Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections

Kshitij Marwah[1] Gordon Wetzstein[1] Yosuke Bando[2,1] Ramesh Raskar[1]
[1]MIT Media Lab [2]Toshiba Corporation

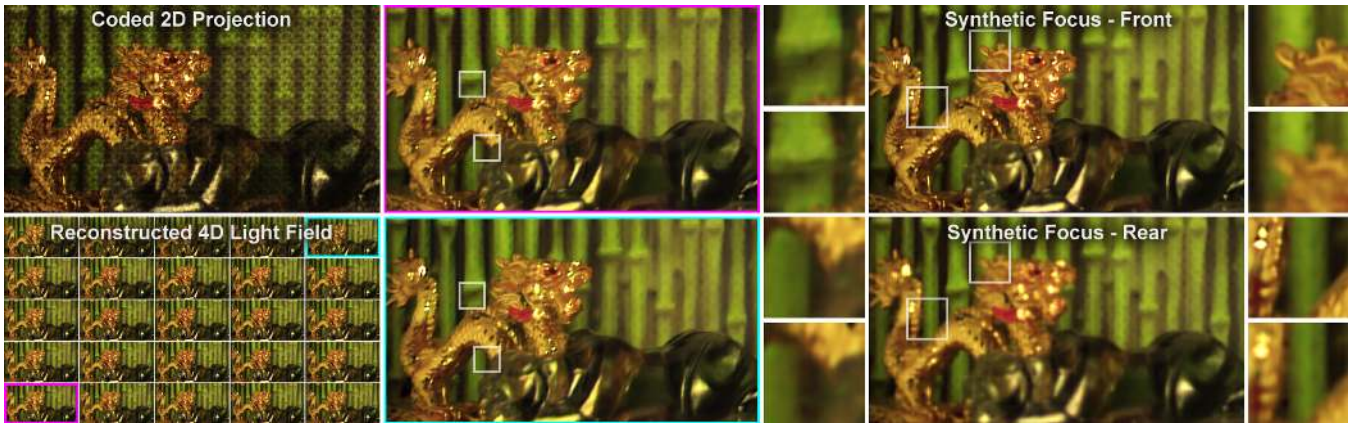

Figure 1: *Light field reconstruction from a single coded projection. We explore sparse reconstructions of 4D light fields from optimized 2D projections using light field atoms as the fundamental building blocks of natural light fields. This example shows a coded sensor image captured with our camera prototype (upper left), and the recovered 4D light field (lower left and center). Parallax is successfully recovered (center insets) and allows for post-capture refocus (right). Even complex lighting effects, such as occlusion, specularity, and refraction, can be recovered, being exhibited by the background, dragon, and tiger, respectively.*

Abstract

Light field photography has gained a significant research interest in the last two decades; today, commercial light field cameras are widely available. Nevertheless, most existing acquisition approaches either multiplex a low-resolution light field into a single 2D sensor image or require multiple photographs to be taken for acquiring a high-resolution light field. We propose a compressive light field camera architecture that allows for higher-resolution light fields to be recovered than previously possible from a single image. The proposed architecture comprises three key components: light field atoms as a sparse representation of natural light fields, an optical design that allows for capturing optimized 2D light field projections, and robust sparse reconstruction methods to recover a 4D light field from a single coded 2D projection. In addition, we demonstrate a variety of other applications for light field atoms and sparse coding techniques, including 4D light field compression and denoising.

Keywords: computational photography, light fields, compressive sensing

Links: DL PDF WEB VIDEO DATA CODE

1 Introduction

Since the invention of the first cameras, photographers have been striving to capture moments on film. Today, camera technology is on the verge of a new era. With the advent of mobile digital photography, consumers can easily capture, edit, and share moments with friends online. Most recently, light field photography was introduced to the consumer market as a technology facilitating novel user experiences, such as digital refocus, and 3D imaging capabilities, thereby capturing moments in greater detail. The technological foundations of currently available light field cameras, however, are more than a century old and have not fundamentally changed in that time. Most currently available devices trade spatial resolution for the ability to capture different views of a light field, oftentimes reducing the final image resolution by orders of magnitude compared to the raw sensor resolution. Unfortunately, this trend directly counteracts increasing resolution demands of the industry—the race for megapixels being the most significant driving factor of camera technology in the last decade.

We propose a computational light field camera architecture that allows for high resolution light fields to be reconstructed from a single coded camera image. This is facilitated by exploring the co-design of camera optics and compressive computational processing; we give three key insights into both optical and computational camera design parameters. First, the fundamental building blocks of natural light fields—light field atoms—can be captured in dictionaries that represent such high-dimensional signals more sparsely than previous representations. Second, this sparsity is directly exploited by nonlinear sparse coding techniques that allow high-resolution light fields to be reconstructed from a single coded projection. Third, the optical system can be optimized to provide incoherent measurements, thereby optically preserving the information content of light field atoms in the recorded projections and improving the reconstruction process.

Figure 2 illustrates these insights for a synthetic light field. Approximated with only a few coefficients, the 4D light field atoms introduced in this paper provide better quantitative compression for this example than previously employed basis representations (second row). Qualitatively, we compare compressibility of a small 4D light field patch with the discrete cosine transform (DCT) and with light field atoms (third row); the former cannot capture edges and junctions, which are crucial for applications such as refocus. Finally, we simulate a single-shot coded projection on a 2D sensor followed by sparse reconstruction with both DCT and our dictionaries (bottom row)—light field atoms significantly improve reconstruction quality.

1.1 Benefits and Contributions

We explore compressive light field photography and evaluate optical and computational design parameters. In particular, we make the following contributions:

- We propose compressive light field photography as a system combining optically-coded light field projections and nonlinear computational reconstructions that utilize overcomplete dictionaries as a sparse representation of natural light fields.
- We introduce light field atoms as the essential building blocks of natural light fields; these atoms are not only useful for high-resolution light field reconstruction from coded projections but also for compressing and denoising 4D light fields.
- We analyze existing compressive light field cameras and evaluate sparse representations for such high-dimensional signals. We demonstrate that the proposed atoms combined with optimized optical codes allow for light field reconstruction from a single photograph.
- We build a prototype compressive light field camera and demonstrate successful recovery of partially-occluded environments, refractions, reflections, and animated scenes.

1.2 Overview of Limitations

The proposed acquisition setup requires a coded attenuation mask between sensor and camera lens. As any mask-based camera system, this optical coding strategy sacrifices light transmission in the capture process. Our computational camera architecture requires significantly increased processing times compared to previous light field cameras. We demonstrate successful light field reconstructions from a single photograph; taking multiple shots, however, further increases image fidelity. Finally, light field atoms are only guaranteed to sparsely represent light fields that exhibit sufficiently similar structures as the training set that they were learned from.

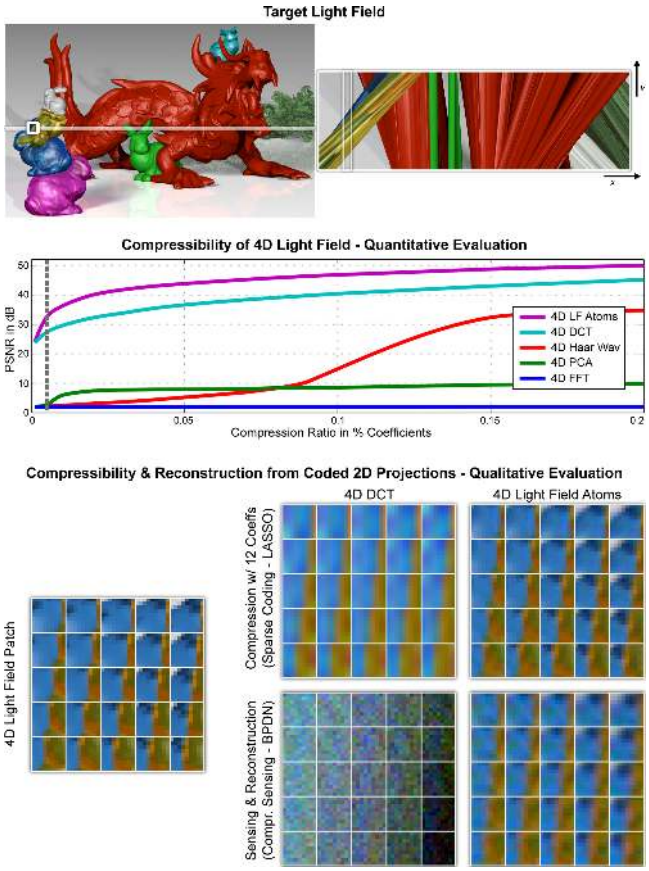

Figure 2: *Compressibility of a 4D light field in various high-dimensional bases. As compared to popular basis representations, the proposed light field atoms provide better compression quality for natural light fields (plots, second row). Edges and junctions are faithfully captured (third row); for the purpose of 4D light field reconstruction from a single coded 2D projection, the proposed dictionaries combined with sparse coding techniques perform best in this experiment (bottom row).*

