samples. For $M < N$, estimating $\mathbf{l}$ from the linear measurements is an ill-conditioned problem. However, when $\mathbf{l}$ is K sparse in the basis Ψ, then CS enables recovery of $\mathbf{s}$ (or alternatively, $\mathbf{l}$, since $\mathbf{l} = \Psi\mathbf{s}$) from $M = O(K\log(N/K))$ measurements, for certain classes of matrices Θ. The guarantees on the recovery of signals extend to the case when $\mathbf{s}$ is not exactly sparse but compressible. A signal is termed compressible if its sorted transform coefficients delay according to power-law, i.e, the sorted coefficient of $\mathbf{s}$ decay rapidly in magnitude [Haupt and Nowak 2006].

Signal recovery: Estimating K-sparse vectors that satisfy the measurement equation of (6) can be formulated as the following ℓ_0 optimization problem:

$$(P0): \ \min \|\mathbf{s}\|_0 \text{ s.t. } \|\mathbf{i} - \Phi\Psi\mathbf{s}\|_2 \leq \epsilon. \tag{7}$$

with ϵ being a bound for the measurement noise e in (6). While this is a NP-hard problem in general, the equivalence between ℓ_0 and ℓ_1 norm for such systems [Donoho 2006b] allows us to reformulate the problem as one of ℓ_1 norm minimization.

$$(P1): \ \widehat{\mathbf{s}} = \arg\min \|\mathbf{s}\|_1 \text{ s.t. } \|\mathbf{i} - \Phi\Psi\mathbf{s}\| \leq \epsilon \tag{8}$$

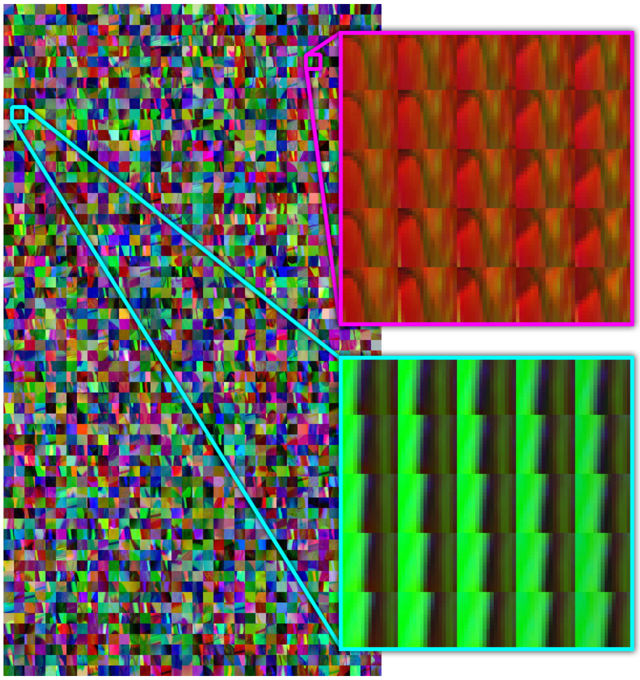


Figure 6: *Learned dictionaries capture the essential building blocks of natural light fields. The dictionary is a collection of small four-dimensional patches (closeups) representing the basic spatio-angular building blocks of a large light field database. The mosaic shows of the central views of light field patches in a dictionary, whereas the closeups magnify two 4D light field patches. Both horizontal and vertical parallax is clearly visible in structures that slightly move over the different viewpoints in each patch.*

It can be shown that, when $M = O(K \log(N/K))$, the solution to the $(P1)$ — $\widehat{\mathbf{s}}$ — is, with overwhelming probability, the K-sparse solution to $(P0)$. In particular, the estimation error can be bounded as follows:

$$||\mathbf{s} - \widehat{\mathbf{s}}||_2 \leq C_0 ||\mathbf{s} - \mathbf{s}_K|| / \sqrt{K} + c_1 \epsilon \tag{9}$$

where $\mathbf{s}_K$ is the best K-sparse approximation of $\mathbf{s}$.

There exist a wide range of algorithms that solve $(P1)$ to various approximations or reformulations [Candès and Tao 2005; Tibshirani 1996]. One class of algorithms model $(P1)$ as a convex problem, and recast it as a linear program (LP) or a second order cone program (SOCP) for which there exist efficient numerical techniques. Another class of algorithms employ greedy methods [Needell and Tropp 2009] which can potentially incorporate other problem-specific properties such as structured supports [Baraniuk et al. 2010]. It has been shown that for overcomplete basis such as dictionaries reweighted L1, which solves several sequential L1 minimization problems each using weights computed from the solution of the previous problem provides with the best solution for $(P1)$.

