

MIT Open Access Articles

A switchable light field camera architecture with Angle Sensitive Pixels and dictionary-based sparse coding

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Hirsch, Matthew, Sriram Sivaramakrishnan, Suren Jayasuriya, Albert Wang, Alyosha Molnar, Ramesh Raskar, and Gordon Wetzstein. "A Switchable Light Field Camera Architecture with Angle Sensitive Pixels and Dictionary-Based Sparse Coding." 2014 IEEE International Conference on Computational Photography (ICCP) (May 2014), Santa Clara, CA. p.1-10.

As Published: <http://dx.doi.org/10.1109/ICCPHOT.2014.6831813>

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Persistent URL: <http://hdl.handle.net/1721.1/92727>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



A Switchable Light Field Camera Architecture with Angle Sensitive Pixels and Dictionary-based Sparse Coding

Matthew Hirsch^{1*} Sriram Sivaramakrishnan^{2*} Suren Jayasuriya^{2*}
Albert Wang² Alyosha Molnar² Ramesh Raskar¹ Gordon Wetzstein¹

¹MIT Media Lab
Cambridge MA, USA

²Cornell University
Ithaca NY, USA

Abstract

We propose a flexible light field camera architecture that is at the convergence of optics, sensor electronics, and applied mathematics. Through the co-design of a sensor that comprises tailored, Angle Sensitive Pixels and advanced reconstruction algorithms, we show that—contrary to light field cameras today—our system can use the same measurements captured in a single sensor image to recover either a high-resolution 2D image, a low-resolution 4D light field using fast, linear processing, or a high-resolution light field using sparsity-constrained optimization.

1. Introduction

Over the last few years, light field acquisition has become one of the most widespread computational imaging techniques. By capturing the 4D spatio-angular radiance distribution incident on a sensor, light field cameras offer unprecedented flexibility for data processing. Post-capture image refocus, depth and 3D volume reconstruction, de-scattering, and synthetic aperture effects are only a few example applications. These unique capabilities make light field imaging an emerging technology that could soon be ubiquitous in consumer photography and scientific imaging, such as microscopy [18].

However, commercial devices offering light field capture modes have only had limited success thus far. One of the main reasons for this may be that conventional light field cameras are subject to the spatio-angular resolution trade-off. Whereas angular light information is captured to enable a variety of new modalities, this usually comes at the cost of severely reduced image resolution. Recent efforts have paved the way for overcoming the resolution tradeoff using sparse coding [23] or super-resolution techniques [30, 7].

*The indicated authors acknowledge equal contributions by sharing first authorship.

Although these methods improve the resolution of 4D light fields, it is still significantly lower than that offered by a regular camera sensor with the same pixel count. One may argue that light field cameras would be most successful if they could seamlessly switch between high-resolution 2D image acquisition and 4D light field capture modes.

In this paper, we explore such a switchable light field camera architecture. The required capabilities are facilitated by an emerging sensor design that uses Angle Sensitive Pixels (ASPs) [31, 32]. As shown in Figure 1, ASPs use special pixel structures that allow for angular radiance information to be captured without the need for additional microlenses [20] or light-blocking masks [13]. The physical principle behind ASPs is the Talbot effect: light incident on a pixel strikes two periodic diffraction gratings that are manufactured using commodity CMOS processes at a slight offset in front of the photodiodes. Whereas several ASP chip designs have been proposed in previous work [31, 27], we combine ASP hardware with modern techniques for compressive light field reconstruction and other processing modes into what we believe to be the most flexible light field camera architecture to date.

1.1. Contributions

In particular, we make the following contributions:

- We present a switchable camera allowing for high-resolution 2D image and 4D light field capture. These capabilities are facilitated by combining ASP sensors with modern signal processing techniques.
- We analyze the imaging modes of this architecture and demonstrate that a single image captured by the proposed camera provides either a high-resolution 2D image using little computation, a medium-resolution 4D light field using a moderate amount of computation, or a high-resolution 4D light field using more compute-intensive compressive reconstructions.



Figure 1: Prototype angle sensitive pixel camera (left). The data recorded by the camera prototype can be processed to recover a high-resolution 4D light field (center). As seen in the close-ups on the right, parallax is recovered from a single camera image.

- We evaluate system parameters and compare the proposed camera to existing light field camera designs. We also show results from a prototype camera system.

1.2. Overview of Limitations

Though the proposed reconstruction techniques allow for a variety of imaging modes, high-resolution light field reconstruction via nonlinear processing significantly increases the computational load compared to conventional photography. The prototype sensor has a relatively low pixel count and we observe slight optical aberrations.

2. Related Work

2.1. Light Field Acquisition

Light field cameras were invented more than a century ago. Early prototypes either used a microlens array [20] or a light-blocking mask [13] to multiplex the rays of a 4D light field onto a 2D sensor. In the last decades, significant improvements have been made to these basic designs, i.e. microlens-based systems have become digital [1, 24] and mask patterns more light efficient [29, 15]. However, the achievable resolution is fundamentally limited by the spatio-angular resolution tradeoff: spatial image resolution is sacrificed for capturing angular information with a single sensor. Detailed discussions of this topic can be found in the literature (e.g., [16, 35]).

Two common approaches seek to overcome this trade-off: using camera arrays [36, 30] or capturing multiple images of the scene from different perspectives [17, 12, 19]. However, camera arrays are usually bulky and expensive whereas multi-shot approaches restrict photographed scenes to be static. It is also possible to combine a regular camera and a microlens-based light field camera [21]; again, multiple devices are necessary. In this paper, we

Technique	Light Transmission	Image Resolution	Single Shot	Single Device	Comp. Complexity	High-Res 2D Image
Microlenses	high	low	yes	yes	low	no
Pinhole Masks	low	low	yes	yes	low	no
Coded Masks (SoS, MURA)	medium	low	yes	yes	medium	no
Scanning Pinhole	low	high	no	yes	low	yes
Camera Array	high	high	yes	no	medium	yes
Compressive LF	medium	high	yes	yes	high	yes*
Proposed Method	high	low/high	yes	yes	medium/high	yes

*With extra computation

Table 1: Overview of benefits and limitations of light field photography techniques. As opposed to existing approaches, the proposed computational camera system provides high light field resolution from a single recorded image. In addition, our switchable camera is flexible enough to provide additional imaging modes that include conventional, high-resolution 2D photography.

present a new camera architecture that uses a single device to recover both a conventional 2D image and a high-resolution 4D light field from a single image.