2 Related Work

Light Field Acquisition Capturing light fields has been an active research area for more than a century. Ives [1903] and Lippmann [1908] were the first to realize that the light field inside a camera can be recorded by placing pinhole or lenslet arrays in front of a film sensor. Recently, lenslet-based systems have been integrated into digital cameras [Adelson and Wang 1992; Ng et al. 2005]; consumer products are now widely available. Light-modulating codes in mask-based systems have evolved to be more light efficient than pinhole arrays [Veeraraghavan et al. 2007; Lanman et al. 2008; Wetzstein et al. 2012a]. Nevertheless, all of these approaches sacrifice image resolution—the number of sensor pixels is the upper limit of the number of light rays captured. Within these limits [Georgiev and Lumsdaine 2006; Levin et al. 2008], alternative designs have been proposed that favor spatial resolution over angular resolution [Lumsdaine and Georgiev 2009]. In order to fully preserve image resolution, current options include camera arrays [Wilburn et al. 2005] or taking multiple photographs with a single camera [Levoy and Hanrahan 1996; Gortler et al. 1996; Liang et al. 2008]. While time-sequential approaches are limited to static scenes, camera arrays are costly and usually bulky. We present a compressive light field camera design that requires only a single photograph to recover a high-resolution light field.

Compressive Computational Photography Compressive sensing has been applied to video acquisition [Wakin et al. 2006; Marcia and Willett 2008; Hitomi et al. 2011; Reddy et al. 2011] and light transport acquisition [Peers et al. 2009; Sen and Darabi 2009]. The idea of compressive light field acquisition itself is not new, either. Kamal et al. [2012] and Park and Wakin [2012], for instance, simulate a compressive camera array. It could also be argued that light field superresolution [Bishop et al. 2009] is a form of compressive light field acquisition; higher-resolution information is recovered from microlens-based measurements under Lambertian scene assumptions. The fundamental resolution limits of microlens cameras, however, are depth-dependent [Perwass and Wietzke 2012]. We show that mask-based camera designs are better suited for compressive light field sensing and derive optimized single-device acquisition setups.

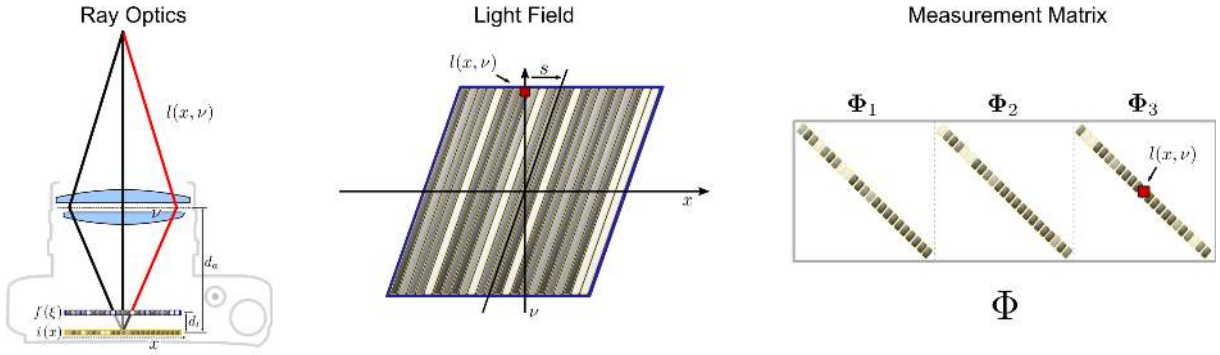

Figure 3: *Illustration of ray optics, light field modulation through coded attenuation masks, and corresponding projection matrix. The proposed optical setup comprises a conventional camera with a coded attenuation mask mounted at a slight offset in front of the sensor (left). This mask optically modulates the light field (center) before it is projected onto the sensor. The coded projection operator is expressed as a sparse matrix $\mathbf{\Phi}$, here illustrated for a 2D light field with three views projected onto a 1D sensor (right).*

Recently, researchers have started to explore compressive light field acquisition with a single camera. Optical coding strategies include coded apertures [Ashok and Neifeld 2010; Babacan et al. 2012], coded lenslets [Ashok and Neifeld 2010], a combination of coded mask and aperture [Xu and Lam 2012], and random mirror reflections [Fergus et al. 2006]. Unfortunately, [Ashok and Neifeld 2010; Babacan et al. 2012] require multiple images to be recorded and are not suitable for reconstructing dynamic scenes, even though they succeed in reducing the number of shots compared to their non-compressive counterparts such as [Liang et al. 2008]. Fergus et al. [2006] require significant changes to the optical setup, so that conventional 2D images are difficult to be captured. The work by Xu and Lam [2012] is most closely related to ours. However, they only show simulated results and employ simple light field priors based on total variation (TV). Furthermore, they propose an optical setup using dual-layer masks, but their choice of mask patterns (random and sum-of-sinusoids) reduces the light efficiency of the optical system to less than 5%.

In this paper, we demonstrate that light field atoms captured in overcomplete dictionaries represent natural light fields more sparsely than previously employed bases. We evaluate a variety of light field camera architectures and show that mask-based approaches provide a good tradeoff between expected reconstruction quality and optical light efficiency; we derive optimized mask patterns with approx. 50% light transmission that allow for high-quality light field reconstructions from a single coded projection. Finally, we show how to recover a conventional 2D photograph from a mask-modulated sensor image.

3 Light Field Capture and Synthesis

This section introduces the mathematical foundations of coded optical light field acquisition, robust computational reconstruction, and learning fundamental building blocks of natural light fields.

3.1 Acquiring Coded Light Field Projections

An image $i(x)$ captured by a camera sensor is the projection of an incident spatio-angular light field $l(x,\nu)$ along its angular dimension ν over the aperture area $\mathcal{V}$:

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

We adopt a two-plane parameterization [Levoy and Hanrahan 1996; Gortler et al. 1996] for the light field where x is the 2D spatial dimension on the sensor plane and ν denotes the 2D position on the aperture plane at distance d_a (see Fig. 3, left). For brevity of notation, the light field in Equation 1 absorbs vignetting and other angle-dependent factors [Ng 2005]. We propose to insert a coded attenuation mask $f(\xi)$ at a distance d_l from the sensor, which optically modulates the light field prior to projection as

$$i(x) = \int_{\mathcal{V}} f(x + s(\nu - x))\, l(x,\nu)\, d\nu, \tag{2}$$

where $s = d_l/d_a$ is the shear of the mask pattern with respect to the light field (see Fig. 3, center). In discretized form, coded light field projection can be expressed as a matrix-vector multiplication:

$$\mathbf{i} = \mathbf{\Phi}\mathbf{l}, \quad \mathbf{\Phi} = \left[\mathbf{\Phi}_1\ \mathbf{\Phi}_2\ \cdots\ \mathbf{\Phi}_{p_\nu^2}\right], \tag{3}$$

where $\mathbf{i} \in \mathbb{R}^m$ and $\mathbf{l} \in \mathbb{R}^n$ are the vectorized sensor image and light field, respectively. All $p_\nu \times p_\nu$ angular light field views $\mathbf{l}_j$ $(j = 1 \dots p_\nu^2)$ are stacked in $\mathbf{l}$. Note that each submatrix $\mathbf{\Phi}_j \in \mathbb{R}^{m \times m}$ is a sparse matrix containing the sheared mask code on its diagonal (see Fig. 3, right). For multiple recorded sensor images, the individual photographs and corresponding measurement matrices are stacked in $\mathbf{i}$ and $\mathbf{\Phi}$.