3.3 Light Field Dictionaries

3.3.1 Learning Overcomplete Dictionaries

In order to effectively apply and exploit principles of sparse representations and compressive sensing, we need to find a dictionary Ψ in which the patches from light fields are sparse. One can possible use non-adaptive dictionaries such as DCT, wavelet or Fourier bases (or a combination of them), but these dictionaries do not model the specific geometry of light field patches. Thus, we learn the dictionary from light field patches themselves. The traditional dictionary learning algorithms such as K-SVD [Aharon et al. 2005] and Focuss [Kreutz-Delgado and Rao 2000; Kreutz-Delgado et al. 2003] are batch methods and hence are not suitable for learning light field patches as the patches are very high-dimensional (of the order 6000). Thus, we use the online dictionary learning approach proposed in [Mairal et al. 2008] to learn our dictionary.For the sake of completeness, we provide a very brief description of the algorithm.

Given a finite training set of light field patches, say $\mathbf{L} = [\mathbf{l_1}, \mathbf{l_2}, ..., \mathbf{l_n}]$, the dictionary learning problem can be formulated as jointly optimizing the dictionary Ψ and the coefficient vectors $\mathbf{S} = [\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_n}]$:

$$\min_{\Psi, \mathbf{S}} \sum_{j=1}^{n} (||\mathbf{l_j} - \Psi \mathbf{s_j}||_2^2 + \lambda ||\mathbf{s_j}||_1)/2n \tag{10}$$

The above equation describes the learning process as the joint optimization problem with respect to the dictionary and the coefficients $\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_n}$. Note that the above optimization problem is a non-convex problem (because of the coupling between Ψ and the coefficients $\mathbf{S}$). However, this is a bi-convex problem, i.e., if we fix one of the variables (say Ψ), then the problem is convex in the other variable ($\mathbf{S}$). The online dictionary approach uses the stochastic gradient algorithm to solve the problem. Once we learn the dictionary Ψ, any new light field patch can be described as a linear combination of the basis elements of the dictionary. Figure 6 shows some of the basis elements of our learned dictionary. It is clear from the figure that the learned dictionary captured the specific structure of the light field data.

3.3.2 Reconstructing Light Fields using Dictionaries

During reconstruction, we extract patches $\mathbf{i_j}, j = 1, 2, ..., m$ from the captured image and reconstruct the corresponding light field patches $\mathbf{l_j}$. The light field patches can in turn be expressed as $\mathbf{l_j} = \Psi \mathbf{s_j}$, where $\mathbf{s_j}$ are the sparse coefficient vectors. To obtain the sparse coeffcent vectors $\mathbf{s_j}$ (and hence $\mathbf{l_j}$), we use use the reweighted L1-norm minimization algorithm [Emmanuel J. Cands and Boyd 2008], which has been shown to have a superior performance than the standard L1-norm algorithm (basis pursuit). The reweighted L1-norm minimization solves the following problem:

$$\min_{\mathbf{s_j}} ||\mathbf{W} \mathbf{s_j}||_1 \quad \text{s.t.} \quad ||\mathbf{i_j} - \mathbf{\Phi} \mathbf{\Psi} \mathbf{s_j}||_2 \leq \epsilon, \tag{11}$$

where $\mathbf{W}$ is a diagonal matrix with the diagonal elements being the weights. In the first few iterations, the largest signal coefficients are identified. The weighting matrix is then updated with these values for identifying the remaining small but non-zero coefficients.

3.3.3 Evaluating Light Field Sparsity

In this section, we evaluate the sparsity of light fields in a variety of commonly used transforms and the over-complete dictionary described in Section 3.3. For conventional transforms, including the Fourer basis (FFT), wavelets, and the dicrete cosine transform (DCT), sparsity of a given light field can be quantified by peak-signal-to-noise ratio (PSNR). For this purpose, the light field is approximated by its K largest coefficients in that basis. Figure 7 plots the PSNR of a synthetic light field for an increasing number of sparse coefficients in a variety of transforms. The compression ratio is given as the ratio between K and the total number of coefficients.

In addition to these conventional transforms, which are all evaluated in their full four-dimensional form, we also plot the sparsity of the same light field in a learned dictionary. Please note that the training set necessary to compute the dictionary does not include the test case. Evaluating the light field sparsity in the dictionary is slightly more involved than for the conventional transforms. In this case, an optimization problem (Eq. 8) has to be solved explicitly to determine the K dictionary elements that best approximate the original light field. Figure 7 plots the sparsity of the test light field in the learned dictionary; this choice of a sparsity basis yields a gain in PSNR by about 5-10 dB as compared to conventional basis.

