Accidental Cameras: Creating Images from Shadows

by

Vickie Ye

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Author .....
Department of Electrical Engineering and Computer Science
February 1, 2018

Certified by .....
William T. Freeman
Professor
Thesis Supervisor

Accepted by .....
Christopher J. Terman
Chairman, Department Committee on Graduate Theses

Accidental Cameras: Creating Images from Shadows

by

Vickie Ye

Submitted to the Department of Electrical Engineering and Computer Science
on February 1, 2018, in partial fulfillment of the
requirements for the degree of
Master of Engineering

Abstract

In this thesis, we explore new imaging systems that arise from everyday occlusions and shadows. By modeling the structured shadows created by various occlusions, we are able to recover hidden scenes. We explore three such imaging systems. In the first, we use a wall corner to recover one-dimensional motion in the hidden scene behind the corner. We show experimental results using this method in several natural environments. We also extend this method to be used in other applications, such as for automotive collision avoidance systems. In the second, we use doorways and spheres to recover two-dimensional images of a hidden scene behind the occlusions. We show experimental results of this method in simulations and in natural environments. Finally, we present how to extend this approach to infer a 4D light field of a hidden scene from 2D shadows cast by a known occluder. 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.

Thesis Supervisor: William T. Freeman
Title: Professor

Contents

List of Figures

Chapter 1

Introduction

Although often not visible to the naked eye, in many environments, light from obscured portions of a scene is scattered over many of the observable surfaces. In certain cases, we can exploit the structured distribution of this light to recover information about the hidden scene. As an example, consider the scenario depicted in Figure 1(A).

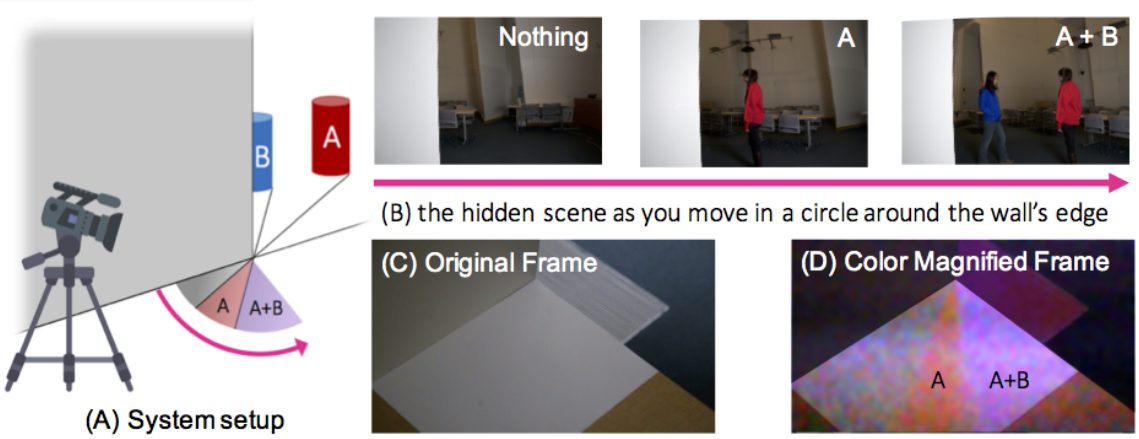

Figure 1-1: We show an example in which information from the scene hidden behind the wall corner can be recovered, using the visible observations at the base of wall corner. (A) show that the light from the hidden scene casts structured shadows on the floor at the base of the corner. This is similar to how, from the point of view of an observer walking around the occluding edge (along the magenta arrow), light from different parts of the hidden scene becomes visible at different angles (B). This subtle variation is not perceptible to the naked eye, such as in (C), but is present, and can be computationally extracted as it is in (D).

In this case, although the scene behind the corner is completely obscured from the

observer, we see in 1(A) and (D) that the light observed on the ground at the base of the corner shows subtle variations that reveal information about the elements behind the corner. From a point near the wall, very little of the hidden scene is visible, and the ground reflects little light from beyond the corner. However, as we move around the corner, along the magenta arrow, the ground begins to reflects red light as A becomes visible, and then purple light as both A and B become visible. This can be seen from the point of view of an observer at the corner in Figure 1(B); as the observer moves along the magenta arrow around the corner, she sees an increasing amount of the scene until eventually, the hidden scene comes fully into view.

We can then think of the shadow we see on the ground as really the sum of several distinct shadows caused by light from various regions in the scene. In this way, the vertical edge occlusion of the wall corner acts as an aperture that filters light from beyond the corner. We can therefore think of the optical system of Figure 1(A) as an *accidental camera* [21] for the out-of-sight region.

1.1 Related Work in Non-Line-of-Sight (NLOS) Imaging

The ability to image such scenes not directly in the line of sight would prove useful in a wide range of applications, from search-and-rescue, to anti-terrorism, to road traffic and collision avoidance [3]. However, this problem is very difficult. In this section we describe previous non-line-of-sight (NLoS) methods. Previous methods used to see past or through occluders have ranged from using WiFi signals [1] to exploiting random specular surfaces [27, 8]. In this summary, we emphasize two primary approaches that have been used to see past occluders and image hidden scenes.