The observed image $\mathbf{i} = \sum_j \mathbf{\Phi}_j \mathbf{l}_j$ sums the light field views, each multiplied with the same mask code but sheared by different amounts. If the mask is mounted directly on the sensor, the shear vanishes ($s = 0$) and the views are averaged. If the mask is located in the aperture ($s = 1$), the diagonals of each submatrix $\mathbf{\Phi}_j$ become constants which results in a weighted average of all light field views. In this case, however, the angular weights do not change over the sensor area. Intuitively, the most random, or similarly incoherent, sampling of different angular samples happens when the mask is located between sensor and aperture; we evaluate this effect in Section 4.6.

Equations 1–3 model a captured sensor image as the angular projection of the incident light field. These equations can be interpreted to either describe the entire sensor image or small neighborhoods of sensor pixels—2D patches—as the projection of the corresponding 4D light field patch. The sparsity priors discussed in the following sections exclusively operate on such small two-dimensional and four-dimensional patches.

3.2 Reconstructing Light Fields from Projections

The inverse problem of reconstructing a light field from a coded projection requires a linear system of equations (Eq. 3) to be in-

verted. For a single sensor image, the number of measurements is significantly smaller than the number of unknowns, i.e. $m \ll n$. We leverage sparse coding techniques to solve the ill-posed underdetermined problem. For this purpose, we assume that natural light fields are sufficiently compressible in some basis or dictionary $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{n \times d}$, such that

$$\mathbf{i} = \boldsymbol{\Phi}\mathbf{l} = \boldsymbol{\Phi}\boldsymbol{\mathcal{D}}\boldsymbol{\alpha}, \tag{4}$$

where most of the coefficients in $\boldsymbol{\alpha} \in \mathbb{R}^d$ have values close to zero. Inspired by recent advances in compressed sensing (e.g., [Donoho 2006; Candès and Wakin 2008]), we seek a robust solution to Equation 4 as

$$\begin{array}{ll} \underset{\{\boldsymbol{\alpha}\}}{\text{minimize}} & \|\boldsymbol{\alpha}\|_1 \\ \text{subject to} & \|\mathbf{i} - \boldsymbol{\Phi}\boldsymbol{\mathcal{D}}\boldsymbol{\alpha}\|_2 \leq \epsilon \end{array} \tag{5}$$

which is known as the basis pursuit denoise (BPDN) problem [Chen et al. 1998]. In general, compressive sensing techniques attempt to solve underdetermined systems by finding the sparsest coefficient vector $\boldsymbol{\alpha}$ that satisfies the measurements, i.e. the ℓ_2-norm of the residual is smaller than the sensor noise level ϵ. In practice, we solve the Lagrangian formulation of Equation 5 as

$$\underset{\{\boldsymbol{\alpha}\}}{\text{minimize}} \; \|\mathbf{i} - \boldsymbol{\Phi}\boldsymbol{\mathcal{D}}\boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_1 . \tag{6}$$

Assuming that the light field is k-sparse, that is it can be well represented by a linear combination of at most k columns in $\boldsymbol{\mathcal{D}}$, a lower bound on the required number of measurements m is $O(k \log(d/k))$ [Candès et al. 2011]. While Equation 6 is not constrained to penalize negative values in the reconstructed light field $\mathbf{l} = \boldsymbol{\mathcal{D}}\boldsymbol{\alpha}$, we have not observed any resulting artifacts in practice.

The two main challenges for any compressive computational photography method are twofold: a "good" sparsity basis has to be known and reconstruction times have to scale up to high resolutions. In the following, we show how to learn dictionaries of small light field atoms that sparsely represent natural light fields. A side effect of using light field atoms is that scalability is intrinsically addressed as follows: instead of attempting to solve a single, large optimization problem, many small and independent problems are solved simultaneously. As discussed in the following, light field atoms model local spatio-angular coherence in the 4D light field sparsely. Therefore, a small 4D light field patch is reconstructed from a 2D image patch centered around each sensor pixel. The recovered light field patches are merged into a single reconstruction. Performance is optimized through parallelization and quick convergence of each subproblem; the reconstruction time grows linearly with increasing sensor resolution.

3.3 Learning Light Field Atoms

Following recent trends in the information theory community (e.g., [Candès et al. 2011]), we propose to learn the fundamental building blocks of natural light fields—light field atoms—in overcomplete dictionaries. We consider 4D spatio-angular light field patches of size $n = p_x \times p_x \times p_\nu \times p_\nu$. Given a large set of such patches, randomly chosen from a collection of training light fields, we learn a dictionary $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{n \times d}$ as

$$\begin{array}{ll} \underset{\{\boldsymbol{\mathcal{D}},\mathcal{A}\}}{\text{minimize}} & \|\mathbf{L} - \boldsymbol{\mathcal{D}}\mathcal{A}\|_F \\ \text{subject to} & \forall j, \|\boldsymbol{\alpha}_j\|_0 \leq k \end{array} \tag{7}$$

where $\mathbf{L} \in \mathbb{R}^{n \times q}$ is a training set comprised of q light field patches and $\mathcal{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q] \in \mathbb{R}^{d \times q}$ is a set of k-sparse coefficient vectors. The Frobenius matrix norm is $\|\mathbf{X}\|_F^2 = \sum_{ij} x_{ij}^2$, the ℓ_0 pseudo-norm counts the number of nonzero elements in a vector, and k ($k \ll d$) is the sparsity level we wish to enforce.

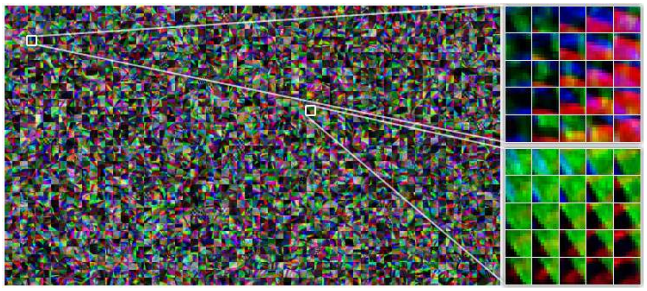

Figure 4: *Visualization of light field atoms captured in an overcomplete dictionary. Light field atoms are the essential building blocks of natural light fields—most light fields can be represented by the weighted sum of very few atoms. We show that light field atoms are crucial for robust light field reconstruction from coded projections and useful for many other applications, such as 4D light field compression and denoising.*

In practice, training sets for the dictionary learning process are extremely large and often contain a lot of redundancy. Solving Equations 7, however, is computationally expensive. *Coresets* have recently been introduced as a means to cheaply reduce large dictionary training sets to manageable sizes. Feigin et al. [2012], for instance, simply pick a subset of training samples in $\mathbf{L}$ that have a sufficiently high variance; we follow their approach.

4 Analysis

In this section, we analyze the structure of light field atoms and dictionaries, evaluate the design parameters of dictionaries, derive optimal modulation patterns for coded projections, evaluate the proposed camera architecture, and compare it with a range of alternative light field camera designs.

4.1 Interpreting Light Field Atoms

As discussed in Section 3.3, overcomplete dictionaries are learned from training sets of natural light fields. The columns of these dictionaries are designed to sparsely represent the respective training set, hence capture their essential building blocks or atoms. Obviously, the structure of these building blocks mainly depends on the specific training set; intuitively, large and diverse collections of natural light fields should exhibit some common structures, just like natural images. Based on recent insights, such as the dimensionality gap [Ng 2005; Levin et al. 2009], one would expect that the increased dimensionality from 2D images to 4D light fields introduces a lot of redundancy. The dimensionality gap is a 3D manifold in the 4D light field space, which successfully models diffuse objects within a certain depth range. Unfortunately, occlusions, specularities, and high-dimensional edges are not accounted for in this prior. In contrast, light field atoms do not model a specific lower-dimensional manifold, rather they sparsely represent the elemental structures of natural light fields.

Figure 4 visualizes an artificially-colored dictionary showing the central views of all its atoms; two of them are magnified and shown as 4D mosaics. We observe that light field atoms capture high-dimensional edges as well as high-frequency structures exhibiting different amounts of rotation and shear. Please note that these atoms also contain negative values; combining a few atoms allows complex lighting effects to be formed, such as reflections and refractions as well as junctions observed in occlusions (see Fig. 2).