The conclusion of this experiment is that bases such as the Fourier transform provide powerful tools for theoretically analyzing computational cameras and upper bounds on their performance (e.g., [Levin et al. 2009]) but for the case of compressive light field sensing, learned dictionaries capturing the essence of natural light fields provide more robust tools for practical computation.

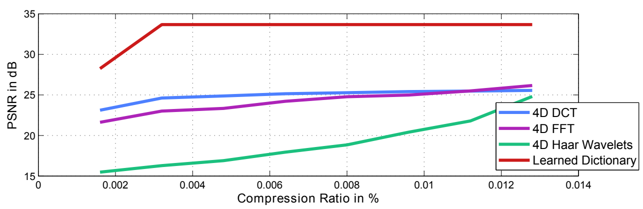


Figure 7: *Sparsity of a light field, measured in PSNR, is evaluated for conventional bases (4D DCT, 4D FFT, 4D Haar wavelets) and a dictionary; the compression ratio is the number of sparse coefficients divided by the total number of basis coefficients. In all tested cases, dictionaries lead to a significant improvement in PSNR, demonstrating that these are usually a better choice for compressive light field reconstruction than conventional transforms.*

4 Analysis

While general compressive reconstructions combined with overcomplete dictionaries, as described in the previous section, are powerful tools for practical computations, deriving analytical performance bounds is difficult. One of the most interesting attributes characterizing a light field camera is the depth of field in which synthetic refocus can be performed. A common approach to such an analysis is the evaluation of the reconstruction performance of a textured diffuse plane at a distance to the focal plane. A major advantage of these assumptions, commonly used for depth of field analysis (e.g., [Levin et al. 2009]), is that the dimensionality of the analysis reduces to three, instead of four dimensions. In the following, we show that Gaussian Mixture Models can analytically describe this special case and be used to derive upper bounds on the depth-dependent reconstruction performance.

Gaussian Mixture Models (GMMs) make a few simplifying assumptions: (1) the scene is lambertian and (2) all objects are within a depth range of $[-tD, tD]$ around the focal plane of the camera, where D is depth of field of traditional camera and $t = 20$. Under these assumptions, which are perfectly valid for the above described depth of field analysis, we can learn a GMM prior for the light field and then use the GMM model to analytically characterize the compressive light field camera. We use the 'minimum mean square error' (MMSE) for GMM priors as a metric to characterize the performance of our camera.

The GMM prior consists of a mixture of Gaussian priors; consider the i^th mixture component $P_i(x) = \mathcal{N}(m_i, \Sigma_i)$, where m_i and Σ_i are the mean and covariance matrix respectively. In practice, we learn separate Gaussian models m_i,Σ_i for a discrete set of sampled depths within the depth range [-tD, tD]. For each depth, we take a set of textures (canonical images such as Lenna, Barbara etc), and place these images at the corresponding depth and generate light fields corresponding to these scenes. We then learn the Gaussian model parameters m_i,Σ_i for this particular depth. We do this over a range of depths and this process results in a GMM. In the following paragraphs, we first present the expression of MMSE for a single Gaussian prior and then for the GMM prior.

Since the compressive camera is a linear system, we can write it as $y = Hx + n$, where x is the unknown light field signal, y is the observed image and n is the noise. If we assume the noise n to be Gaussian $P(n) = \mathcal{N}(0, \Sigma_n)$, then the observation likelihood $P(y|x) = \mathcal{N}(Hx, \Sigma_n)$ is Gaussian. For Gaussian prior $P_i(x) = \mathcal{N}(m_i, \Sigma_i)$, the posterior distribution $P_i(x|y)$ is also Gaussian distributed and the mean square error $mmse_i(H)$ is given by [Kay 1993]:

$$mmse_i(H) = trace(\Sigma_i) - trace(\Sigma_i H^T (H\Sigma_i H^T + \Sigma_n)^{-1} H\Sigma_i). \tag{12}$$

It can be shown that, for GMM prior $P(x) = \sum_{i=1}^{m} \alpha_i P_i(x)$ (where $\alpha_i, i = 1, 2, ..., m$ are the mixture weights) and Gaussian likelihood $P(y|x) = \mathcal{N}(Hx, \Sigma_n)$, the posterior distribution $P(x|y)$ is also a GMM (see [Flam et al. 2011]). The MMSE can be lower bounded as follows [Flam et al. 2011; Anon. 2012]:

$$mmse(H) \geq \sum_{i=1}^{m} \alpha_i mmse_i(H), \tag{13}$$

where, $mmse_i(H)$ are the MMSE for the individual Gaussian priors (12). We use this lower bound on MMSE to charaterize the performance of our camera. Using this expression for MMSE and the the average signal power (which can be computed from the GMM prior $P(x)$), we obtain the expected system SNR. For details regarding the derivation and the expression please see [Flam et al. 2011].

