Inferring Light Fields from Shadows

Manel Baradad[1] Vickie Ye[1] Adam B. Yedidia[1] Frédo Durand[1]
William T. Freeman[1,2] Gregory W. Wornell[1] Antonio Torralba[1]
[1] Massachusetts Institute of Technology [2] Google Research
{mbaradad,vye16,adamy,fredo,billf,gww}@mit.edu, torralba@csail.mit.edu

Abstract

We present a method for inferring a 4D light field of a hidden scene from 2D shadows cast by a known occluder on a diffuse wall. We do this by determining how light naturally reflected off surfaces in the hidden scene interacts with the occluder. By modeling the light transport as a linear system, and incorporating prior knowledge about light field structures, we can invert the system to recover the hidden scene. We demonstrate results of our inference method across simulations and experiments with different types of occluders. For instance, using the shadow cast by a real house plant, we are able to recover low resolution light fields with different levels of texture and parallax complexity. We provide two experimental results: a human subject and two planar elements at different depths.

1. Introduction

Imagine you are peering into a room through an open doorway, and all you can see is a white wall and an object placed just in front of it. Although you cannot see the rest of the room directly, the object is casting complex shadows on the wall (see Fig. 1). What information can be deduced from the patterns these shadows make? Is recovering a 2D picture of the room possible? What about depth recovery?

The problem of imaging scenes that are not within direct line of sight has a wide range of applications, including search-and-rescue, anti-terrorism and transportation [2]. However, this problem is very challenging. Previous approaches to non-line-of-sight (NLoS) imaging have used time-of-flight (ToF) cameras or shadows cast by ambient light to count objects in the hidden scene [28], or to reconstruct approximate low-dimensional images [3, 23].

In this paper, we attempt a far more ambitious class of reconstructions: estimating the 4D light field produced by what remains hidden. The recovered light field would allow us to view the hidden scene from multiple perspectives and thereby recover parallax information that can tell us about the hidden scene's geometry.

Figure 1: Example scenario. An observer peers partially into a room, and she is only able to see two things: a plant and the shadow it casts on the wall. If the observer can fully characterize the geometry of the plant, what can she learn about the rest of the room by analyzing the shadow?

Previous NLoS imaging methods that rely on secondary reflections from a naturally illuminated hidden scene use the known imaging conditions to approximate a *transfer matrix*—a linear relationship between all elements in the scene to observations on the visible reflector. They then invert this transfer matrix while simultaneously incorporating prior knowledge about likely scene structure, yielding a transformation that takes these observations as input and returns a reconstruction of the scene.

Our approach is also modeled on these principles, but prior knowledge plays an even more important role. An image of the white wall only provides a 2D array of observations. If we are going to reconstruct a 4D array of light-field samples from a vastly smaller number of observations, we must rely much more heavily on our

2575-7075/18 $31.00 © 2018 IEEE
DOI 10.1109/CVPR.2018.00656

prior distribution over light fields to obtain meaningful reconstructions. In particular, we can exploit the fact that elements of real scenes have spectra that are concentrated at low frequencies [4]. This allows us to reduce the effective dimensionality of the resulting problem to the point of tractability.

Our work also suggests that complex occluders can help reveal disparity information about a scene, more so than simple ones. This means that a commonplace houseplant that has a complex structure may be a better computational light field camera than a simpler pinspeck occluder. This is consistent with conventional wisdom about occlusions derived from past work in coded-aperture imaging, such as in [15].

Even with these insights, our approach relies on having an accurate estimate of the occluder's geometry and direct illumination over what remains hidden. Nevertheless, we think that our approach provides a strong template for future progress in NLoS imaging and compelling evidence that the world around us is covered with subtle, but very rich, visual information.

2. Background

2.1. NLoS Imaging

Non-line-of-sight (NLoS) imaging is the study of how to infer information about objects that are not directly visible. Many approaches to NLoS imaging involve a combination of active laser illumination and ToF cameras [13, 19, 29]. These methods, called *active methods*, work by illuminating a point on the visible region that projects light into the hidden scene. Then, structure in the scene can be inferred from the time it takes for that light to return [12, 20, 22, 24]. These methods have been used to count hidden people [28], or to infer location, size and motion of objects [9, 11, 18].

Other recent approaches, called *passive methods*, rely on ambient light from the hidden scene or elsewhere for inference. These approaches range from using naturally-occurring pinholes or pinspecks [7, 23] to using edges [3] to resolve the scene. Our work can be thought of as the extension of these same principles to arbitrary known occluders. To our knowledge, this work is the first to demonstrate reconstructions of 2D images from arbitrary known occluders in an NLoS setting, let alone reconstructions of 4D light fields.

2.2. Light field reconstructions

There has been ample prior work on inference of the full light field function for directly-visible scenes. This problem is addressed in [6], [8], [14], [21], [26], [27], among others. Notably, the work of [21] estimates a full 4D light field from a single 2D image of the scene. However, this learning-based method is trained on very constrained domain-specific data and is unable to accurately extend to novel images. This past work, particularly [14], heavily informed our choice of prior over light fields, which we discuss at length in Sec. 5.