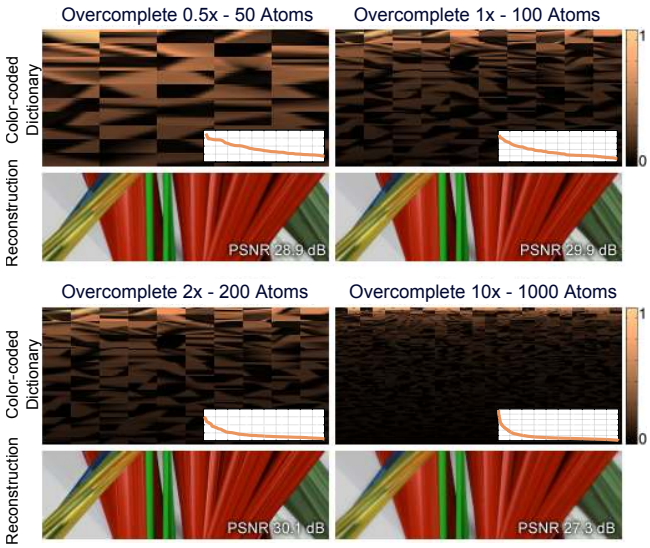

Figure 5: *Evaluating dictionary overcompleteness. Color-coded visualizations of dictionaries (first and third rows) and histograms of their coefficients used to represent the training data (insets). Reconstructions of a 2D light field (see Fig. 2) show that the quality (PSNR) for this example is best for $1-2\times$ overcompleteness and drops below and above. Dictionaries with less than $1\times$ overcompleteness do not perform well, because they simply do not contain enough atoms to sparsely represent the target light field whereas extremely overcomplete dictionaries contain many coefficients that are rarely used (see histograms).*

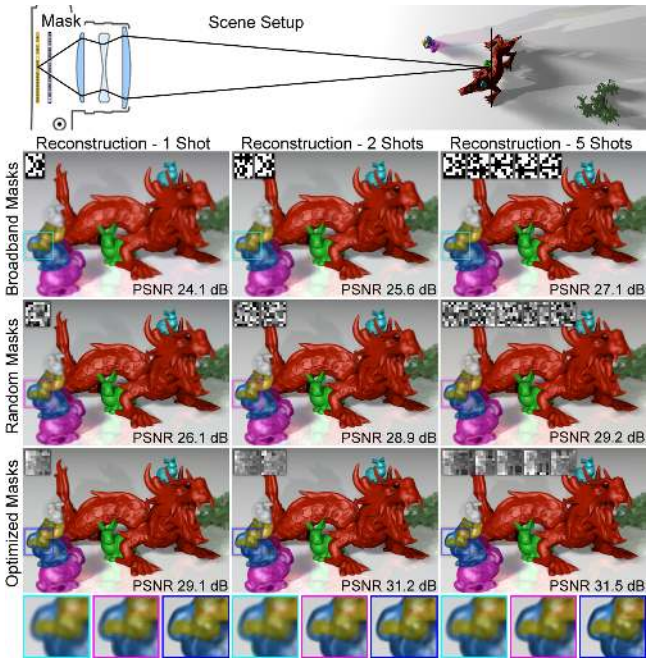

Figure 6: *Evaluating optical modulation codes and multiple shot acquisition. We simulate light field reconstructions from coded projections for one, two, and five captured camera images. One tile of the corresponding mask patterns is shown in the insets. For all optical codes, an increasing number of shots increases the number of measurements, hence reconstruction quality. Nevertheless, optimized mask patterns facilitate single-shot reconstructions with a quality that other patterns can only achieve with multiple shots.*

4.2 Evaluating Dictionary Design Parameters

Light Field Atom Size The size of light field atoms is an important design parameter. Consider an atom with $n = p_x^2 \times p_\nu^2$ pixels—the number of measurements is always $m = p_x^2$. Assuming a constant sparseness k of the light field in $\boldsymbol{\mathcal{D}}$, the number of measurements should follow the general rule $m \geq O(k \log(d/k))$ [Candès et al. 2011]. As the spatial atom size is increased for a fixed angular size and overcompleteness, the recovery problem becomes more well-posed because m grows linearly with the atom size, whereas the right hand side only grows logarithmically because d is directly proportional to n. On the other hand, an increasing atom size may decrease light field compressibility due to reduced local coherence within the atoms. Heuristically, we found $p_x = 11$ to be a good atom size for our applications.

Dictionary Overcompleteness We also evaluate how much overcompleteness dictionaries should have, that is how many atoms should be learned from a given training set. Conventional orthonormal bases in this unit are "$1\times$" overcomplete—$\boldsymbol{\mathcal{D}}$ is square ($d = n$). The overcompleteness of dictionaries, however, can be arbitrarily chosen in the learning process.

We evaluate overcompleteness for a "flatland" 2D example in Figure 5. The color-coded visualizations of the atoms in the respective dictionaries indicate how many times each of the atoms is actually being used to represent the training set (on a normalized scale). The histograms (insets) count how many times an atom was used to represent the training set. We observe that for a growing dictionary size, the redundancy grows as well. While all coefficients in the $0.5\times$ dictionary are being used almost equally often, for $10\times$ overcompleteness most of the coefficients are rarely being used. Since the dictionary size d is proportional to overcompleteness, there is a tradeoff between sparsity level k and dictionary size d. For a fixed number of measurements, an increasingly overcomplete dictionary must also sufficiently increase the sparsity of the signal to maintain or increase reconstruction quality. We found that $1-2\times$ overcomplete dictionaries adequately represent this particular training set while providing a good tradeoff between sparsity and dictionary size; all atoms have a resolution of 5×20 in angle and space. While this is only a heuristic experiment, we use it as an indicator for choosing the overcompleteness of light field dictionaries used in the experiments shown in this paper.

4.3 What are Good Modulation Patterns?

The proposed optical setup consists of a conventional camera with a coded attenuation mask mounted in front of the sensor. A natural question emerges: what should the mask patterns be? In the compressive sensing literature, most often dense sensing matrices of random Gaussian noise are employed. The proposed optical setup, however, restricts the measurement matrix $\boldsymbol{\Phi}$ to be very sparse (see Fig. 3, right). In the following, we discuss several choices of mask codes with respect to both computational and optical properties. A good mask design should facilitate high quality reconstructions while also providing a high light transmission.

Tiled Broadband Codes Broadband codes, such as arrays of pinholes, sum-of-sinusoids (SoS), or MURA patterns, are common choices for light field acquisition with attenuation masks. These patterns are designed to multiplex angular light information into the spatial sensor layout; under bandlimited assumptions, the 4D light field is reconstructed using linear demosaicking [Wetzstein

et al. 2012a]. In previous applications, the number of reconstructed light field elements is limited to the number of sensor pixels. The proposed nonlinear framework allows for a larger number of light rays to be recovered than available sensor pixels; however, these could also be reconstructed from measurements taken with broadband codes and the sparse reconstruction algorithms proposed in this paper. We evaluate such codes in Figure 6 and show that the achieved quality is lower than for random or optimized masks.

Random Mask Patterns For high resolutions, random measurement matrices provide incoherent signal projections with respect to most sparsity bases, including overcomplete dictionaries, with a high probability. This is one of the main reasons why random codes are by far the most popular choice in compressive sensing applications. In our application, the structure of the measurement matrix is dictated by the optical setup—it is extremely sparse. Each sensor pixel integrates over only a few incident light rays, hence the corresponding matrix row only has that many non-zero entries. While random modulation codes are a popular choice in compressive computational photography applications, these are not necessarily the best choice for overcomplete dictionaries, as shown in the following.