4.1 Depth-Dependent Reconstruction Performance

We evaluated the reconstructed SNR for four different cameras keeping the number of sensor pixels constant. The four different cameras we considered in our analysis are: (1) Traditional Camera (2) Pinhole Array based Light Field Camera (3) Micro-lens array based light field Camera (Lytro) and (4) Our compressive Light field camera with GMM prior. For the existing 2 light field imaging architectures (Pinhole array and micro-lens), the reconstructed light field is usually lower resolution. We then use PCA to upsample these light fields to obtain full-resolution light fields. For our proposed compressive light field camera, we the mixing matrix H corresponding to the mask used. We then use the GMM model that we learned $\{m_i, \Sigma_i\}$ and evaluate the lower bound on the mmse given by Equation 13. The results are shown in Figure 8.

When the scene is at the plane of focus of the traditional camera, it is clear that a traditional camera outperforms all other light field cameras. Notice that all the presented light field cameras have a reconstruction performance that is better than a traditional camera, as the scene moves away from the plane of focus. It is also clear that our compressive light field camera design significantly outperforms both micro-lens array based [Adelson and Wang 1992; Ng et al. 2005; Lytro 2012] and the pinhole [Ives 1903; Lippmann 1908; Veeraraghavan et al. 2007; Lanman et al. 2008; Ihrke et al. 2010] based designs for acquiring the light field. Figure 8 also shows that the depth of field of our compressive light field camera is larger than that of other alternatives.

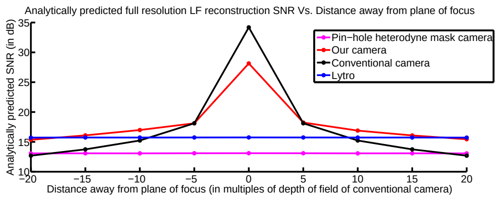


Figure 8: *Analytical estimates (using a GMM model) of the reconstruction SNR for varying light field camera architectures. At the plane of focus, traditional camera provides the best performance. As you move away from plane of focus both Lytro and our architecture provides better performance.*

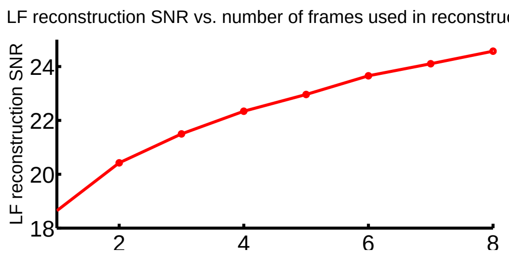


Figure 9: *Analytical estimates of reconstruction SNR (using GMM model) for varying number of captured images.*

4.2 Analysis of Multi-Shot Camera Sequences

If the mask is implemented using an electronically controllable spatial light modulator, this would allow us to acquire multiple frames with different masks. If the scene is static or slow moving during the acquisition time, then multiple images can be used to reconstruct the light field. Since each successive frame provides new additional information about the structure of light field this would presumably improve reconstruction performance. We tested this thoroughly in simulation by varying the number of frames acquired from one to eight using the analytical expression in Equation 13. For the k^th frame, we use a different mask m_k and obtain the corresponding mixing matrix H_k. The combined effect of all these frames is equivalent to stacking these mixing matrices to obtain an effective mixing matrix $H = [H_1; H_2; H_3; ...; H_K]$, where the symbol ; represents vertical concatenation. The results are shown in Figure 9, clearly showing that significant benefit is obtained by increasing the number of frames used during reconstruction.

5 Implementation and Assessment

5.1 Implementation

5.1.1 Software

As described in Section 3.3.1, we use the implementation of online sparse coding [Mairal et al. 2009] algorithm available as a part of SPAMS(Sparse Modeling Software) package. Dictionaries with varying patch sizes from $8\times8\times3\times3$ to $16\times16\times5\times5$ are learned. Learned basis are ten times overcomplete for patches with angular resolution of 5×5. For lower angular resolution of 3×3 we are able to learn dictionaries that are hundred times overcomplete. We find in simulation that due to high coherency of light fields a coherency factor $10\times$ induces enough sparsity for a compressive reconstruction. For our reconstructions on real scenes, it takes about 6 hours to learn a dictionary with a patch size of $8\times8\times3\times3$ overcomplete by $10\times$ resulting in about 6000 dictionary elements.