2.2. Overcoming the Device/Resolution Tradeoff

It is well-understood that light fields of natural scenes contain a significant amount of redundancy. Most objects are diffuse; a textured plane at some depth, for instance, will appear in all views of a captured light field, albeit at slightly different positions. This information can be fused using super-resolution techniques, which compute a high-resolution image from multiple subpixel-shifted, low-resolution images [28, 26, 5, 22, 25, 30, 7, 34].

With the discovery of compressed sensing [8, 9], a new generation of compressive light field camera architectures is emerging that goes far beyond the improvements offered by super-resolution. For example, the spatio-angular resolution tradeoff in single-device light field cameras [3, 4, 37, 23] can be overcome or the number of required cameras in arrays reduced [14]. Compressive ap-

proaches rely on increased computational processing with sparsity priors to provide higher image resolutions than otherwise possible.

The camera architecture proposed in this paper is well-suited for compressive reconstructions, for instance with dictionaries of light field atoms [23]. In addition, our flexible approach allows for high-quality 2D image and lower-resolution light field reconstruction from the *same measured data* without numerical optimization.

2.3. Angle Sensitive Pixels

Whereas light field cameras typically rely on modern algorithms applied to data captured with off-the-shelf opto-electronic systems, recent advances in complementary metal-oxide-semiconductor (CMOS) processes have created opportunities for more specialized sensors. In particular, angle sensitive pixels (ASPs) have recently been proposed to capture spatio-angular image information [31]. These pixel architectures use a pair of near-wavelength gratings in each pixel to tune the angular response of each sensor element using the Talbot effect. Creating a sensor of tiled ASPs with pre-selected responses enables range imaging, focal stacks [32], and lensless imaging [11]. Optically optimized devices, created with phase gratings and multiple interdigitated diodes can achieve quantum efficiency comparable to standard CMOS imagers [27].

ASPs represent a promising sensor topology, as they are capable of reconstructing both sensor-resolution conventional 2D images and space/angle information from a single shot (see Sec. 3). However, general light field reconstruction techniques have not previously been described with this hardware. We analyze ASPs in the context of high-resolution, compressive light field reconstruction and explore flexible image modalities for an emerging class of cameras based on ASP sensors.

3. Method

This section introduces the image formation model for ASP devices. In developing the mathematical foundation for these camera systems, we entertain two goals: to place the camera in a framework that facilitates comparison to existing light field cameras, and to understand the plenoptic sampling mechanism of the proposed camera.

3.1. Light Field Acquisition with ASPs

The Talbot effect created by periodic gratings induces a sinusoidal angular response from ASPs [27]. For a one-dimensional ASP, this can be described as

$$\rho^{(\alpha,\beta)}(\theta) = 1/2 + m/2 \cos(\beta\theta + \alpha). \quad (1)$$

Here, α and β are phase and frequency, respectively, m is the modulation efficiency, and θ is the angle of inci-

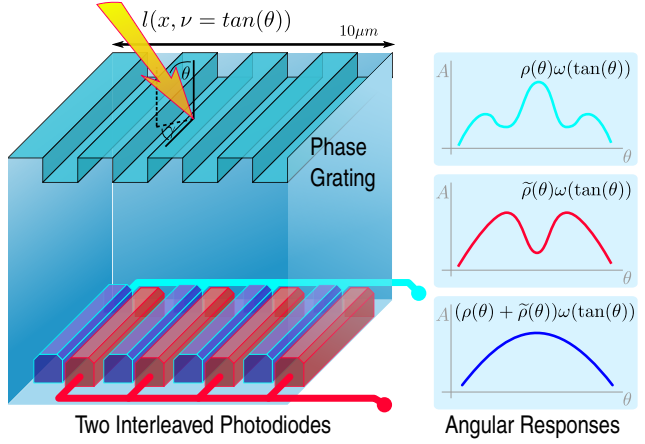


Figure 2: Schematic of a single angle sensitive pixel. Two interleaved photodiodes capture a projection of the light field incident on the sensor (left). The angular responses of these diodes are complementary: a conventional 2D image can be synthesized by summing their measurements digitally (right).

dent light. Specific values of these parameters used in our experimental setup can be found in Section 5.1. Both α and β can be tuned in the sensor fabrication process [32]. Common implementations choose ASP types with $\alpha \in 0, \pi/2, \pi, 3\pi/4$. We note that prior publications describe the ASP response without the normalization constant of $1/2$ introduced here. Normalizing Equations 1 and 2 simplifies the discussion of 2D image recovery using ASPs.

Similarly, 2D ASP implementations exhibit the resulting angular responses for incident angles θ_x and θ_y :

$$\rho^{(\alpha,\beta,\gamma)}(\theta) = 1/2 + m/2 \cos(\beta(\cos(\gamma)\theta_x + \sin(\gamma)\theta_y) + \alpha), \quad (2)$$

where α is phase, β frequency, and γ grating orientation.