Optimizing Mask Patterns Most recently, research has focused on deriving optimal measurement matrices for a given dictionary [Duarte-Carvajalino and Sapiro 2009]. The intuition here is that projections of higher-dimensional signals should be as orthogonal as possible in the lower-dimensional projection space. Poor choices of codes would allow high-dimensional signals to project onto the same measurement, whereas optimal codes remove such ambiguities as best as possible. Mathematically, this optimality criterion can be expressed as

$$\begin{aligned} \underset{\{\mathbf{f}\}}{\text{minimize}} \quad & \left\|\mathbf{I} - \mathbf{G}^T\mathbf{G}\right\|_F \\ \text{subject to} \quad & 0 \le f_i \le 1,\ \forall i \\ & \textstyle\sum_i f_i/m \ge \tau \end{aligned} \tag{8}$$

where $\mathbf{G}$ is $\mathbf{\Phi}\mathcal{D}$ with normalized columns and $\mathbf{f} \in \mathbb{R}^m$ is the mask pattern along the diagonals of the submatrices in $\mathbf{\Phi}$ (see Fig. 3, right). Hence, each column of $\mathbf{G}$ is the normalized projection of one light field atom into the measurement basis. The individual elements of $\mathbf{G}^T\mathbf{G}$ are inner products of each of these projections, hence measuring the distance between them. Whereas diagonal elements of $\mathbf{G}^T\mathbf{G}$ are always one, the off-diagonal elements correspond to mutual distances between projected light field atoms. To maximize these distances, the objective function attempts to make $\mathbf{G}^T\mathbf{G}$ as close to identity as possible. To further optimize for light efficiency of the system, we add an additional constraint τ on the mean light transmission of the mask code $\mathbf{f}$.

4.4 Are More Shots Better?

We strongly believe that the most viable light field camera design would be able to reconstruct a high-quality and high-resolution light field from a single photograph. Nevertheless, it may be argued that more measurements may give even better results. This argument is supported by the experiments shown in Figure 6, where we evaluate multi-shot reconstructions for different mask patterns. In all cases, quality measured in peak signal-to-noise ratio (PSNR) is improved for an increasing number of shots, each captured with a different mask pattern. However, after a certain number of shots reconstruction quality is not significantly increased further—in the shown experiment, the gain from two to five shots is rather low. We also observe that a "good" choice of modulation codes equally improves reconstruction quality. In particular, optimized mask patterns allow for a single-shot reconstruction quality that can only be achieved with multiple shots otherwise. Capturing multiple shots with optimized mask patterns does not significantly improve image quality.

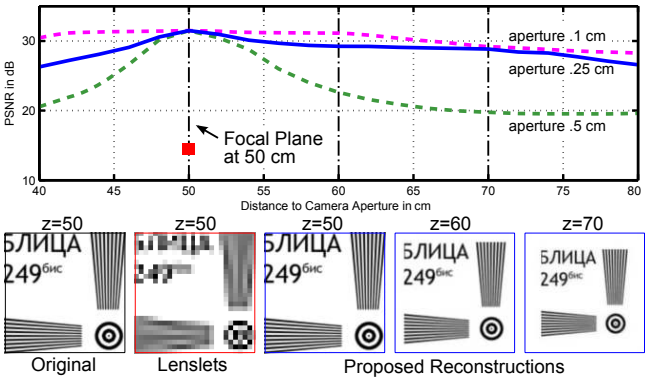

Figure 7: *Evaluating depth of field. As opposed to lenslet arrays, the proposed approach preserves most of the image resolution at the focal plane. Reconstruction quality, however, decreases with distance to the focal plane. Central views are shown (on focal plane) for full-resolution light field, lenslet acquisition, and compressive reconstruction; compressive reconstructions are also shown for two other distances. The three plots evaluate reconstruction quality for varying aperture diameters with a dictionary learned from data corresponding to the blue plot (aperture diameter 0.25 cm).*

4.5 Evaluating Depth of Field

We evaluate the depth of field achieved with the proposed method in Figure 7. For this experiment, we render light fields containing a single planar resolution chart at different distances to the camera's focal plane (located at 50 cm). Each light field has a resolution of 128×128 pixels and 5×5 views. The physical distances correspond to those in our camera prototype setup described in Section 5.1. While the reconstruction quality is high when the chart is close to the focal plane, it decreases with an increasing distance. Compared to capturing this scene with a lenslet array, however, the proposed approach results in a significantly increased image resolution.

The training data for this experiment contains white planes with random text at different distances to the focal plane, rendered with an aperture diameter of 0.25 cm. Whereas parallax within the range of the training data can be faithfully recovered (magenta and blue plots), a drop in reconstruction quality is observed when parallax exceeds that of the training data (green plot).

4.6 Comparing Computational Light Field Cameras

Two criteria are important when comparing different light field camera designs: optical light efficiency and expected quality of computational reconstruction. Light efficiency is measured as the mean light transmission of the optical system τ, whereas the value $\mu = \left\|\mathbf{I} - \mathbf{G}^T\mathbf{G}\right\|_F$ quantifies the expected reconstruction quality based on Equation 8 (lower value is better).

We compare lenslet arrays [Lippmann 1908; Ng et al. 2005], randomly coded lenslet arrays and coded apertures [Ashok and Neifeld 2010], coded broadband masks [Ives 1903; Veeraraghavan et al. 2007; Lanman et al. 2008] (we only show URA masks as the best general choice for this resolution), random masks and optimized masks, as proposed in this paper, as well as randomly coded apertures combined with a coded mask [Xu and Lam 2012]. All optical camera designs are illustrated in Figure 8. Optically, lenslet arrays

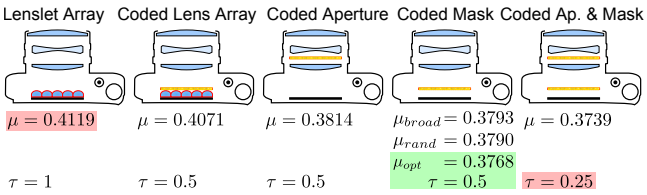

Figure 8: *Illustration of different optical light field camera setups with a quantitative value μ for the expected reconstruction quality (lower value is better). While lenslet arrays have the best light transmission τ (higher value is better), reconstructions are expected to be of lower quality. Masks coded with random or optimized patterns perform best of all systems with 50% or more transmission. Two masks are expected to perform slightly better with our reconstruction, but at the cost of reduced light efficiency.*

perform best with little loss of light; most mask-based designs have a light transmission of approx. 50%, except for pinholes. Combining randomly coded apertures with a modulation mask results in an overall transmission of about 25%, although Xu and Lam's choice of sum-of-sinusoids masks results in transmissions of less than 5%.

Figure 8 also shows the quantitative value μ for expected quality. Under this aspect, lenslet arrays perform worst, followed by coded lenslet arrays, coded apertures, and previously proposed tiled broadband codes (μ_{broad}). Random modulation masks (μ_{rand}) and the optimized patterns (μ_{opt}) proposed in Section 4.3 have the best expected quality of all setups with a mean transmission of 50% or higher. Although the dual-layer design proposed by Xu and Lam [2012] has a lower μ value, their design is significantly less light efficient than ours. While the quantitative differences between μ-values of these camera designs are subtle, qualitative differences of reconstructions are much more pronounced, as shown in Figures 6 and 7 and in the supplemental video. The discussed comparison is performed by assuming that all optical setups use the reconstruction method and overcomplete dictionaries proposed in this paper, as opposed to previously proposed PCA sparsity bases [Ashok and Neifeld 2010] or simple total variation priors [Xu and Lam 2012]. We hope that the optical and computational optimality criteria derived in this paper help find better optical camera configurations in the future.

5 Implementation

5.1 Hardware

For experiments with real scenes, it is necessary to easily change mask patterns for calibration and capturing training light fields. To this end, we implement a capture system using a liquid crystal on silicon (LCoS) display (SiliconMicroDisplay ST1080). An LCoS acts as a mirror where each pixel can independently change the polarization state of incoming light. In conjunction with a polarizing beam splitter and relay optics, as shown in Figure 9, the optical system emulates an attenuation mask mounted at an offset in front of the sensor. As a single pixel on the LCoS cannot be well resolved with the setup, we treat blocks of 4×4 LCoS pixels as macropixels, resulting in a mask resolution of 480×270. The SLR camera lens (Nikon 105 mm f/2.8D) is not focused on the LCoS but in front of it, thereby optically placing the (virtual) image sensor behind the LCoS plane. A Canon EF 50 mm f/1.8 II lens is used as the imaging lens and focused at a distance of 50 cm; scenes are placed within a depth range of 30–100 cm. The f-number of the system is the maximum of both lenses (f/2.8).