We used POVRAY a freely available raytracing software to render several synthetic light fields. We divided the synthetic light fields we rendered into two non-overlapping sets – a training set and a test set. Patches from the training set were used to train the dictionary learning algorithm, while simulation experiments were performed on the test set of light fields. An example light field from the test set is the dice dataset shown in Figure 11. Our reconstruction algorithm described in Section 3.3.2 leverages on the software base made available by NESTA [Becker et al. 2009] that implements reweighted L1 optimizations. All implementations of dictionary learning and L1 minimization are done in MATLAB. The reconstruction algorithm takes about four hours for reconstructing a $256\times256\times5\times5$ light field on a desktop personal computer.

5.1.2 Hardware

Figure 4 (a) shows our prototype compressive light field camera. We fabricated a mask holder that fits into the sensor housing of a Lumenera Lw11059 camera, and attached a film with a random mask pattern, where each dot had an intensity uniformly drawn from [0,1] range. As the printer guaranteed 25 μm resolution, we conservatively picked a mask resolution of 50 μm, which roughly corresponded to 6×6 pixels on the sensor. We therefore downsampled the sensor image by 6, and cropped out the center 320×240 region for light field reconstruction in order to avoid mask holder reflection and vignetting. The distance between the mask and the sensor was 1.6mm. A Canon EF 50mm f/1.8 II lens was used and focused at a distance of 35cm.

Calibration: In order to be able to perform the reconstruction, we need to know the mixing matrix $\mathbf{\Phi}$. Since the mask is only approximately positioned at about 1.6 mm away from the sensor, it becomes necessary to calibrate and measure the effective mixing matrix $\mathbf{\Phi}$. To do this, we placed an LCD in front of the camera as shown in Figure 4 (b) to obtain control over angular samples of incoming light fields. We used the full aperture size of the lens (8×8 mm) and divided it into 3×3 sub-apertures. For calibration, we placed a monitor displaying a white image at the plane of focus (35 cm depth), and captured that white image modulated by the mask for each sub-aperture. We also normalized each of these images by an image captured without the mask in order to obtain the effective mixing matrix. Once the system is calibrated, i.e., we have $\mathbf{\Phi}$ measured, then we can perform reconstruction on real scenes from a single captured image. Note that the calibration process needs to be done only once and need not be repeated for every dataset.

5.2 Experimental Results

This section assesses the quality of reconstructed results for four examples: the teaser scene including a number of diffuse objects with specularities, two book scenes (Figs. 2, 10), and a synthetic scene that contains translucencies, occlusions, specularities and other challenging illumination effects (Fig. 11).

The scene in Figure 1 contains three objects arranged on three distinct distances from the camera. The sensor image shows the effect of the random attenuation pattern created by the mask in front of the camera. Several views of the reconstructed light field (top row) are visualized along with a small mosaic showing all 3×3 reconstructed light field views (top right). Parallax is visible, as are specularities on the bear's eyes and occlusions between the eye of the yellow bird and the blue bear's hand. The light field can be refocused by shearing the views and averaging them [Ng 2005] (bottom row). From left to right, we see the car in the foreground focused, then the blue bear, and finally the yellow bird in the background.