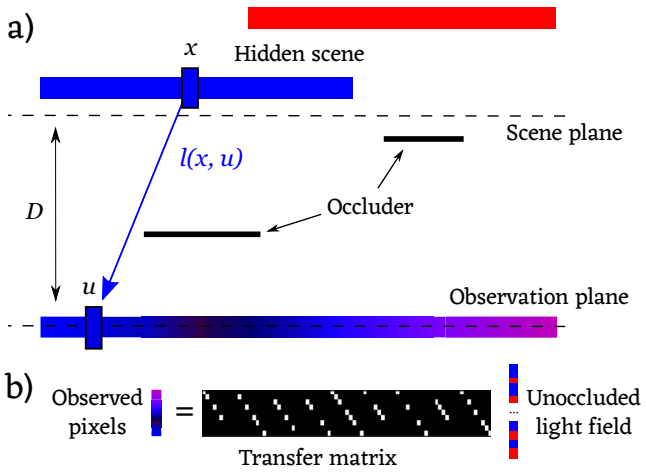

Figure 2: a) Simplified 2D scenario, depicting all the elements of the scene (occluder, hidden scene and observation plane) and the parametrization planes for the light field (dashed lines). (b) Discretized version of the scenario, with the light field and the observation encoded as the discrete vectors $\mathbf{x}$ and $\mathbf{y}$, respectively. The transfer matrix is a sparse, row-deficient matrix that encodes the occlusion and reflection in the system.

3. Overview

In order to reconstruct light fields using secondary reflections from the scene, our imaging method has two main components. The first is a linear forward model that computes observations from light fields, i.e. a transfer matrix $\mathbf{A}$, that has many columns but is sparse. The transfer matrix for an arbitrary scene is depicted schematically in Fig. 2.

The second is a prior distribution on light fields that allows reducing the effective dimensionality of the inverse problem, turning this ill-posed problem into one that is well-posed and computationally feasible. This strategy is better than other methods for reducing the dimensionality of the inverse problem (for example, naively downsampling the forward model and inverting). Light field sampling theory [5] and novel light field priors [14] inform how we reduce the dimensionality of the inverse problem given mild assumptions of the elements that produce the light field to be recovered.

In Section 4, we describe our model for how light propagates in the scene as well as our mathematical representations for each of its components. These choices

inform the design of our procedure for inferring the light field, which we explain in detail in Section 5. Finally, in Section 6, we present the results of our system on both simulated and real data.

4. Model

Our model of the system consists of three elements: a hidden scene, an occluder, and an observation plane that diffusely reflects light (see Fig. 2). We assume that the observation plane is fully visible and the geometry of the occluder is fully known, but that we do not have direct line of sight of any element of the hidden scene. For simplicity, we first formulate the model in a 2D world with 1D observations. The 2D analysis can be easily extended to the 4D case.

The hidden scene is assumed to be composed of mostly planar and parallel elements, which emit diffuse light with a band-limited spectrum. These elements are positioned at least a distance D away from the observation plane and at a maximum distance that is unknown but bounded. We assume that these elements are non-occlusive, i.e. that they do not block light coming from other elements of the scene.

The occluder is a set of non-reflective objects of arbitrary shape placed between both planes. It interacts with the light field by blocking some of the light rays from the scene. This produces a secondary light field, which we refer to as an *occluded light field*.

The observation plane is a uniform and Lambertian surface that diffusely reflects the light from the scene not blocked by the occluder. The 1D projection of the 2D light field on the plane (in the case of the 4D light field, we observe a 2D projection) constitutes our observation.

4.1. Light field parametrization

We parameterize the 2D light field $l(x, u)$ using the two-plane parametrization [16], as shown in Fig. 2. For simplicity, we place the angular plane (u) in the same position as the observation plane, and the scene plane (x) at a distance D from the observation plane and parallel to this. To simplify the analysis, we neglect the distance attenuation of light, and assume that the power of a ray remains constant traveling along a direction (x, u). In practice, we take these effects into consideration.

4.2. Occlusion

Without occluders, the scene alone would produce an unoccluded light field $l(x, u)$. When occluders are introduced, they generate a binary visibility function $v(x, u)$, which is 1 if x on the scene plane has an unobstructed view of u on the observation plane, and 0 otherwise. We express the *occluded light field* $l_{occ}(x, u)$ as follows:

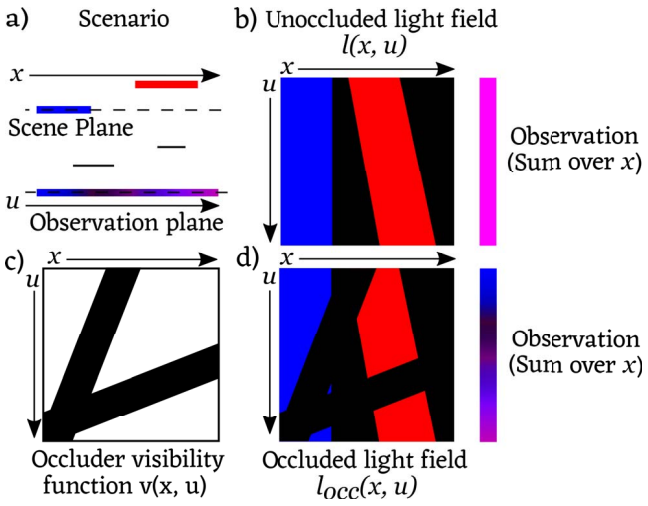

Figure 3: a) Sketch of a 2D imaging scenario. (b) Resulting unoccluded light field function $l(x, u)$ and observation. (c) The occluder visibility function $v(x, u)$. (d) Resulting occluded light field function $l_{occ}(x, u) = v(x, u)l(x, u)$ and observation. Note that the unoccluded observation is constant in u, which is not true of the occluded observation. The partial occlusion of the light field makes recovery of the scene possible from a diffuse observation.

$$l_{occ}(x, u) = v(x, u) \cdot l(x, u) \tag{1}$$

In the Fourier domain, this is equivalent to:

$$L_{occ} = \mathcal{F}(v(x, u) \cdot l(x, u)) = V(f_x, f_u) * L(f_x, f_u) \tag{2}$$

Thus, we can express the occluded light field spectrum as a convolution of the light field spectrum $L(x, u)$ with the visibility function spectrum $V(x, u)$. This allows recovering albedo information of the incoming light field from ideal diffuse reflections, as we explain in the following section.