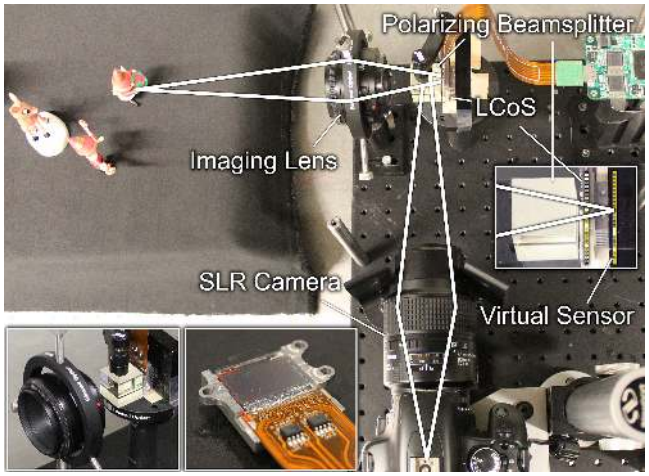

Figure 9: *Prototype light field camera. We implement an optical relay system that emulates a spatial light modulator (SLM) being mounted at a slight offset in front of the sensor (right inset). We employ a reflective LCoS as the SLM (lower left insets).*

Adjusting the Mask–Sensor Distance The distance d_l between the mask (LCoS plane) and the virtual image sensor is adjusted by changing the focus of the SLR camera lens. For capturing light fields with $p_\nu \times p_\nu$ angular resolution ($p_\nu = 5$ in our experiments), the distance is chosen as that of a conventional mask-based method that would result in the desired angular resolution albeit at lower spatial resolution [Veeraraghavan et al. 2007]. Specifically, we display a pinhole array on the LCoS where adjacent pinholes are p_ν macropixels apart while imaging a white calibration object. We then adjust the focus of the SLR camera lens so that disc-shaped blurred images under the pinholes almost abut each other. In this way, angular light field samples impinging on each sensor pixel pass through distinct macropixels on the LCoS with different attenuation values before getting integrated on the sensor.

Capturing Coded Light Field Projections We capture mask-modulated light field projections by displaying a pattern on the LCoS macropixels and resizing the sensor images accordingly.

Capturing Training Light Fields For the dictionary learning stage, we capture a variety of scenes using a traditional pinhole array. For this purpose, $p_\nu \times p_\nu$ $(= 25)$ images are recorded with shifting pinholes on the LCoS to obtain full-resolution light fields.

Calibration We measure the projection matrix $\mathbf{\Phi}$ by capturing the light field of a uniform white cardboard scene modulated by the mask pattern. This scene is captured in multiple shots with a shifting pinhole array on the LCoS, where each pinhole is additionally modulated by the corresponding mask value. The measurement matrix only has to be captured once.

5.2 Software

The algorithmic framework is a two step process involving an offline dictionary learning stage and a nonlinear reconstruction.

Dictionary Learning We capture five training light fields, each captured with an aperture setting of approx. 0.5 cm (f/2.8), with our prototype setup and randomly extract one million 4D light field patches, each with a spatial resolution of 11×11 pixels and 5×5

angular samples. After applying coreset reduction [Feigin et al. 2012], 50,000 remaining patches are used to learn a $1.7\times$ overcomplete dictionary consisting of 5,000 light field atoms. The memory footprint of this learned dictionary is about 111 MB. We employ the Sparse Modeling Software [Mairal et al. 2009] to learn this dictionary on an workstation equipped with a 24-core Intel Xeon processor and 200 GB RAM in about 10 hours. This is a one-time preprocessing step.

Sparse Reconstruction For the experiments discussed in Section 6, each light field is reconstructed with 5×5 views from a single sensor image with a resolution of 480×270 pixels. For this purpose, the coded sensor image is divided into overlapping 2D patches, each with a resolution of 11×11 pixels, by centering a sliding window around each sensor pixel. Subsequently, a small 4D light field patch is recovered for each of these windows. The reconstruction is performed in parallel on an 8-core Intel i7 workstation with 16 GB RAM. We employ the fast ℓ_1-relaxed homotopy method described by Yang et al. [2010] with sparsity penalizing parameter λ set to 10, tolerance to 0.001 and iterations to be 10,000; reconstructions for three color channels take about 18 hours for each light field. The reconstructed overlapping 4D patches are merged with a median filter.

Additional hardware and software implementation details, timings, and an evaluation of solvers for both dictionary learning and sparse reconstruction can be found in the supplemental material.

6 Results

All results discussed in this section are captured with our prototype compressive light field camera and reconstructed from a single sensor image. This image is a coded projection of the light field and the employed mask pattern is optimized for the computed dictionary with the technique described in Section 4.3. The same optical code and dictionary is used in all examples, the latter being learned from captured light fields that do not include any of the shown objects. All training sets and captured data are publicly available on the project website or upon request.

Layered Diffuse Objects Figure 10 shows results for a set of cards at different distances to the camera's focal plane. A 4D light field with 5×5 views (upper right) is reconstructed from a single coded projection (upper left). Parallax for out-of-focus objects is observed (center row). By shearing the 4D light field and averaging all views, a synthetically refocused camera image can be computed in post-processing (bottom row).

Partly-occluded Environments Reconstructed views of a scene exhibiting more complex structures are shown in Figure 11. The toy is partly occluded by a high-frequency shrub; occluded areas of the toy are faithfully reconstructed.

Reflections and Refractions Complex lighting effects, such as reflections and refractions exhibited by the dragon and the tiger in Figure 1, are successfully reconstructed with the proposed technique. In this scene, parts of the dragon are refracted through the head and shoulders of the glass tiger, whereas its back reflects and refracts the green background. We show images synthetically focused on foreground and background objects as well.

Animated Scenes The proposed algorithms allow 4D light fields to be recovered from a single 2D sensor image. Dynamic events can be recovered this way; to demonstrate this capability, we show several frames of an animated miniature carousel in Figure 12. Coded sensor images and reconstructions are shown.

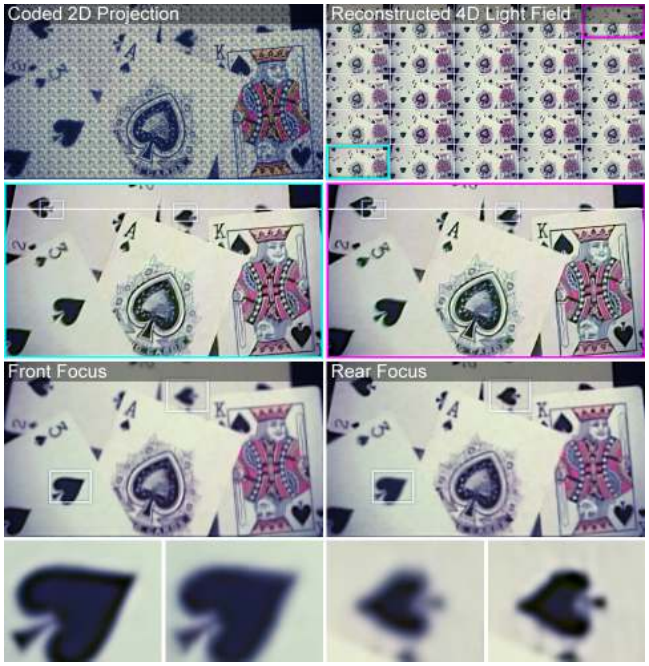

Figure 10: *Light field reconstruction from a single coded 2D projection. The scene is composed of diffuse objects at different depths; processing the 4D light field allows for post-capture refocus.*

Figure 11: *Reconstruction of a partly-occluded scene. Two views of a light field reconstructed from a single camera image. Areas occluded by high-frequency structures can be recovered by the proposed methods, as seen in the close-ups.*

7 Additional Applications

In this section, we outline a variety of additional applications for light field dictionaries and sparse coding techniques. In particular, we show applications in 4D light field compression and denoising. We also show how to remove the coded patterns introduced by modulation masks so as to retrieve a conventional 2D photograph. While an in-depth exploration of all of these applications is outside the scope of this paper, we hope to stimulate further research on related topics.