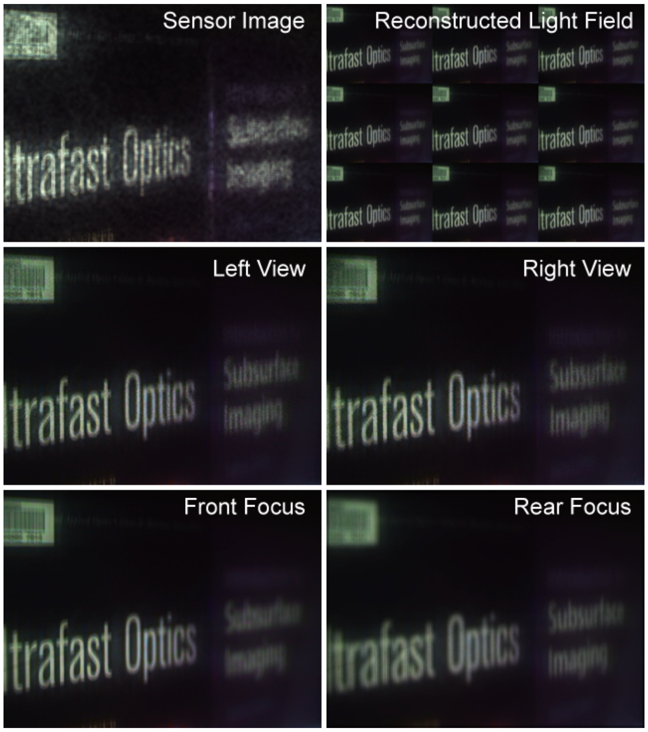


Figure 10: *Light field reconstruction from prototype camera. The sensor image (upper left) is optically modulated prior to capture by a random attenuation mask; using the algorithms described in this paper, we reconstruct the light field (upper right). While the individual views of the light field (center row) exhibit slight reconstruction noise, these artifacts are barely visible in the refocused images (lower row).*

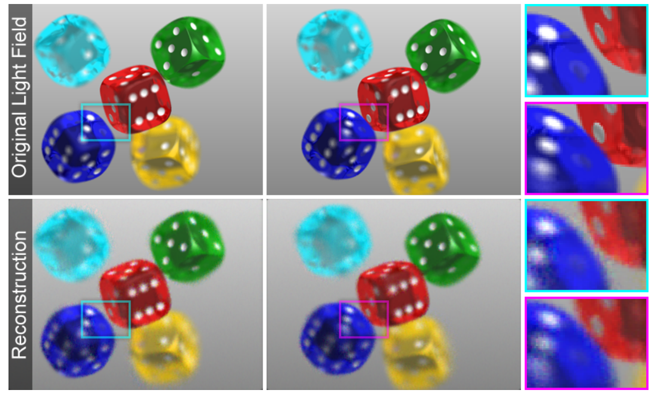


Figure 11: *As seen in this simulated reconstruction, our algorithm handles occlusions and translucencies as well as specularities (Fig. 1) among other effects not captured by previous light field super-resolution approaches.*

Figure 2 shows a refocused scene containing two books at distinct distances in front of the camera. The photograph on the left is focused on the front book, while the right image is focused on the rear. As visible in these examples, the limited angular resolution of the reconstructions, in this case 3×3 views, introduces a limited depth of field for each view corresponding to a finite-sized sub-aperture. The image resolution in the refocused images is limited to the depth of field of the individual views.

A single book, slanted in depth, is shown in Figure 10. In addition to the captured sensor image (top left) we show a mosaic of the reconstructed light field (top right), two of the light field view (center row), and two images with synthetic refocus applied (bottom row). While slight reconstruction artifacts in the light field views prevail, the refocus operation averages all of them and, hence, mitigating any such artifacts.

Finally, in Figure 11 we show a simulation using a povray rendered dataset. This result demonstrates that even challenging scenes with strong occlusions, specularities, and translucent objects can successfully be reconstructed with the proposed approach. Effects such as these are not handled directly using existing light field priors such as the dimensionality gap [Levin et al. 2009; Levin and Durand 2010].

6 Discussion

In summary, we present a novel approach to single-shot, resolution-preserving light field acquisition. Facilitated by the joint design of optical light modulation and compressive computational reconstruction, our approach has the potential to overcome one of the major limiting factors of current light field camera technology: the inherent resolution tradeoff. Our technique is the first to explore overcomplete dictionaries learned from a database of synthetic light fields; we show that these capture the essential building blocks of natural light fields and allow for sparser representations and higher-quality reconstructions as compared to conventional high-dimensional bases used in the compressive sensing literature. Using Gaussian Mixture Models, we derive upper bounds for the expected reconstruction quality of diffuse scenes at a varying distance to the focal plane; this analysis allows for intuitive interpretations of the camera's expected depth of field. Using a prototype camera, we demonstrate the practicality of our approach.

6.1 Benefits and Limitations

While humble in its initial image quality, we demonstrate the first compressive camera architecture that allows for compressive reconstructions of real world data. Full parallax, four-dimensional light fields are recovered from two-dimensional sensor image. One of the key insights of this paper is that mask-based camera designs offers more flexibility for processing recorded data as aliasing, which is critical for compressive reconstructions, is optically preserved. Light attenuating masks are less costly than high-quality refractive optical elements, more robust to misalignment, and avoid refractive errors such as spherical and chromatic aberrations. Furthermore, the optical parameters of lenslets mounted on the sensor have to match the main lens aperture [Ng et al. 2005], whereas our mask-based approach is more flexible in supporting varying main camera lenses. In theory, the proposed compressive camera design allows for a significant increase in image resolution as compared to both lenslet-based systems and previously proposed mask cameras for in-focus image regions as well as refocused parts of the scene. The key advantage of our approach is the use of overcomplete dictionaries that capture the essence of natural light fields and allow for robust sparse reconstructions.