4.3. Diffuse reflection

For a diffuse surface, the power reflected at each position u_0 on the observation plane can be expressed as

$$\begin{aligned} P(u_0) &= \int_x l_{occ}(x, u_0)dx \\ &= \int_x \int_u l_{occ}(x, u)\delta(u - u_0)dudx \\ &\overset{(1)}{=} \int_{f_x} \int_{f_u} L^*_{occ}(f_x, f_u)\mathcal{F}(\delta(u - u_0))df_u df_x \\ &= \int_{f_u} L^*_{occ}(0, f_u)e^{-j2\pi u_0 f_u} df_u \end{aligned} \tag{3}$$

where (1) follows from Parseval's theorem.

In Eq. 3, we see that nonzero spatial frequencies have no effect on the observation; that is, if the reflected light field is constant in u, the observed reflectance is uniform.

We show this effect for the 2D case in Fig. 3. For a diffuse surface, the observation at u integrates incoming light $l(x, u)$ from all x. In 3b, the unoccluded observation is uniform. In 3d, the occluder introduces variations in the observation that encodes scene information.

4.4. Discretization

We discretize the unoccluded light field $l(x, u)$ and visibility $v(x, u)$ as rasterized vectors $\mathbf{x}$ and $\mathbf{v}$. We construct both $\mathbf{x}$ and $\mathbf{v}$ using M uniform samples from the scene and N uniform samples from the observation plane.

As the number of samples is the same for both functions, it must meet the sampling requirements for both: those imposed by the expected bandwidth of the light field as well as those imposed by the bandwidth of the visibility function.

To ensure that the number of samples we need is finite, we assume that the visibility function is band-limited. We know that the unoccluded light field is band-limited because the depth of objects in the scene is assumed to be bounded and the emission spectrum of the scene elements is assumed to be band-limited.

In practice, we fix the amount of samples, taking into account practical considerations such as the calibration and capture methods (for real scenes) or the maximum resolution of the generated light fields (for simulations).

After discretization, we can express an observation $\mathbf{y}$ in terms of the unoccluded light field vector $\mathbf{x}$ as $\mathbf{y} = \mathbf{A}\mathbf{x}$ for the following $N \times MN$ transfer matrix $\mathbf{A}$ (schematically depicted in Fig. 2b):

$$\mathbf{A}_{k,Ni+j} = \begin{cases} c_{Ni+j} \cdot v_{Ni+j} & k = j \\ 0 & \text{else} \end{cases} \tag{4}$$

where c_{Ni+j} takes into account distance attenuation and other near-field effects. $\mathbf{A}$ encodes the fact that a diffuse reflector integrates over all the rays that arrive at a given observation point (when $k = j$) but only the rays that are not blocked by the occluder (when $v_{Ni+j} \neq 0$.)

5. Linear light field inference

5.1. Likelihood

Given an observation $\mathbf{y}$, we seek to find an estimate for the light field that maximizes the posterior probability $P(\mathbf{x}|\mathbf{y})$ (i.e. producing a MAP estimate $\hat{\mathbf{x}}$ for the light field), with $\mathbf{x}, \hat{\mathbf{x}} \in \mathbb{C}^{MN}$ and $\mathbf{y} \in \mathbb{C}^{N}$. We model $\mathbf{x}$ and $\mathbf{y}$ as complex variables to keep the notation straightforward when reformulating the problem in the Fourier domain.

We assume that our observation $\mathbf{y}$ is corrupted by additive white Gaussian noise. Thus, the likelihood of an observation $\mathbf{y}$ is:

$$P(\mathbf{y}|\mathbf{x}) \sim N(\mathbf{A}\mathbf{x}, \sigma^2\mathbf{I}) \tag{5}$$

Given the spectrum of $\mathbf{x}$ ($\mathbf{f}_\mathbf{x}$), expressed using the DFT matrix $\mathbf{F}$ as:

$$\mathbf{f}_\mathbf{x} = \mathbf{F}\mathbf{x} \tag{6}$$

$$\mathbf{f}_{\mathbf{x}\,Nk+l} = \sum_{m=1}^{M}\sum_{n=1}^{N} \mathbf{x}_{Nm+n} e^{-j2\pi\frac{m}{M}k} e^{-j2\pi\frac{n}{N}l} \tag{7}$$