7.1 "Undappling" Images with Coupled Dictionaries

Although the optical acquisition setup proposed in this paper allows for light fields to be recovered from a single sensor image, a photographer may want to capture a conventional 2D image as well. In a commercial implementation, this could be achieved if the proposed optical system was implemented with programmable spatial light modulators or modulation masks that can be mechanically moved out of the optical path. As an alternative to optical solutions, we propose a computational approach to "undappling" a

Figure 12: *Light field reconstructions of an animated scene. We capture a coded sensor image for multiple frames of a rotating carousel (left) and reconstruct 4D light fields for each of them. The techniques explored in this paper allow for higher-resolution light field acquisition than previous single-shot approaches.*

Figure 13: *"Undappling" a mask-modulated sensor image (left). The known projection of the mask pattern can be divided out; remaining noise patterns in out-of-focus regions are further reduced using a coupled dictionary method (right).*

mask-modulated sensor image. For this purpose, we multiply the sensor image $\mathbf{i}$ by the inverse of the known 2D projection of the mask code $c_i = 1/\sum_j \phi_{ij}$ and process the resulting 2D image using sparse coding:

$$\underset{\{\boldsymbol{\alpha}\}}{\text{minimize}} \; \|\mathbf{i} \cdot \mathbf{c} - \boldsymbol{\mathcal{D}}_{dap}\boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_1 \tag{9}$$

Following [Yang et al. 2012], we learn a coupled dictionary $\boldsymbol{\mathcal{D}}_{2D} = \left[\boldsymbol{\mathcal{D}}_{dap}^T \; \boldsymbol{\mathcal{D}}_{undap}^T\right]^T$ from a training set containing projected light fields both with and without the mask patterns. One part of that dictionary $\boldsymbol{\mathcal{D}}_{dap}$ is used for reconstructing the 2D coefficients $\boldsymbol{\alpha}$, whereas the other is used to synthesize the "undappled" image as $\boldsymbol{\mathcal{D}}_{undap}\boldsymbol{\alpha}$. As opposed to the framework discussed in Section 3, the dictionaries and coefficients for this application are purely two-dimensional. Additional details of image "undappling" can be found in the supplemental document.

7.2 Light Field Compression

We illustrate compressibility of a 4D light field in Figure 2 both quantitatively and, for a single 4D patch, also qualitatively. Compression is achieved by finding the best representation of a light field with a fixed number of coefficients. This representation can be found by solving the LASSO [Natarajan 1995] problem

$$\begin{aligned} \underset{\{\boldsymbol{\alpha}\}}{\text{minimize}} \quad & \|\mathbf{l} - \boldsymbol{\mathcal{D}}\boldsymbol{\alpha}\|_2 \\ \text{subject to} \quad & \|\boldsymbol{\alpha}\|_0 \leq \kappa \end{aligned} \tag{10}$$

In this formulation, $\mathbf{l}$ is a 4D light field patch that is represented by at most κ atoms. As opposed to sparse reconstructions from coded 2D projections (Sec. 3), light field compression strives to reduce the required data size of a given 4D light field. As this technique

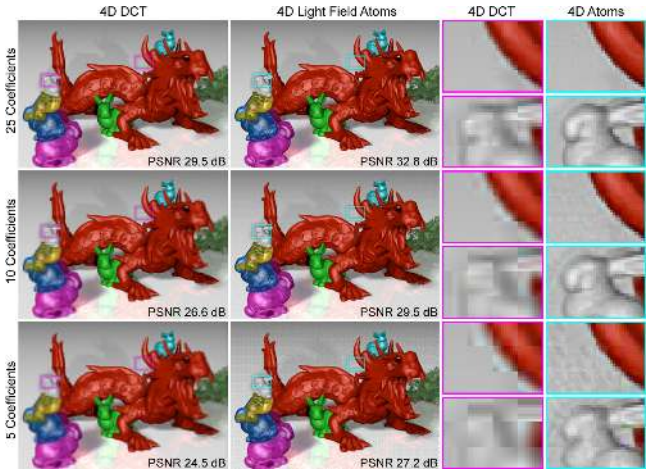

Figure 14: *Light field compression. A light field is divided into small 4D patches and represented by only few coefficients. Light field atoms achieve a higher image quality than DCT coefficients.*

is independent of the light field acquisition process, we envision future applications in high-dimensional data storage and transfer.

Figure 14 compares an example light field compressed into a fixed number of DCT coefficients and light field atoms. For this experiment, a light field with 5×5 views is divided into distinct $9 \times 9 \times 5 \times 5$ spatio-angular patches that are individually compressed. Light field atoms allow for high image quality with a low number of coefficients and smooth transitions between neighboring patches. While this experiment demonstrates improved compressibility of a single example light field using atoms, further investigation is required to analyze the suitability of overcomplete dictionaries for compressing a wide range of different light fields.

7.3 Light Field Denoising

Another popular application of dictionary-based sparse coding techniques is image denoising [Elad and Aharon 2006]. Following this trend, we apply sparse coding techniques to denoise 4D light fields. Similar to light field compression, the goal of denoising is not to reconstruct higher-resolution data from a smaller number of measurements, but to represent a given 4D light field by a linear combination of a small number of noise-free atoms. In practice,

Figure 15: *Light field denoising. Sparse coding and the proposed 4D dictionaries can remove noise from 4D light fields.*

this can be achieved in the same way as compression, i.e. by applying Equation 10. For the case of a noisy target light field $\mathbf{l}$, this effectively applies a nonlinear four-dimensional denoising filter to the light field.

Figure 15 shows the central view and close-ups of one row of a noisy 4D light field and its denoised representation. We hope that 4D light field denoising will find applications in emerging commercial light field cameras, as this technique is independent of the proposed compressive reconstruction framework and could be applied to light fields captured with arbitrary optical setups.

8 Discussion

In summary, this paper explores compressive light field acquisition by analyzing and evaluating sparse representations of natural light fields, optimized optical coding strategies, robust high-dimensional light field reconstruction from lower-dimensional coded projections, and additional applications such as 4D light field compression and denoising.

Compressive light field acquisition is closely related to emerging compressive light field displays [Wetzstein et al. 2011; Lanman et al. 2011; Wetzstein et al. 2012b]. These displays are compressive in the sense that the display hardware has insufficient degrees of freedom to exactly represent the target light field and relies on an optimization process to determine a perceptually acceptable approximation. Compressive cameras are constrained in their degrees of freedom to capture each ray of a light field and instead record coded projections with subsequent sparsity-exploiting reconstructions. We envision future compressive image acquisition and display systems to be a single, integrated framework that exploits the duality between computational light acquisition and display. Most recently, researchers have started to explore such ideas for display-adaptive rendering [Heide et al. 2013].

8.1 Benefits and Limitations

The primary benefits of the proposed computational camera architecture compared to previous techniques are increased light field resolution and a reduced number of required photographs. We show that reconstructions from coded light field projections captured in a single image can achieve a high quality; this is facilitated by the proposed co-design of optical codes, nonlinear reconstruction techniques, and sparse representations of natural light fields.

However, the achieved resolution of photographed objects decreases at larger distances to the camera's focal plane. Attenuation masks lower the light efficiency of the optical system as compared to refractive optical elements, such as lenslet arrays. Yet, they are less costly than lenslet arrays. The mask patterns are fundamentally limited by diffraction. Dictionaries have to be stored along with sparse reconstructions, thereby increasing memory requirements. Processing times of the discussed compressive camera design are higher than those of most other light field cameras. While these seem prohibitive at the moment, each small 4D patch is reconstructed independently; the computational routines discussed in this paper are well suited for parallel implementation, for instance on GPUs.

The camera prototype exhibits a number of artifacts, including angle-dependent color and intensity nonlinearities as well as limited contrast. Observed color shifts are intrinsic to the LCoS, due to birefringence of the liquid crystals; this spatial light modulator (SLM) is designed to work with collimated light, but we operate it outside its designed angular range so as to capture ground truth light fields for evaluation and dictionary learning. Current resolution limits of the captured results are imposed by the limited contrast of the LCoS—multiple pixels have to be binned. Simple coded transparencies or alternative SLMs could overcome these optical limitations in future hardware implementations.

Atoms captured in overcomplete dictionaries are shown to represent light fields more sparsely than other basis representations. However, these atoms are adapted to the training data, including its depth range, aperture diameter, and general scene structures, such as occlusions and high-frequency textures. We demonstrate that even a few training light fields that include reflections, refractions, texture, and occlusions suffice to reconstruct a range of scene types. Nevertheless, we expect reconstruction quality to degrade for scenes that contain structures not captured in the training data, as for instance shown for parallax exceeding that of the training data in Section 4.5. A detailed analysis of how target-specific light field atoms are w.r.t. all possible parameters, however, is left for future work.