The captured sensor image i is then a projection of the incident light field l weighted by the angular responses of a mosaic of ASPs:

$$i(\mathbf{x}) = \int_{\mathcal{V}} l(\mathbf{x}, \nu) \rho(\mathbf{x}, \tan^{-1}(\nu)) \omega(\nu) d\nu. \quad (3)$$

In this formulation, $l(\mathbf{x}, \nu)$ is the light field inside the camera behind the main lens. We describe the light field using a relative two-plane parameterization [10], where $\nu = \tan(\theta)$. The integral in Equation 3 contains angle-dependent vignetting factors $\omega(\nu)$ and the aperture area \mathcal{V} restricts the integration domain. Sensor noise is discounted in this idealized representation, though it is addressed during discretization below. Finally, the spatial coordinates $\mathbf{x} = \{x, y\}$ are defined on the sensor pixel-level; the geometrical microstructure of ASP gratings and photodiodes is not observable at the considered scale. In practice, the spatially-varying pixel response function $\rho(\mathbf{x}, \theta)$ is a periodic mosaic of a few different ASP types. A common example of such a layout for color imaging is the Bayer filter array that

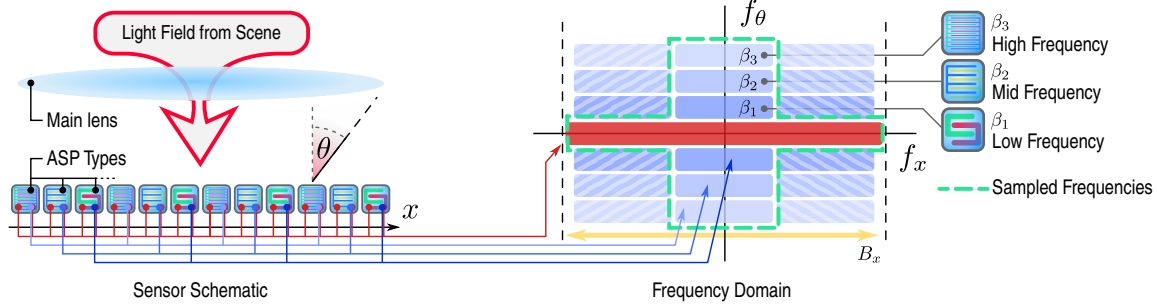


Figure 3: Illustration of ASP sensor layout (left) and sampled spatio-angular frequencies (right). The pictured sensor interleaves three different types of ASPs. Together, they sample all frequencies contained in the dashed green box (right). A variety of light field reconstruction algorithms can be applied to these measurements, as described in the text.

interleaves red, green, and blue subpixels. ASPs with different parameters (α, β, γ) can be fabricated following this scheme. Mathematically, this type of spatial multiplexing is formulated as

$$\rho(\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^N \left(\text{III}^{(k)}(\mathbf{x}) * \rho^{(\zeta(k))}(\boldsymbol{\theta}) \right), \quad (4)$$

where $*$ is the convolution operator and $\text{III}^{(k)}(\mathbf{x})$ is a sampling operator consisting of a set of Dirac impulses describing the spatial layout of one type of ASP. A total set of N types is distributed in a regular grid over the sensor. The parameters of each are given by the mapping function $\zeta(k) : \mathbb{N} \rightarrow \mathbb{R}^3$ that assigns a set of ASP parameters (α, β, γ) to each index k .

Whereas initial ASP sensor designs use two layered, attenuating diffraction gratings and conventional photodiodes underneath [31, 32, 11], more recent versions enhance the quantum efficiency of the design by using a single phase grating and an interleaved pair of photodiodes [27]. For the proposed switchable light field camera, we illustrate the latter design with the layout of a single pixel in Figure 2.

In this sensor design, each pixel generates two measurements: one that has an angular response described by Equation 2 and another one that has a complementary angular response $\tilde{\rho} = \rho^{(\alpha+\pi, \beta, \gamma)}$ whose phase is shifted by π . The discretized version of the two captured images can be written as a simple matrix-vector product:

$$\mathbf{i} = \Phi \mathbf{l} + \epsilon, \quad (5)$$

where $\mathbf{i} \in \mathbb{R}^{2p}$ is a vector containing both images $i(\mathbf{x})$ and $\tilde{i}(\mathbf{x})$, each with a resolution of p pixels, and $\Phi \in \mathbb{R}^{2p \times \mathbb{R}^n}$ is the projection matrix that describes how the discrete, vectorized light field $\mathbf{l} \in \mathbb{R}^n$ is sensed by the individual photodiodes. In Equation 5, sensor noise is modeled as Gaussian, i.i.d., and represented by ϵ .

3.2. Image and Light Field Synthesis

In this section, we propose three alternative ways to process the data recorded with an ASP sensor.

3.2.1 Direct 2D Image Synthesis

As illustrated in Figure 2, the angular responses of the complementary diodes in each pixel can simply be summed to generate a conventional 2D image, i.e. $\rho^{(\alpha, \beta, \gamma)} + \tilde{\rho}^{(\alpha, \beta, \gamma)}$ is a constant. Hence, Equation 3 reduces to the conventional photography equation:

$$i(\mathbf{x}) + \tilde{i}(\mathbf{x}) = \int_{\mathcal{V}} l(\mathbf{x}, \boldsymbol{\nu}) \omega(\boldsymbol{\nu}) d\boldsymbol{\nu}, \quad (6)$$

which can be implemented in the camera electronics. Equation 6 shows that a conventional 2D image can easily be generated from an ASP sensor. While this may seem trivial, existing light field camera architectures using microlenses or coded masks cannot easily synthesize a conventional 2D image for in-focus and out-of-focus objects.

3.2.2 Linear Reconstruction for Low-resolution 4D Light Fields

Using a linear reconstruction framework, the same data can alternatively be used to recover a low-resolution 4D light field. We model light field capture by an ASP sensor as Equation 5 where the rows of Φ correspond to vectorized 2D angular responses of different ASPs. These angular responses are either sampled uniformly from Equation 2 or fit empirically from measured impulse responses. The approximate orthonormality of the angular wavelets (see Sec. 5) implies $\Phi^T \Phi \approx \mathbf{I}$. Consequently $\Sigma = \text{diag}(\Phi^T \Phi)$ is used as a preconditioner for inverting the capture equation: $\mathbf{l} = \Sigma^{-1} \Phi^T \mathbf{i}$.