The likelihood of $\mathbf{f}_\mathbf{x}$ is:

$$P(\mathbf{y}|\mathbf{f}_\mathbf{x}) \sim N(\mathbf{B}\mathbf{f}_\mathbf{x}, \sigma^2\mathbf{I}) \tag{8}$$

where:

$$\mathbf{B} = \mathbf{A}\mathbf{F}^{-1} \tag{9}$$

5.2. Prior on light fields

Thus, we can now express the MAP problem in terms of the spectral variables. Supposing that we use some zero-mean Gaussian prior on $\mathbf{f}_\mathbf{x}$ defined by its covariance matrix $\mathbf{C}_{\mathbf{f}_\mathbf{x}}$, the MAP estimate of the spectrum $\hat{\mathbf{f}}_\mathbf{x}$ can be expressed as:

$$P(\mathbf{f}_\mathbf{x}) \sim N(0, \mathbf{C}_{\mathbf{f}_\mathbf{x}}) \tag{10}$$

$$\hat{\mathbf{f}}_\mathbf{x} = \underset{\mathbf{f}_\mathbf{x}}{\operatorname{argmax}}\, P(\mathbf{f}_\mathbf{x}|\mathbf{y}) = \underset{\mathbf{f}_\mathbf{x}}{\operatorname{argmax}}\, P(\mathbf{y}|\mathbf{f}_\mathbf{x})P(\mathbf{f}_\mathbf{x}) \tag{11}$$

To find the $\mathbf{f}_\mathbf{x}$ that maximizes $P(\mathbf{f}_\mathbf{x}|\mathbf{y})$, we set the gradient of $P(\mathbf{f}_\mathbf{x}|\mathbf{y})$ over the conjugate of $\mathbf{f}_\mathbf{x}$ ($\mathbf{f}_\mathbf{x}^*$) to be 0:

$$\nabla_{\mathbf{f}_\mathbf{x}^*} P(\mathbf{f}_\mathbf{x}|\mathbf{y}) = 0 \tag{12}$$

$$\hat{\mathbf{f}}_\mathbf{x} = \left(\frac{1}{\sigma^2}\mathbf{B}^H\mathbf{B} + \mathbf{C}_{\mathbf{f}_\mathbf{x}}^H\mathbf{C}_{\mathbf{f}_\mathbf{x}}\right)^{-1}\mathbf{B}^H\mathbf{y} \tag{13}$$

Given the severely ill-posed nature of the light-field reconstruction problem, we have found the choice of prior, defined by $\mathbf{C}_{\mathbf{f}_\mathbf{x}}$, to be particularly important. Past (direct-line-of-sight) approaches to capturing light fields also rely heavily on a well-chosen prior for accurate reconstruction [6, 26, 27]. A few use plenoptic cameras, such as [1] and [17]. These cameras are designed to be able to capture the light rays reaching the aperture from different directions separately. Although prior distributions over the possible hidden scenes could be used to inform these sampling methods, these cameras usually sample uniformly across each dimension. For our application, this isotropic prior is sub-optimal.

The prior we propose instead is derived using the following simple assumptions: the scene is composed of mostly planar and diffuse elements that do not occlude each other, whose textures are produced by a random process that is independent of their position.

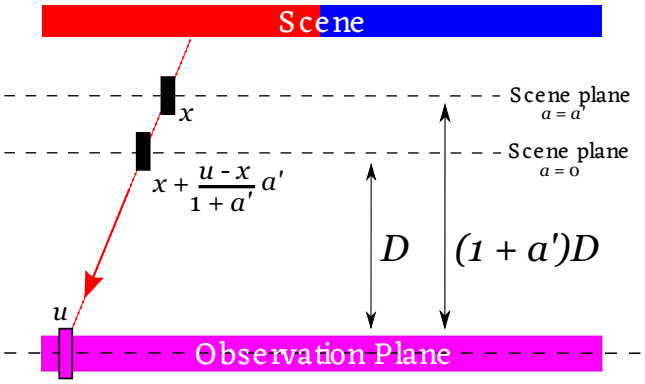

Figure 4: Light field reparametrization by a displacement of the scene plane. For any u, x, and a, $l_a(x,u) = l(x + \frac{u-x}{1+a}a, u)$.

These assumptions can be easily formulated in the Fourier domain and factored into two Gaussian and independent terms, which we refer to as a 3D manifold term (P_m) and a texture term (P_t), respectively. With this, the PDF of each spectrum coefficient $\mathbf{f}^i_\mathbf{x}$ is also Gaussian and independent, following:

$$P(\mathbf{f}^i_\mathbf{x}) \propto P_m(\mathbf{f}^i_\mathbf{x}) P_t(\mathbf{f}^i_\mathbf{x}) \tag{14}$$

Reparametrization of light fields. To characterize our assumptions, it is necessary to model light field reparametrization of the scene plane. Given a light field $l(x,u)$, we define the reparametrized light field $l_a(x,u)$ as that containing the same radiance as $l(x,u)$, but whose scene plane x is displaced a distance aD further away from the u plane as seen in Figure 4. The relation between l_a and $l(x,u)$ is:

$$l_a(x,u) = l\left(x + \frac{u-x}{1+a}a, u\right) \tag{15}$$

This reparametrization causes a shear of the spectrum in the angular domain and a scaling in the spatial domain (similar to the effect of a refocus, and first derived in [10]), following:

$$L_a(f_x, f_u) = (1+a)L((1+a)f_x, f_u - af_x) \tag{16}$$

Dimensionality gap of 4D light fields. The 2D light field parametrized at a' ($L_{a'}$) created by a single 1D diffuse texture at depth $a = a'$ has non-zero power only at $f_u = 0$ (as it is constant over u under this parametrization). Following the light field reparametrization derived in 16, the spectrum for this same light field but parametrized at depth $a = 0$, is a sheared (and scaled) version of $L_{a'}$. The spectrum L for this texture is thus non-zero only in the region corresponding to that shear (i.e. $f_x = a' f_u$)