8.2 Future Work

Our current prototype camera is designed as a multipurpose device capturing coded projections as well as reference light fields for dictionary learning and evaluating reconstructions. Future devices will decouple this process. Whereas coded projections can be recorded with conventional cameras enhanced by coded masks, the dictionary learning process will rely increasingly on large online datasets of natural light fields. These are likely to appear as a direct result of the commercial success of light field cameras on the consumer market. Such developments have two advantages. First, a larger range of different training data will make light field dictionaries more robust and better adapted to specific applications. Second, widely available dictionaries will fuel research on novel optical camera designs or commercial implementations of compressive light field cameras.

While we evaluate a range of existing light field camera designs and devise optimal coding strategies for them, we would like to explore new optical setups in the future. Evaluation with alternative error metrics to PSNR, such as perceptually-driven strategies, is an interesting avenue of future work. Finally, we plan to explore compressive acquisitions of the full plenoptic function, adding temporal and spectral light variation to the proposed framework. While this increases the dimensionality of the dictionary and reconstruction problem, we believe that exactly this increase in dimensionality will further improve compressibility and sparsity of the underlying visual signals.

9 Conclusion

The proposed compressive camera architecture is facilitated by the synergy of optical design and computational processing. We believe that the exploration of sparse representations of high-

dimensional visual signals has only just begun; fully understanding the latent structures of the plenoptic function, including spatial, angular, spectral, and temporal light variation, seems one step closer but still not within reach. Novel optical designs and improved computational routines both for data analysis and reconstruction will have to be devised, placing future camera systems at the intersection of scientific computing, information theory, and optics engineering. We believe that this paper provides many insights indispensable for future computational camera designs.

Acknowledgements

We thank the reviewers for valuable feedback and the following people for insightful discussions and support: Ashok Veeraraghavan, Rahul Budhiraja, Kaushik Mitra, Matthew O'Toole, Austin Lee, Sunny Jolly, Ivo Ihrke, Wolfgang Heidrich, Guillermo Sapiro, and Silicon Micro Display. Gordon Wetzstein was supported by an NSERC Postdoctoral Fellowship and the DARPA SCENICC program. Ramesh Raskar was supported by an Alfred P. Sloan Research Fellowship and a DARPA Young Faculty Award.

References

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

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

BABACAN, S., ANSORGE, R., LUESSI, M., MATARAN, P., MOLINA, R., AND KATSAGGELOS, A. 2012. Compressive Light Field Sensing. *IEEE Trans. Im. Proc. 21*, 12, 4746 –4757.

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

CANDÈS, E., AND WAKIN, M. B. 2008. An Introduction to Compressive Sampling. *IEEE Signal Processing 25*, 2, 21–30.

CANDÈS, E. J., ELDAR, Y. C., NEEDELL, D., AND RANDALL, P. 2011. Compressed Sensing with Coherent and Redundant Dictionaries. *Appl. and Comp. Harmonic Analysis 31*, 1, 59–73.

CHEN, S. S., DONOHO, D. L., MICHAEL, AND SAUNDERS, A. 1998. Atomic Decomposition by Basis Pursuit. *SIAM J. on Scientific Computing 20*, 33–61.

DONOHO, D. 2006. Compressed Sensing. *IEEE Trans. Inform. Theory 52*, 4, 1289–1306.

DUARTE-CARVAJALINO, J., AND SAPIRO, G. 2009. Learning to sense sparse signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization. *IEEE Trans. Im. Proc. 18*, 7, 1395–1408.

ELAD, M., AND AHARON, M. 2006. Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries. *IEEE Trans. Im. Proc. 15*, 12, 3736–3745.

FEIGIN, M., FELDMAN, D., AND SOCHEN, N. A. 2012. From High Definition Image to Low Space Optimization. In *Scale Space and Var. Methods in Comp. Vision*, vol. 6667, 459–470.

FERGUS, R., TORRALBA, A., AND FREEMAN, W. T. 2006. Random Lens Imaging. Tech. Rep. TR-2006-058, MIT.

GEORGIEV, T., AND LUMSDAINE, A. 2006. Spatio-angular Resolution Tradeoffs in Integral Photography. *Proc. EGSR*, 263–272.

GORTLER, S., GRZESZCZUK, R., SZELINSKI, R., AND COHEN, M. 1996. The Lumigraph. In *Proc. ACM SIGGRAPH*, 43–54.

HEIDE, F., WETZSTEIN, G., RASKAR, R., AND HEIDRICH, W. 2013. Adaptive Image Synthesis for Compressive Displays. *ACM Trans. Graph. (SIGGRAPH) 32*, 4, 1–11.

HITOMI, Y., GU, J., GUPTA, M., MITSUNAGA, T., AND NAYAR, S. K. 2011. Video from a Single Coded Exposure Photograph using a Learned Over-Complete Dictionary. In *Proc. IEEE ICCV*.

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

KAMAL, M., GOLBABAEE, M., AND VANDERGHEYNST, P. 2012. Light Field Compressive Sensing in Camera Arrays. In *Proc. ICASSP*, 5413 –5416.

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.

LANMAN, D., WETZSTEIN, G., HIRSCH, M., HEIDRICH, W., AND RASKAR, R. 2011. Polarization Fields: Dynamic Light Field Display using Multi-Layer LCDs. *ACM Trans. Graph. (SIGGRAPH Asia) 30*, 1–9.

LEVIN, A., FREEMAN, W. T., AND DURAND, F. 2008. Understanding Camera Trade-Offs through a Bayesian Analysis of Light Field Projections. In *Proc. ECCV*, 88–101.

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.

MAIRAL, J., BACH, F., PONCE, G., AND SAPIRO, G. 2009. Online Dictionary Learning For Sparse Coding. In *International Conference on Machine Learning*.

MARCIA, R. F., AND WILLETT, R. M. 2008. Compressive coded aperture video reconstruction. In *EUSIPCO*.

NATARAJAN, B. K. 1995. Sparse Approximate Solutions to Linear Systems. *SIAM J. Computing 24*, 227–234.

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., Stanford University.

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

PARK, J. Y., AND WAKIN, M. B. 2012. A geometric approach to multi-view compressive imaging. *EURASIP Journal on Advances in Signal Processing 37*.

PEERS, P., MAHAJAN, D. K., LAMOND, B., GHOSH, A., MATUSIK, W., RAMAMOORTHI, R., AND DEBEVEC, P. 2009. Compressive Light Transport Sensing. *ACM Trans. Graph. 28*, 3.

PERWASS, C., AND WIETZKE, L. 2012. Single Lens 3D-Camera with Extended Depth-of-Field. In *Proc. SPIE 8291*, 29–36.

REDDY, D., VEERARAGHAVAN, A., AND CHELLAPPA, R. 2011. P2C2: Programmable Pixel Compressive Camera for High Speed Imaging. In *Proc. IEEE CVPR*, 329–336.

SEN, P., AND DARABI, S. 2009. Compressive Dual Photography. *Computer Graphics Forum 28*, 609–618.

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.

WAKIN, M. B., LASKA, J. N., DUARTE, M. F., BARON, D., SARVOTHAM, S., TAKHAR, D., KELLY, K. F., AND BARANIUK, R. G. 2006. Compressive imaging for video representation and coding. In *Picture Coding Symposium*.

WETZSTEIN, G., LANMAN, D., HEIDRICH, W., AND RASKAR, R. 2011. Layered 3D: Tomographic Image Synthesis for Attenuation-based Light Field and High Dynamic Range Displays. *ACM Trans. Graph. (SIGGRAPH)*.

WETZSTEIN, G., IHRKE, I., AND HEIDRICH, W. 2012. On Plenoptic Multiplexing and Reconstruction. *IJCV*, 1–16.

WETZSTEIN, G., LANMAN, D., HIRSCH, M., AND RASKAR, R. 2012. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (SIGGRAPH) 31*, 1–11.

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.

XU, Z., AND LAM, E. Y. 2012. A High-resolution Lightfield Camera with Dual-mask Design. In *Proc. SPIE 8500*, 85000U.

YANG, A., GANESH, A., SASTRY, S., AND MA, Y. 2010. Fast L1-Minimization Algorithms and An Application in Robust Face Recognition: A Review. Tech. rep., UC Berkeley.

YANG, J., WANG, Z., LIN, Z., COHEN, S., AND HUANG, T. 2012. Coupled Dictionary Training for Image Super-Resolution. *IEEE Trans. Im. Proc. 21*, 8, 3467–3478.