The proposed systems has the potential to overcome resolution limits inherent in current plenoptic camera design; due to limited computational resources, current results demonstrate the concept at a reduced resolution. With the growing availability of cloud computing, we hope to significantly increase the size of the datasets we can practically process. Currently, processing times take about $1 - 2$ hours for a light field with moderate resolution (e.g., $256 \times 256 \times 5 \times 5$) on a standard workstation. Although mask-based camera designs have many advantages over lens arrays, they also reduce the optical light transmission. Random attenuation patterns, as used in our experiments, practically reduce the image brightness by half. Diffraction certainly limits the lower bound of mask pixel size. Finally, calibration of the capture setup is critical but only needs to be performed once as a pre-processing step.

6.2 Future Work

In the future, we plan to explore compressive acquisitions of the full plenoptic function, adding temporal and spectral light variation to the equation. While significantly increasing the dimensionality of the dictionary learning and reconstruction problem, we believe that exactly this increase in dimensionality will further improve compressibility and sparsity of the underlying signal. For this purpose, dynamically changing attenuations patterns and programmable spectral transmission as well as more efficient dictionary learning and reconstruction routines will have to be explored. Another avenue of future work is the exploration of content-adaptive sensing. Can optimal attenuation masks or, more generally, plenoptic sensing codes be derived for particular materials or different scene properties?

7 Conclusion

The proposed camera architecture is an integral step toward the "ultimate" camera, which can be argued to be a device capable of capturing the full plenoptic function, including spatial, angular, and temporal light variation as well as the color spectrum, at a high resolution with a single image. We believe that the joint design of camera optics and compressive computational processing of the recorded data is the key to facilitate next-generation camera technology; in combination with dictionary learning and reconstruction techniques discussed in this paper, compressive computational photography paves the road for practical exploitation of the correlations between the plenoptic dimensions—the future of plenoptic camera technology.

References

ADELSON, E., AND WANG, J. 1992. Single Lens Stereo with a Plenoptic Camera. *IEEE Trans. PAMI 14*, 2, 99–106.

AHARON, M., ELAD, M., AND BRUCKSTEIN, A. 2005. K-svd: Design of dictionaries for sparse representation. *Proceedings of SPARS 5*, 9–12.

ANON. 2012. Effect of noise, scene priors and multiplexing in computational imaging systems. *submitted to European Conference on Computer Vision*.

ASHOK, A., AND NEIFELD, M. A. 2010. Compressive Light Field Imaging. In *Proc. SPIE 7690*, 76900Q.

BABACAN, S., ANSORGE, R., LUESSI, M., MOLINA, R., AND KATSAGGELOS, A. 2009. Compressive sensing of light fields. In *Proc. ICIP*, 2337–2340.

BARANIUK, R., CEVHER, V., DUARTE, M., AND HEGDE, C. 2010. Model-based compressive sensing. *IEEE Trans. Inf. Theory 56*, 4, 1982–2001.

BECKER, S., BOBIN, J., AND CANDES, E. 2009. Nesta: A fast anad accurate first-order method for sparse recovery. In *Applied and Computational Mathematics*.

BISHOP, T., ZANETTI, S., AND FAVARO, P. 2009. Light-Field Superresolution. In *Proc. ICCP*, 1–9.

CANDÈS, E., AND TAO, T. 2005. Decoding by linear programming. *IEEE Trans. Inf. Theory 51*, 12, 4203–4215.

CANDÈS, E., AND TAO, T. 2006. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory 52*, 12, 5406–5425.

CANDÈS, E., ROMBERG, J., AND TAO, T. 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory 52*, 2, 489–509.

DAVIS, A., LEVOY, M., AND DURAND, F. 2012. Unstructured Light Fields. vol. 31, 1–10.

DONOHO, D. 2006. Compressed sensing. *IEEE Trans. Inf. Theory 52*, 4, 1289–1306.

DONOHO, D. 2006. For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics 59*, 6, 797–829.

EMMANUEL J. CANDES, M. B. W., AND BOYD, S. P. 2008. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*.