Extending this to the 4D case with planar and parallel textures, the spectrum $L_{a'}(f_x, f_y, f_u, f_v)$ of a texture at depth $a = a'$ has non-zero power when both $f_u = 0$ and $f_x = 0$. The shear caused by the reparametrization L is proportional to a' in both dimensions, and follows:

$$\begin{aligned} f_x &= a' f_u \\ f_y &= a' f_v \end{aligned} \tag{17}$$

Considering only planar and parallel textures at any a, Eqs. 17 describe a 3D manifold of the full 4D light field space, as first derived in [14]. Using this relationship, we model spectrum coefficients outside this manifold as zero-mean Gaussian variables with low variance, while coefficients inside the manifold are modeled with high variance. If we know that the objects of interest within the scene lie within a known range of possible depths (that is, knowing the set of a's where the objects of interest are) we can further limit the coefficients with high variance using Eqs. 17.

Prior on textures. Our model assumes that the hidden scene consists of parallel textures, or planes of unknown albedo, at arbitrary depths. We model the textures of these planes as realizations of the same random process and assume that the process is independent of the planes' depths.

Following [4], we model the albedo of each texture as a realization of a random process with $1/f^\gamma$ spectral density (for $\gamma = 2$). Taking into account that these textures are diffuse, the light field l_a created by a single texture at the scene plane has power given by:

$$\left\{ \begin{array}{ll} 1/f_x^\gamma & \text{if } f_u = 0 \\ 0 & \text{else} \end{array} \right\} \tag{18}$$

Thus the probability density function for each spectral component of this light field is given by:

$$P(L_a(f_x, 0)) = N(0, 1/f_x^\gamma) \tag{19}$$

This equation expresses the fact that, if we had a light field created by a single planar element at depth a and we placed the scene plane at depth a, the equivalent light field $l_a(x,u)$ would consist of a typical texture with expected frequency distribution $1/f_x^\gamma$ over x, for all u.

Taking into account the reparametrization of light fields formulated in Eq. 16 and the relationship between a, f_x and f_u ($a = \frac{f_x}{f_u}$), we extend the principle of Eq. 20 to the whole light field, assuming no mutual occlusions between parallel textures occur (even though the real scene may have mutual occlusions).

$$L(f_x, f_u) = \frac{1}{1+a} L_a \left(\frac{1}{1+a} f_x, 0 \right)$$
$$P_t(L(f_x, f_u)) = N\left(0, \left(\frac{1}{1+\frac{f_x}{f_u}}\right)^{1-\gamma} f_x^{-\gamma}\right) \quad (20)$$

The previous probability density function is extended to the 4D case, assuming the spectral density of 2D textures is $(f_x^2 + f_y^2)^{-\gamma/2}$:

$$P_t(L(f_x, f_y, f_u, f_v)) = N\left(0, \left(\frac{1}{1+\frac{f_x}{f_u}}\right)^{1-\gamma} (f_x^2 + f_y^2)^{-\gamma/2}\right) \quad (21)$$

5.2.1 Inverse problem limitations

Finally, we find it informative to illustrate the ill-posed nature of the inverse problem, and how it is affected by the assumed depth range of the elements in the scene. It is also worth discussing the advantages and disadvantages of even attempting a 4-dimensional reconstruction as opposed to something more conservative.

First, we note that the system we propose can be easily used to obtain a single diffuse 2D reconstruction of the hidden scene at no additional cost. When the depth range at which we reconstruct is negligible compared to the distance to the first element, we can assume that the spectrum is non-zero only for $a = 0$, which corresponds to $f_u = f_v = 0$, derived from Eqs. 17. This effectively corresponds to a 2D reconstruction, where all the points in the observation plane see the same image.

As we increase the depth range at which elements can be placed, we give our model more expressive power with which to describe a wider range of scenarios, but we expose ourselves to an increasingly ill-posed problem—a variety of equally-plausible light fields, only one of which is true, all explaining the data equally well.

6. Experimental results

6.1. Practical considerations

Using the closed-form solution to the MAP problem derived in Eq. 13 for the full light-field spectrum is computationally infeasible. To overcome this, we solve the inference problem by assuming that only the K highest-variance spectral components of L have non-zero variance, and that the remaining $NM - K$ coefficients have exactly 0 variance.

To help prevent the appearance of artifacts caused by the elimination of spectral components, we assume that the original light field is zero-padded in the scene plane and mirror-padded in the observation plane. When we compute the Fourier-domain transfer matrix $\mathbf{B}$ from $\mathbf{A}$ (Eq. 9), we replace the inverse DFT matrix $\mathbf{F}^{-1}$ with its padded version.

For both real experiments and simulations, we reconstruct a volume with depth range between $a = 0$ and $a = 0.5$ (meaning that all elements of interest are assumed to be within a factor of 1.5 of the minimum depth D). This allows us to recover reasonable light fields for certain occluders (like spheres and plants). However, the enforced depth range yields poor reconstructions for some occluders (like the single pinhole) that work well for 2D images at known depth or in the far-field.

6.2. Simulations

For our simulations, we use an observation plane of 150×225 samples and scene plane of 30×30 samples. The complete light field thus has approximately 3×10^7 light ray components. We use a rendering engine to generate the light fields of simulated scenes at this resolution. However, during inference, we limit our reconstructions to $K = 10^4$ spectral components, for computational feasibility.