The main benefit of a linear reconstruction is its computational performance. However, the spatial resolution of the resulting light field will be approximately k -times lower

than that of the sensor ($k = n/p$) since the different ASPs are grouped into tiles on the sensor. Similarly to demosaicing from color filter arrays, different angular measurements from the ASP sensor can be demosaiced using interpolation and demultiplexing [35] to improve visual appearance. In addition, recent work on light field super-resolution has demonstrated that resolution loss can be slightly mitigated for the particular applications of image refocus [30] and volume reconstruction [7].

3.2.3 Sparse Coding for High-resolution Light Fields

Finally, we can choose to follow Marwah et al. [23] and apply nonlinear sparse coding techniques to recover a high-resolution 4D light field from the same measurements. This is done by representing the light field using an overcomplete dictionary as $\mathbf{l} = \mathcal{D}\chi$, where $\mathcal{D} \in \mathbb{R}^{n \times d}$ is a dictionary of light field atoms and $\chi \in \mathbb{R}^d$ are the corresponding coefficients. Natural light fields have been shown to be sparse in such dictionaries [23], i.e. the light field can be represented as a weighted sum of a few light field atoms (columns of the dictionary). For robust reconstruction, a basis pursuit denoise problem (BPDN) is solved

$$\begin{aligned} & \underset{\{\chi\}}{\text{minimize}} && \|\chi\|_1 \\ & \text{subject to} && \|\mathbf{i} - \Phi\mathcal{D}\chi\|_2 \leq \epsilon, \end{aligned} \quad (7)$$

where ϵ is the sensor noise level. Whereas this approach offers significantly increased light field resolution, it comes at an increased computational cost. Note that Equation 7 is applied to a small, sliding window of the recorded data, each time recovering a small 4D light field patch rather than the entire 4D light field at once. In particular, window blocks with typical sizes of 9×9 pixels are processed in parallel to yield light field patches with $9 \times 9 \times 5 \times 5$ rays each. See Section 5.2 for implementation details.

4. Analysis

In this section, we analyze the proposed methods and compare them to alternative light field sensing approaches.

4.1. Frequency Analysis

As discussed in the previous section, Angle Sensitive Pixels sample a light field such that a variety of different reconstruction algorithms can be applied to the same measurements. To understand the information contained in the measurements, we can turn to a frequency analysis. Figure 3 (left) illustrates a one-dimensional ASP sensor with three interleaved types of ASPs sampling low, mid, and high angular frequencies, respectively. As discussed in Section 3.2.1, the two measurements from the two interdigitated diodes in each pixel can be combined to synthesize a

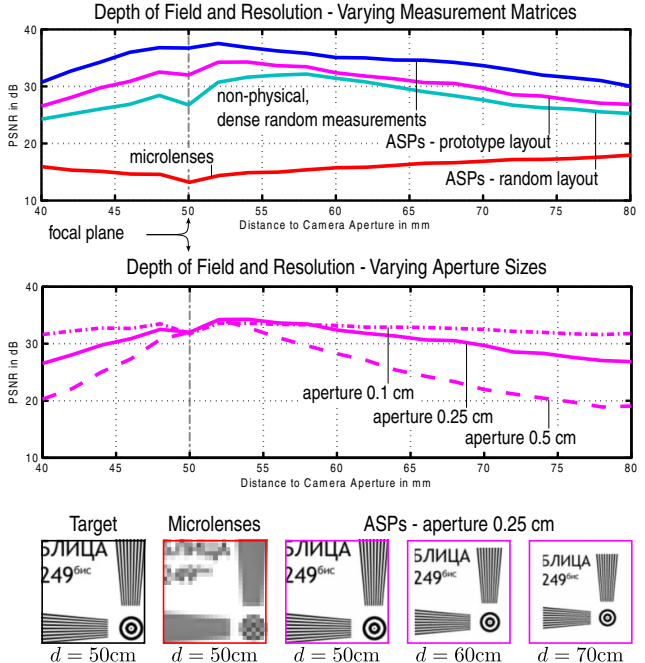


Figure 4: Evaluating depth of field. Comparing the reconstruction quality of several different optical setups shows that the ASP layout in the prototype camera is well-suited for sparsity-constrained reconstructions using overcomplete dictionaries (top). The dictionaries perform best when the parallax in the photographed scene is smaller or equal to that of the training light fields (center). Central views of reconstructed light fields are shown in the bottom.

conventional 2D image. This image has no angular information but samples the entire spatial bandwidth B_x of the sensor (Fig. 3 right, red box).

The measurements of the individual photodiodes contain higher angular frequency bands, but only for lower spatial frequencies due to the interleaved sampling pattern (Fig. 3 right, solid blue boxes). A linear reconstruction (Sec. 3.2.2) would require an optical anti-aliasing filter to be mounted on top of the sensor, as is commonly found in commercial sensors. In the absence of an optical anti-aliasing filter, aliasing is observed. For the proposed application, aliasing results in downmixing of high spatio-angular frequencies (Fig. 3 right, hatched blue boxes) into lower spatial frequency bins. As spatial frequencies are sampled by an ASP sensor while angular frequencies are measured continuously, aliasing occurs only among spatial frequencies. The region of the spatio-angular frequency plane sampled by the ASP sensor in Figure 3 is highlighted by the dashed green box. Although aliasing makes it difficult to achieve high-quality reconstructions with simple linear demosaicing, it is crucial in preserving information for non-linear, high-resolution reconstructions based on sparsity-constrained optimization (Sec. 3.2.3).

4.2. Depth of Field

To evaluate the depth of field that can be achieved with the proposed sparsity-constrained reconstruction methods, we simulate a two-dimensional resolution chart at multiple different distances to the camera’s focal plane. The results of our simulations are documented in Figure 4. The camera is focused at 50 cm, where no parallax is observed in the light field. At distances closer to the camera or farther away the parallax increases—we expect the reconstruction algorithms to achieve a lower peak signal-to-noise ratio (PSNR). The PSNR is measured between the depth-varying target 4D light field and the reconstructed light field.