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

GEORGIEV, T., AND LUMSDAINE, A. 2010. Reducing Plenoptic Camera Artifacts. *Computer Graphics Forum 29*, 6, 1955–1968.

GEORGIEV, T., INTWALA, C., BABACAN, S., AND LUMSDAINE, A. 2008. Unified Frequency Domain Analysis of Lightfield Cameras. In *Proc. ECCV*, 224–237.

HAUPT, J., AND NOWAK, R. 2006. Signal reconstruction from noisy random projections. *IEEE Trans. Inf. Theory 52*, 9, 4036–4048.

IHRKE, I., WETZSTEIN, G., AND HEIDRICH, W. 2010. A Theory of Plenoptic Multiplexing. In *Proc. IEEE CVPR*, 1–8.

IVES, H., 1903. Parallax Stereogram and Process of Making Same. US patent 725,567.

KAY, S. M. 1993. Fundamentals of statistical signal processing: Estimation theory. *Prentice-Hall, USA*.

KREUTZ-DELGADO, K., AND RAO, B. 2000. Focuss-based dictionary learning algorithms. In *Proceedings of SPIE*, vol. 4119, 459.

KREUTZ-DELGADO, K., MURRAY, J., RAO, B., ENGAN, K., LEE, T., AND SEJNOWSKI, T. 2003. Dictionary learning algorithms for sparse representation. *Neural computation 15*, 2, 349–396.

LANMAN, D., RASKAR, R., AGRAWAL, A., AND TAUBIN, G. 2008. Shield Fields: Modeling and Capturing 3D Occluders. *ACM Trans. Graph. (Siggraph Asia) 27*, 5, 131.

LEVIN, A., AND DURAND, F. 2010. Linear View Synthesis Using a Dimensionality Gap Light Field Prior. In *Proc. IEEE CVPR*, 1–8.

Levin, A., Hasinoff, S. W., Green, P., Durand, F., and Freeman, W. T. 2009. 4D Frequency Analysis of Computational Cameras for Depth of Field Extension. *ACM Trans. Graph. (Siggraph) 28*, 3, 97.

Levoy, M., and Hanrahan, P. 1996. Light Field Rendering. In *Proc. ACM Siggraph*, 31–42.

Liang, C.-K., Lin, T.-H., Wong, B.-Y., Liu, C., and Chen, H. H. 2008. Programmable Aperture Photography: Multiplexed Light Field Acquisition. *ACM Trans. Graph. (Siggraph) 27*, 3, 1–10.

Lippmann, G. 1908. La Photographie Intégrale. *Academie des Sciences 146*, 446–451.

Lumsdaine, A., and Georgiev, T. 2009. The Focused Plenoptic Camera. In *Proc. ICCP*, 1–8.

Lytro, I., 2012. Lytro Light Field Camera. `www.lytro.com`.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. 2008. Supervised dictionary learning. *Arxiv preprint arXiv:0809.3083*.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. 2009. Online dictionary learning for sparse coding. *International Conference on Machine Learning*.

Needell, D., and Tropp, J. 2009. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harm. Anal. 26*, 3, 301–321.

Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. 2005. Light field photography with a hand-held plenoptic camera. Tech. Rep. Computer Science CSTR 2005-02, Stanford University.

Ng, R. 2005. Fourier Slice Photography. *ACM Trans. Graph. (Siggraph) 24*, 3, 735–744.

Sivaramakrishnan, S., Wang, A., Gill, P., and Molnar, A. 2011. Enhanced angle sensitive pixels for light field imaging. In *Electron Devices Meeting (IEDM), 2011 IEEE International*, IEEE, 8–6.

Taguchi, Y., Agrawal, A., Veeraraghavan, A., Ramalingam, S., and Raskar, R. 2010. Axial-Cones: Modeling Spherical Catadioptric Cameras for Wide-Angle Light Field Rendering. *ACM Trans. Graph. 29*, 172:1–172:8.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *J. Royal Statist. Soc B 58*, 1, 267–288.

Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., and Tumblin, J. 2007. Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocussing. *ACM Trans. Graph. (Siggraph) 26*, 3, 69.

Wang, A., Gill, P., and Molnar, A. 2011. An angle-sensitive cmos imager for single-sensor 3d photography. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, IEEE, 412–414.

Wilburn, B., Joshi, N., Vaish, V., Talvala, E.-V., Antunez, E., Barth, A., Adams, A., Horowitz, M., and Levoy, M. 2005. High Performance Imaging using Large Camera Arrays. *ACM Trans. Graph. (Siggraph) 24*, 3, 765–776.