We compute the transfer matrix $\mathbf{A}$ using the fully known geometry of the occluder. For simple occluders (e.g. spheres), we compute the visibility $\mathbf{v}$ analytically; for complex occluders (e.g. plants) we do this by rendering the equivalent $\mathbf{v}$ from a CAD model of the occluder. These transfer matrices are then used to project generated light fields and generate the observations, to which we add Gaussian noise to achieve a peak signal-to-noise ratio (PSNR) of 20dB.

In Fig. 5, we consider a simple hidden scene of two squares, one green placed at a depth $a = 0$ (no parallax), and one red placed at depth $a = 0.3$ (some parallax). We compare the reconstruction quality of light fields recovered with different occluders. Our results indicate that the quality of the recovery method improves as we use more complex occluders. This is consistent with the findings of past work on coded-aperture cameras, which found that complicated occlusion patterns generally yield better reconstructions than simple ones [15, 30].

We further show in Fig. 7 the recovered light field of a more complex hidden scene consisting of a single head, for the best performing occluder previously tested (a set of plants).

6.3. Real experiments

Experimental Setup. For all experimental measures, we use a Canon EOS 6D with a Canon EF 24-105mm f/4L lens, set at ISO 100 and f/4.0. We construct $\mathbf{y}$ by averaging 5 RAW captures with 5s exposure and then subtracting a background image taken without the hidden scene in place.

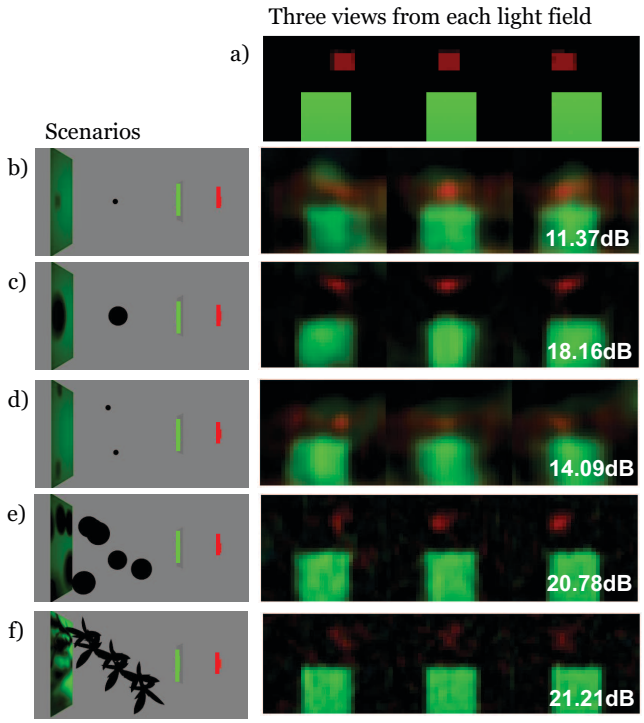

Figure 5: Recovered light fields for a simulated scene and different occluders. For each occluder, we provide three views of a horizontal slice of the recovered light field and the PSNR of these views. (a) Ground-truth. (b)-(f) Simulated scenario and recovered views for (b) a pinspeck, (c) a sphere, (d) two pinspecks at different depths, (e) multiple spheres, and (f) multiple plants.

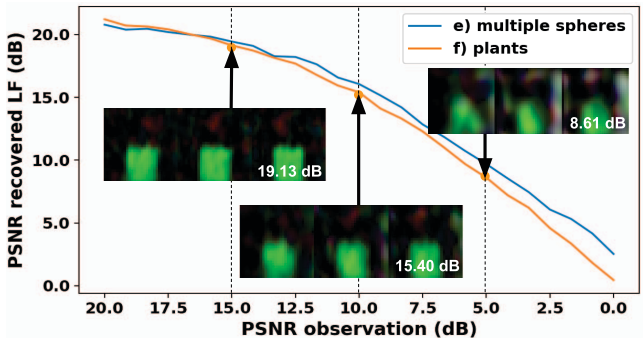

Figure 6: Reconstruction PSNR plotted against observation PSNR, for simulated scenarios (e) and (f) from Fig. 5. We also show three reconstructions of the scene using the setup from (f), for three different values of the PSNR.

The hidden scene is illuminated with two Lowel ID-Light 100W placed between both parametrization planes, one on each side. The observation is 24×36in, the scene plane is 14×23in and D is equivalent to 45in. The camera is

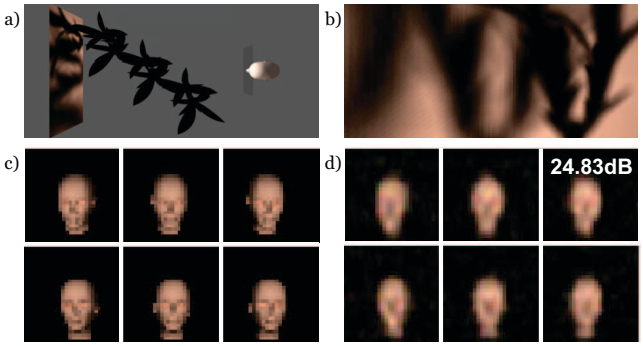

Figure 7: Recovered light field of another simulated scene, using the same occluder as in scenario (f) of Fig. 5. (a) Simulated scenario and (b) the resulting observation. In (c) we show selected views from the ground truth light field and in (d) the same views for the recovered light field. This illustrates the performance of our method for an arbitrary scene, with self-occlusions and non-planar elements.