Figure 4 (top) compares sparsity-constrained reconstructions using different measurement matrices and also a direct sampling of the low-resolution light field using microlenses (red plot). Slight PSNR variations in the latter are due to the varying size of the resolution chart in the depth-dependent light fields, which is due to the perspective of the camera (cf. bottom images). Within the considered depth range, microlenses always perform poorly.

The different optical setups tested for the sparsity-constrained reconstructions include the ASP layout of our prototype (magenta plot, described in Sec. 5.1), ASPs with completely random angular responses that are also randomized over the sensor (green plot), and also a dense random mixing of all light rays in each of the light field patches (blue plot). A dense random mixing across a light field patch requires that each measurement within the patch is a random mixture of all spatial and angular samples that fall within the patch. Though such a mixture is not physically realizable, it does yield an intuition of the approximate achievable upper performance bounds. Unsurprisingly, such a dense, random measurement matrix Φ performs best. What is surprising, however, is that random ASPs are worse than the choice of regularly-sampled angular wavelet coefficients in our prototype (see Sec. 5.1). For compressive sensing applications, the rows of the measurement matrix Φ should be as incoherent (or orthogonal) as possible to the columns of the dictionary \mathcal{D} . For the particular dictionary used in these experiments, random ASPs seem to be more coherent with the dictionary. These findings are supported by Figure 5. We note that the PSNR plots are content-dependent and also dependent on the employed dictionary.

The choice of dictionary is critical. The one used in Figure 4 is learned from 4D light fields showing 2D planes with random text within the same depth range as the resolution chart. If the aperture size of the simulated camera matches that used in the training set (0.25 cm), we observe high reconstruction quality (solid line, center plots). Smaller aperture sizes will result in less parallax and can easily be recovered as well, but resolution charts rendered at larger aperture sizes also contain a larger amount of parallax than any of the

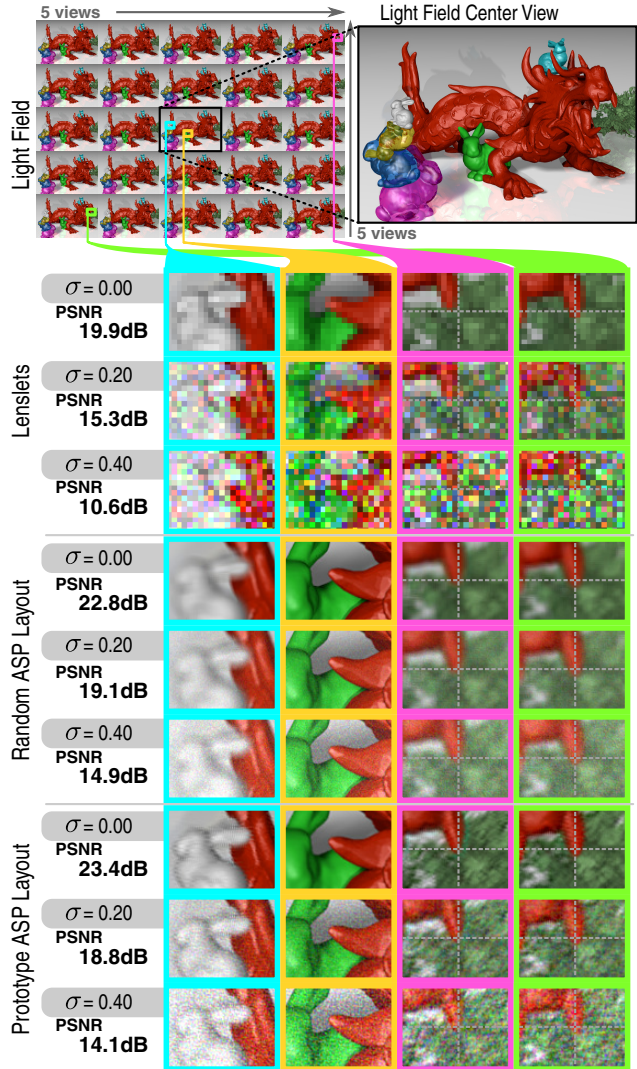


Figure 5: Simulated light field reconstructions from a single coded sensor image for different levels of noise and three different optical sampling schemes. For the ASP layout in the prototype camera (bottom), high levels of noise result in noisy reconstructions—parallax is faithfully recovered (dragon’s teeth, lower right, fiducials added). A physically-realizable random ASP layout (center) does not measure adequate samples for a sparse reconstruction to recover a high-quality light field from a single sensor image; the reconstructions look more blurry and parallax at the views is poorly recovered (center, right). A standard lenslet-based reconstruction (top) subsamples spatial information. Noise is more apparent in the lenslet case as BPDN attenuates noise in the other cases. In all cases, the peak sensor measurement magnitude is normalized on $[0\ 1]$ prior to adding Gaussian noise.

training data. The reconstruction quality in this case drops rapidly with increasing distance to the focal plane (Fig. 4, center plots).

4.3. Resilience to Noise

Finally, we evaluate the sparse reconstruction algorithm proposed in Section 3.2.3 w.r.t. noise and compare three different optical sampling schemes. Figure 5 shows a synthetic light field with 5×5 different views. We simulate sensor images with zero-mean i.i.d. Gaussian noise and three different standard deviations $\sigma = \{0.0, 0.2, 0.4\}$. In addition, we compare the ASP layout of the prototype (see Sec. 5.1) with a random layout of ASPs that each also have a completely random angular response. Confirming the depth of field plots in Figure 4, a random ASP layout achieves a lower reconstruction quality than sampling wavelet-type angular basis functions on a regular grid. Again, this result may be counter-intuitive because most compressive sensing algorithms perform best when random measurement matrices are used. However, these usually assume a dense random matrix Φ (simulated in Fig. 4), which is not physically realizable in an ASP sensor. One may believe that a randomization of the available degrees of freedom of the measurement system may be a good approximation of the fully random matrix, but this is clearly not the case. We have not experimented with optical layouts that are optimized for a particular dictionary [23], but expect such codes to further increase reconstruction quality.

5. Implementation

5.1. Angle Sensitive Pixel Hardware

A prototype ASP light field camera was built using an angle sensitive pixel array sensor [33]. The sensor consists of 24 different ASP types, each of which has a unique response to incident angle described by Equation 2. Since a single pixel generates a pair of outputs, a total of 48 distinct angular measurements are read out from the array. Recall from Section 3 that ASP responses are characterized by the parameters α , β , γ , and m which define the phase, two dimensional angular frequency, and modulation efficiency of the ASP. The design includes three groups of ASPs that cover low, medium, and high frequencies with β values of 12, 18 and 24, respectively. The low and high frequency groups of ASPs have orientations (γ in degrees) of 0° , 90° and $\pm 45^\circ$ whereas the mid frequency group is staggered in frequency space with respect to the other two and has γ values of $\pm 22.5^\circ$ and $\pm 67.5^\circ$. Individual ASPs are organized into a rectangular unit cell that is repeated to form the array. Within each tile, the various pixel types are distributed randomly so that any patch of pixels has a uniform mix of orientations and frequencies as illustrated in Figure 6. The modulation efficiency, m , is a process parameter and typical values are measured to be near 0.5 with some dependence on wavelength [31]. The die size is 5×5 mm which accommodates a 96×64 grid of tiles, or 384×384 pixels.

In addition to the sensor chip, the only optical compo-

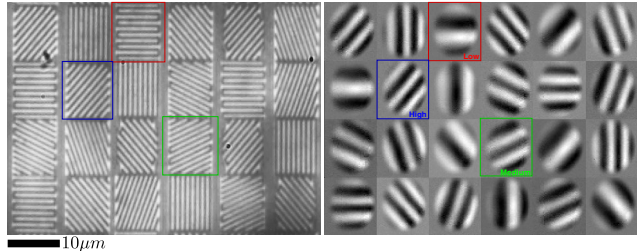


Figure 6: Microscopic image of a single 6×4 pixel tile of the ASP sensor (left). We also show captured angular point spread functions (PSFs) of each ASP pixel type (right).

nent in the camera is the focusing lens. We used a commercial 50 mm Nikon manual focus lens at an aperture setting of $f/1.2$. The setup, consisting of the data acquisition boards that host the imager chip, and the lens, can be seen in Figure 1. The target imaging area was staged at a distance of 1m from the sensor which provided a 10:1 magnification. Calibration of the sensor response was performed by imaging a 2mm diameter, back-illuminated hole positioned far away from the focal plane. Figure 6 shows the captured angular point spread function for all 24 ASP types. These responses were empirically fitted and resampled to form the rows of the projection matrix Φ for both the linear and non-linear reconstructions on captured data.

5.2. Software Implementation

The compressive part of our software pipeline closely follows that of Marwah et al. [23]. Conceptually, nonlinear reconstructions depend on an offline dictionary learning phase, followed by an online reconstruction over captured data. To avoid the challenges of large-scale data collection with our prototype hardware, we used the dictionaries provided by Marwah et al. to reconstruct light fields from the prototype hardware. Dictionaries used to evaluate depth of field in Figure 4 were learned using KSVD [2].

Online reconstruction was implemented by the Alternating Direction Method of Multipliers (ADMM) [6] with parameters $\lambda = 10^{-5}$, $\rho = 1$, and $\alpha = 1$, to solve the ℓ_1 -regularized regression (BPDN) of Equation 7. ASP sensor images were subdivided into sliding, 9×9 pixel windows; small 4D light field patches were reconstructed for each window, each with 5×5 angles. The sliding reconstruction window was translated in one pixel increments over the full 384×384 pixel sensor image and the results were integrated with an average filter. Reconstructions were computed on an 8-core Intel Xeon workstation with 16GB of RAM. Average reconstruction time for experiments in Section 6 was 8 hours. Linear reconstruction algorithms are significantly faster, taking less than one minute for each result.

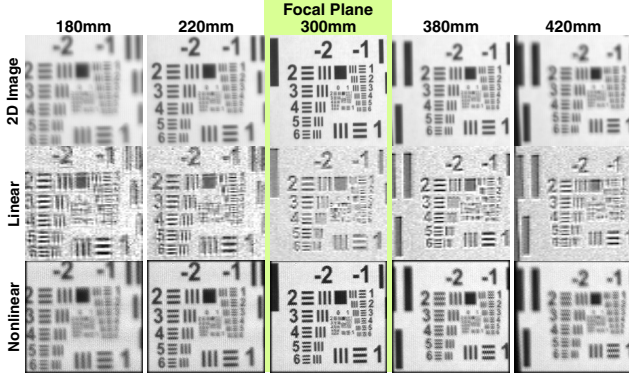


Figure 7: Evaluation of prototype resolution. We capture images of a resolution target at different depths and compare the 2D image (top), center view of the linearly reconstructed light field (center), and center view of the nonlinearly reconstructed light field (bottom).

6. Results

This section shows an overview of experiments with the prototype camera. In Figure 7, we evaluate the resolution of the device for all three proposed reconstruction algorithms. As expected for a conventional 2D image, the depth of field is limited by the f-number of the imaging lens, resulting in out-of-focus blur for a resolution chart that moves away from the focal plane (top row). The proposed linear reconstruction recovers the 4D light field at a low resolution (center row). Due to the lack of an optical anti-aliasing filter in the camera, aliasing is observed in the reconstructions. The anti-aliasing filter would remove these artifacts but also decrease image resolution. The resolution of the light field recovered using the sparsity-constrained nonlinear methods has a resolution comparable to the in-focus 2D image. Slight artifacts in the recovered resolution charts correspond to those observed in noise-free simulations (cf. Fig. 5). We believe these artifacts are due to the large compression ratio—25 light field views are recovered from a single sensor image via sparsity-constrained optimization.