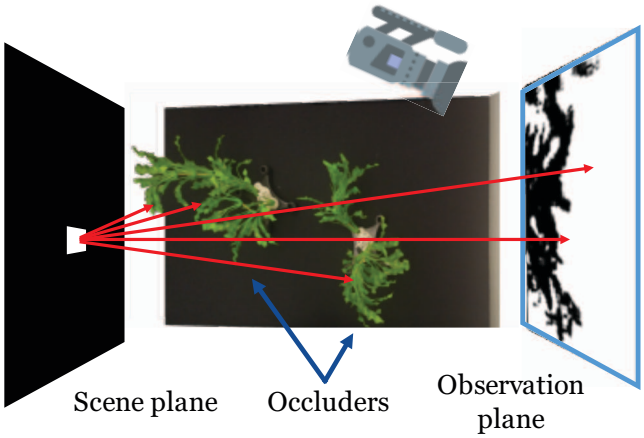

Figure 8: Experimental setup used to calibrate complex occluders. An impulse at a single position on an LCD screen placed at the scene plane casts a shadow. This shadow is used to compute the visibility function of that scene position and all observation positions. This process is done for each possible position in the scene.

positioned as seen in Fig. 8, having a full unoccluded view of the observation plane and at the height of the setup.

Calibration of the visibility function. In these experiments, we use two desk plants as occluders, as shown in Fig. 8. The geometry of these occluders is not readily available, and we must empirically measure each element of the visibility vector $\mathbf{v}$. To do so, we first capture the observation obtained by lighting up each location on the scene plane with an LCD monitor, and then convert these observations into binary masks by thresholding pixel

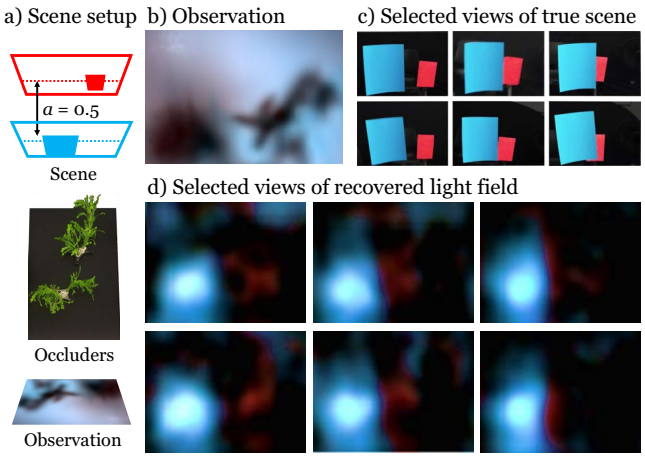

Figure 9: Reconstructions of an experimental scene with two rectangles, one blue at $a = 0$ and one red at $a = 0.5$. (a) Schematic of the setup. (b) Observation plane after background subtraction. (c) Six views of the true scene, shown to demonstrate of what the true light field would look like. These are taken with a standard camera from equivalent positions on the observation wall. (d) Reconstructions of the light field for these views. The blue and red targets measure 8×12in and 6×8in.

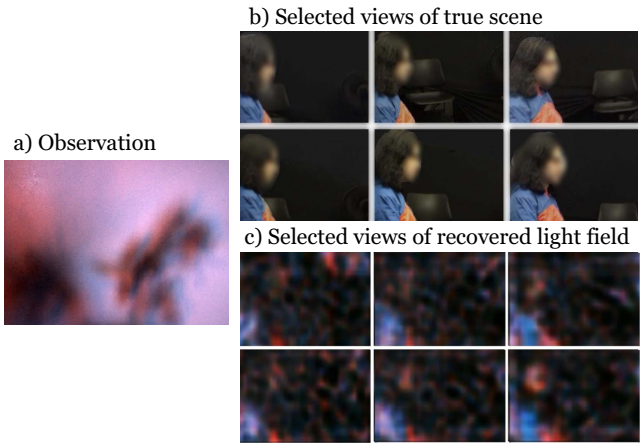

Figure 10: Reconstructions of an experimental scene with a seated subject at the scene plane. (a) Observation plane after background subtraction. (b) Six views of the true scene, captured in the same manner as in Fig. 9 (c) Reconstructions of the light field for the same views as in (b).

intensities. These masks approximately encode which light field rays reach the observation plane, and thus $\mathbf{v}$. The full transfer matrix is computationally derived according to Eq. 4 using the measured $\mathbf{v}$ and using $c_k = \cos^2(\theta)/r^2$ to model near-field effects, where θ is the angle between the light field ray k and the observation plane and r is the distance between the two points at which the ray intersects the parametrization planes. We note that directly using the captured observations to construct the full transfer matrix would capture the directivity spectrum of the calibration device (in our case, the LCD screen).

Real reconstructions. In Figs. 9 and 10, we show six selected views of recovered light fields for two hidden scenes, with different levels of spatial (texture-related) and angular (parallax-related) complexity. Both are obtained using the same prior, without adapting the depth range to recover the particulars of the scene. The results demonstrate consistent reconstruction of both texture and parallax information with respect to the expected light fields of the hidden scenes. In Fig. 9, we see the blue and red squares of the true scene. Moreover, we see the red square being increasingly hidden by the blue square as the "camera" moves right, and the red square moving lower relative to the blue square as the "camera" moves down. Similarly, in Fig. 10, we see the blue and orange of the man's coat, and as the "camera" moves right and down, we can see an increasing amount of his face.