We show additional comparisons of the three reconstruction methods for a more complex scene in Figure 8. Though an analytic comparison of resolution improvement by our nonlinear method is not currently possible, referring to Figure 4 (top) at the focal plane depth yields a numerical comparison for a simulated resolution chart.

Figure 9 shows several scenes that we captured in addition to those already shown in Figures 1 and 8. Animations of the recovered light fields for all scenes can be found in the supplementary video. We deliberately include a variety of effects in these scenes that are not easily captured in alternatives to light field imaging (e.g., focal stacks or range imaging), including occlusion, refraction, and translucency. Specular highlights, as for instance seen on the glass piglet in the two scenes on the right, often lead to sensor saturation

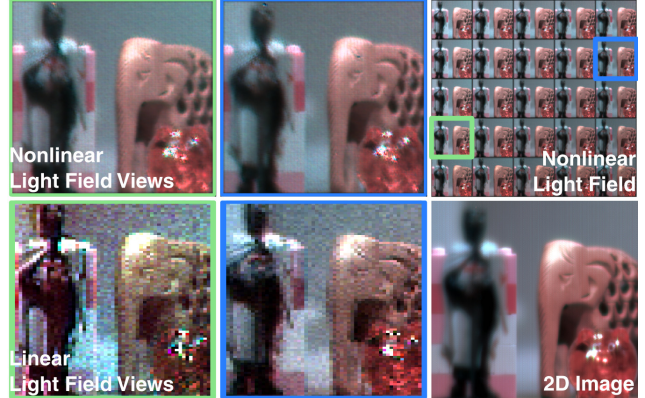


Figure 8: Comparison of different reconstruction techniques for the same captured data. We show reconstruction of a 2D image (bottom right), a low-resolution light field via linear reconstruction (bottom left and center), and a high-resolution light field via sparsity-constrained optimization with overcomplete dictionaries (top). Whereas linear reconstruction trades angular for spatial resolution—thereby decreasing image fidelity—nonlinear reconstructions can achieve an image quality that is comparable to a conventional, in-focus 2D image for each of 25 recovered views.

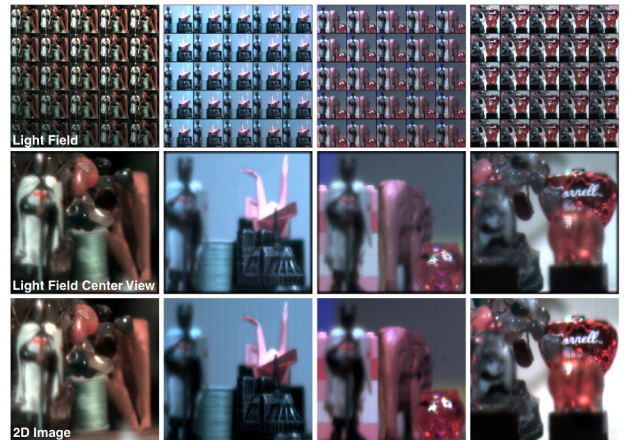


Figure 9: Overview of captured scenes showing mosaics of light fields reconstructed via sparsity-constrained optimization (top), a single view of these light fields (center), and corresponding 2D images (bottom). These scenes exhibit a variety of effects, including occlusion, refraction, specularity, and translucency. The resolution of each of the 25 light field views is similar to that of the conventional 2D images.

tion, which causes artifacts in the reconstructions. This is a limitation of the proposed reconstruction algorithms.

Finally, we show in Figure 10 that the recovered light fields contain enough parallax to allow for post-capture image refocus. Chromatic aberrations in the recorded sensor image and a limited depth of field of each recovered light field view place an upper limit on the resolvable resolution of the knight (right).



Figure 10: Refocus of the “Knight & Crane” scene.

7. Discussion

In summary, we present a flexible light field camera architecture that combines Angle Sensitive Pixel sensors with modern mathematical techniques for sparse signal recovery. We evaluate system parameters in simulation, present a frequency analysis of the camera design, and demonstrate experimentally that the recorded data facilitates an unprecedented flexibility for post-processing. In particular, we show conventional 2D image reconstruction, fast reconstruction of low-resolution 4D light fields, and also more computationally intensive, sparsity-constrained reconstructions of high-resolution 4D light fields.

Limitations The low resolution of the prototype chip is not comparable with modern, commercial image sensors offering tens of megapixels. A color filter array is not integrated into the chip; at the moment, we capture color results in three photographs, each using a different color filter in front of the main lens. Our chip was fabricated in a commercial mixed signal complementary metal-oxide-semiconductor (CMOS) process that was not optimized for imaging and, as such, exhibits a lower signal-to-noise ratio (SNR) than commercial image sensors. However, the proposed ASP chip design can be replicated in more optimized fabrication processes, yielding significant improvements in quantum efficiency and SNR.

Future Work The Talbot effect creating the angular responses of the pixels is based on diffraction and therefore wavelength-dependent. This results in slightly different angular frequencies captured in different color channels. We plan to exploit cross-spectral information for enhanced signal recovery in the future and also extend the employed overcomplete 4D dictionaries to include the spectral domain in addition to space and angle. Finally, we would also like to explore new spatial layouts of ASP subpixels and tailor angular responses of individual pixels to the employed dictionaries.

Conclusion Computational photography is at the intersection of optics, sensor electronics, applied mathematics, and high performance computing. In this paper, we propose a system that couples the design of all of these aspects to achieve an unprecedented amount of flexibility in computational light field imaging. We hope to inspire the community to follow similar strategies for other applications and unlock the true potential of next-generation computational cameras.