7. Conclusions

Our proposed method allows linear inference of light fields in passive NLoS settings, using the geometry and position of a known occluder and plausible assumptions about typical light fields. Ours is the first method known to us to infer full light fields when the scene is totally hidden from direct view, and the first to use arbitrary occluders to reconstruct even 2D representations of scenes.

We believe that these methods are an exciting first step towards a richer class of NLoS reconstructions. Past efforts in NLoS imaging, including active approaches, have generally focused on learning only part of the information available in the scene. For instance, [11], [18], and [25] focus on extracting the size, motion, and shape of objects, while [3] attempts a 1D reconstruction of a moving scene. In contrast, our proposed method provides a reconstruction for the light field, which can then be used to approximate some of the previous methods, either computationally or by human inspection.

Our methods and experimental results demonstrate that the hidden patterns of shadows on walls can, if they are properly interpreted, reveal a rich world of previously invisible visual information.

8. Acknowledgements

This work was supported in part by the DARPA REVEAL Program under Contract No. HR0011-16-C-0030. We thank Jeffrey H. Shapiro, Franco N. C. Wong and Vivek K Goyal for helpful discussions.

References

[1] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:99–106, 1992. 4

[2] P. Borges, A. Tews, and D. Haddon. Pedestrian detection in industrial environments: Seeing around corners. *Intelligent Robots and Systems (IROS). IEEE/RSJ International Conference.*, 2012. 1

[3] K. Bouman, V. Ye, A. Yedidia, F. Durand, G. Wornell, A. Torralba, and W. T. Freeman. Turning corners into cameras: Principles and methods. *International Conference on Computer Vision*, 2017. 1, 2, 8

[4] G. Burton and I. R. Moorhead. Color and spatial structure in natural scenes applied optics 26 157-170. 26:157–70, 01 1987. 2, 5

[5] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum. Plenoptic sampling. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 307–318, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. 2

[6] W. chao Chen, J.-Y. Bouguet, M. H. Chu, and R. Grzeszczuk. Light field mapping: Efficient representation and hardware rendering of surface light fields. In *ACM Transactions on Graphics*, pages 447–456, 2002. 2, 4

[7] A. L. Cohen. Anti-pinhole imaging. *Optica Acta: International Journal of Optics*, 29(1):63–67, 1982. 2

[8] K. Egan, F. Hecht, F. Durand, and R. Ramamoorthi. Frequency analysis and sheared filtering for shadow light fields of complex occluders. *ACM Transactions on Graphics*, 30(2):9, 2011. 2

[9] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 2015. 2

[10] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 297–306, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. 5

[11] A. Kadambi, H. Zhao, B. Shi, and R. Raskar. Occluded imaging with time-of-flight sensors, 2016. ACM Transactions on Graphics. 2, 8

[12] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar. Looking around the corner using transient imaging, 09 2009. 2

[13] M. Laurenzis, A. Velten, and J. Klein. Dual-mode optical sensing: three-dimensional imaging and seeing around a corner. *Optical Engineering*, 2017. 2

[14] A. Levin and F. Durand. Linear view synthesis using a dimensionality gap light field prior. In *In Proc. IEEE CVPR*, pages 1–8, 2010. 2, 5

[15] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26(3):70, 2007. 2, 6

[16] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 31–42, New York, NY, USA, 1996. ACM. 3

[17] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. 2005. 4

[18] R. Pandharkar, A. Velten, A. Bardagjy, E. Lawson, M. Bawendi, and R. Raskar. Estimating motion and size of moving non-line-of-sight objects in cluttered environments, 2011. 2, 8

[19] D. Shin, A. Kirmani, V. Goyal, and J. Shapiro. Photon-efficient computational 3-d and reflectivity imaging with single-photon detectors. *IEEE Transactions on Computational Imaging*, 2015. 2

[20] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. Computational imaging with multi-camera time-of-flight systems. *ACM Transactions on Graphics (TOG)*, 2016. 2

[21] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng. Learning to synthesize a 4d rgbd light field from a single image. *International Conference on Computer Vision*, 2017. 2

[22] C. Thrampoulidis, G. Shulkind, F. Xu, W. T. Freeman, J. H. Shapiro, A. Torralba, F. N. C. Wong, and G. W. Wornell. Exploiting Occlusion in Non-Line-of-Sight Active Imaging. *ArXiv e-prints*, Nov. 2017. 2

[23] A. Torralba and W. T. Freeman. Accidental pinhole and pinspeck cameras. *International Journal of Computer Vision*, 110(2):92–112, Nov 2014. 1, 2

[24] C.-Y. Tsai, K. N. Kutulakos, S. G. Narasimhan, and A. C. Sankaranarayanan. The geometry of first-returning photons for non-line-of-sight imaging. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2

[25] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3(3):745, 2012. ACM Transactions on Graphics. 8

[26] B. Wilburn, N. Joshi, V. Vaish, E. ville Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Trans. Graph*, pages 765–776, 2005. 2, 4

[27] D. N. Wood. Surface light fields for 3d photography, 2004. 2, 4

[28] L. Xia, C. Chen, and J. Aggarwal. Human detection using depth information by kinect. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011. 1, 2

[29] F. Xu, D. Shin, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. Goyal, F. Wong, and J. Shapiro. Photon-efficient computational imaging with a single-photon camera. *Computational Optical Sensing and Imaging*, 2016. 2

[30] C. Zhou, S. Lin, and S. Nayar. Coded aperture pairs for depth from defocus and defocus deblurring. *International Journal of Computer Vision*, 93(1):53–72, 2011. 6