Acknowledgements

We thank the anonymous reviewers for their insightful feedback. We recognize the Camera Culture group for helpful discussions and support. This work was supported in part by NSF Grant IIS-1218411, NSF Grant IIS-1116452, and MIT Media Lab consortia funding. Suren Jayasuriya was supported by an NSF Graduate Research Fellowship. Gordon Wetzstein was supported by an NSERC Postdoctoral Fellowship. Ramesh Raskar was supported by the Alfred P. Sloan Research Fellowship and DARPA Young Faculty Award.

References

- [1] E. H. Adelson and J. Y. Wang. Single lens stereo with a plenoptic camera. *IEEE Trans. PAMI*, 14(2):99–106, 1992.
- [2] M. Aharon, M. Elad, and A. Bruckstein. K-svd: Design of dictionaries for sparse representation. *Proceedings of SPARS*, 5:9–12, 2005.
- [3] A. Ashok and M. A. Neifeld. Compressive light field imaging. In *SPIE Defense, Security, and Sensing*, pages 76900Q–76900Q. International Society for Optics and Photonics, 2010.
- [4] D. Babacan, R. Ansorge, M. Luessi, P. Ruiz, R. Molina, and A. Katsaggelos. Compressive light field sensing. 2012.
- [5] T. E. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *Proc. ICCP*, pages 1–9. IEEE, 2009.
- [6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Matlab scripts for alternating direction method of multipliers. Technical report, Technical report, <http://www.stanford.edu/boyd/papers/admm>, 2012.
- [7] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy. Wave optics theory and 3-d deconvolution for the light field microscope. *Optics express*, 21(21):25418–25439, 2013.
- [8] E. Candès, J. Romberg, and T. Tao. Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

- [9] D. Donoho. Compressed Sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [10] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. X. Sillion. A frequency analysis of light transport. In *ACM Trans. Graph. (SIGGRAPH)*, volume 24, pages 1115–1126, 2005.
- [11] P. R. Gill, C. Lee, D.-G. Lee, A. Wang, and A. Molnar. A microscale camera using direct fourier-domain scene capture. *Optics letters*, 36(15):2949–2951, 2011.
- [12] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. SIGGRAPH*, pages 43–54, 1996.
- [13] H. Ives. Parallax Stereogram and Process of Making Same. US patent 725,567, 1903.
- [14] M. Kamal, M. Golbabaee, and P. Vanderghenst. Light Field Compressive Sensing in Camera Arrays. In *Proc. ICASSP*, pages 5413–5416, 2012.
- [15] D. Lanman, R. Raskar, A. Agrawal, and G. Taubin. Shield fields: modeling and capturing 3d occluders. In *ACM Trans. Graph. (SIGGRAPH)*, volume 27, page 131, 2008.
- [16] A. Levin, W. T. Freeman, and F. Durand. Understanding camera trade-offs through a bayesian analysis of light field projections. In *Proc. ECCV*, 2008.
- [17] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. SIGGRAPH*, pages 31–42, 1996.
- [18] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz. Light field microscopy. *ACM Trans. Graph. (SIGGRAPH)*, 25(3):924–934, 2006.
- [19] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *ACM Trans. Graph. (SIGGRAPH)*, volume 27, page 55, 2008.
- [20] G. Lippmann. La Photographie Intégrale. *Academie des Sciences*, 146:446–451, 1908.
- [21] C.-H. Lu, S. Muenzel, and J. Fleischer. High-resolution light-field microscopy. In *Proc. OSA COSI*, 2013.
- [22] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *Proc. ICCP*, pages 1–8, 2009.
- [23] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graph. (TOG)*, 32(4):46, 2013.
- [24] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11), 2005.
- [25] C. Perwass and L. Wietzke. Single Lens 3D-Camera with Extended Depth-of-Field. In *Proc. SPIE 8291*, pages 29–36, 2012.
- [26] P. M. Shankar, W. C. Hasenplaugh, R. L. Morrison, R. A. Stack, and M. A. Neifeld. Multiaperture imaging. *Appl. Opt.*, 45(13):2871–2883, 2006.
- [27] S. Sivaramakrishnan, A. Wang, P. R. Gill, and A. Molnar. Enhanced angle sensitive pixels for light field imaging. In *Proc. IEEE IEDM*, pages 8–6, 2011.
- [28] J. Tanida, T. Kumagai, K. Yamada, S. Miyatake, K. Ishida, T. Morimoto, N. Kondou, D. Miyazaki, and Y. Ichioka. Thin observation module by bound optics (tombo): Concept and experimental verification. *Appl. Opt.*, 40(11):1806–1813, 2001.
- [29] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph. (SIGGRAPH)*, 26(3):69, 2007.
- [30] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar. Picam: an ultra-thin high performance monolithic camera array. *ACM Trans. Graph. (SIGGRAPH Asia)*, 32(6):166, 2013.
- [31] A. Wang, P. Gill, and A. Molnar. Light field image sensors based on the talbot effect. *Applied optics*, 48(31):5897–5905, 2009.
- [32] A. Wang, P. R. Gill, and A. Molnar. An angle-sensitive cmos imager for single-sensor 3d photography. In *Proc. IEEE Solid-State Circuits Conference (ISSCC)*, pages 412–414. IEEE, 2011.
- [33] A. Wang, S. Sivaramakrishnan, and A. Molnar. A 180nm cmos image sensor with on-chip optoelectronic image compression. In *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pages 1–4, 2012.
- [34] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. PAMI*, 2013.
- [35] G. Wetzstein, I. Ihrke, and W. Heidrich. On Plenoptic Multiplexing and Reconstruction. *IJCV*, 101:384–400, 2013.
- [36] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph. (SIGGRAPH)*, 24(3):765–776, 2005.
- [37] Z. Xu and E. Y. Lam. A high-resolution lightfield camera with dual-mask design. In *SPIE Optical Engineering+Applications*, pages 85000U–85000U, 2012.