Time-of-Flight Sensors: Past approaches to see around corners have largely involved using time-of-flight (ToF) cameras [19, 26, 13]. These methods involve using a laser to illuminate a point that is visible to both the observable and hidden scene,

and measuring how long it takes for the light to return [26, 20]. By measuring the light's time of flight, one can infer the distance to objects in the hidden scene, and by measuring the light's intensity, one can often learn about the reflectance and curvature of the objects [18]. Past work has used ToF methods to infer the location [11], size and motion [17, 9], and shape [22] of objects in the hidden scene. These methods have also been used to count hidden people [25].

ToF cameras work well in estimating the depths of hidden objects, however, they have a number of limitations. First, they require specialized and comparatively expensive detectors with fine temporal resolution. Second, they are limited in how much light they can introduce in the scene to support imaging. Third, they are vulnerable to interference from ambient light sources. Unlike passive methods, which make use of ambient illumination, ToF systems must introduce every photon that is measured. By contrast, our proposed passive technique operates in unpredictable indoor and outdoor environments with comparatively inexpensive consumer cameras.

Imaging Using RGB Videos: Other work has previously considered the possibility of using structures naturally present in the real world as cameras. Naturally occurring pinholes (such as windows) or pinspecks have been previously used for non-line-of-sight imaging [21, 7]. In addition, specular reflections off of human eyes have been used to image hidden scenes [16]. Although these accidental cameras can be used to reconstruct 2-D images, they require a more specialized accidental camera scenario than the simple edges we propose to use in this work.

Our work also bears some resemblance to [24], which focuses on detecting and visualizing small, often imperceptible, color changes in video. However, in our work, rather than just visualize these tiny color changes, we interpret them in order to reconstruct a video of a hidden scene.

1.2 Contributions and Outline

In this thesis, we explore the optical systems formed by commonly occurring occluders that allow us to recover information from a hidden or obstructed scene. In Chapter 2, we introduce the general framework to model and invert these optical systems. In this chapter, we state the assumptions we make of the imaging environment... In the following chapters, we examine specific instances of these accidental camera systems using this general framework.

In Chapter 3 explore the system introduced in Figure 1, the *corner camera*, and show that it allows us to recover motion behind a corner in one dimension. We show that our method is capable of recovering hidden motion in diverse natural environments, with several different lighting conditions and surfaces.

In Chapter 4, we show this system in two specific applications. We extend the corner camera to allow a moving observer to detect the motion behind a corner, and implement a real-time publicly available corner camera system. We also use the stereo views offered by adjacent corner cameras to track two dimensional motion behind a doorway that creates a binocular corner camera.

In Chapter 5, we explore the imaging capacity of two other occluders, spheres and rectangles, to reconstruct 2D projections of hidden scenes. The higher-dimensional occlusions allow us to recover static 2D images of hidden scenes. For instance, the structured light filtered through a doorway onto the ceiling can allow us to recover a 2D view of the scene outside the door.

In Chapter 6, we further expand this approach to recover to use complex occluders the 4D light field of a hidden scene [2]. The recovered light field would allow us to view the hidden scene from multiple perspectives and thereby recover depth information within the scene. This is an ambitious task, and we show that proper calibration and a good choice of a prior distribution over light fields is crucial to obtaining meaningful reconstructions.

Finally, in Chapter 7, we discuss next steps that can be taken to better understand the imaging power of these accidental cameras.

Chapter 2

General Imaging System

Our model of the systems we are interested in consists of three elements: a hidden scene, an occluder, and an observation plane. We assume that the observation plane is fully visible and the geometry of the occluder is fully known, but that the camera does not have direct line of sight of any element of the hidden scene.

In all cases we consider, the occluder is a non-reflective object of known geometry, placed between the observation plane and the hidden scene. By occluding light from the hidden scene in a known way, it acts as an aperture that allows us to reconstruct elements of the hidden scene, solely from the light we see on the observation plane. We assume the observation plane is a uniform Lambertian surface that diffusely reflects the light not blocked by the occluder. We can therefore formulate the following model to relate the illumination of the observation plane with that of the hidden scene.

2.1 Light Transport Model

For a uniform Lambertian surface with albedo a, the light L_o reflected from the observation plane at the point $\mathbf{p}$ is a function of the incoming light L_i':

$$L_o(\mathbf{p}) = a \int L_i'(\mathbf{p}, \widehat{\mathbf{v}}_i)(\widehat{\mathbf{v}}_i \cdot \widehat{\mathbf{n}}) d\widehat{\mathbf{v}}_i, \tag{2.1}$$

where $\widehat{\mathbf{v}}_i$ is the vector of the incoming light, and $\widehat{\mathbf{n}}$ is the surface normal of the observation plane. To simplify subsequent expressions, we assume an albedo of $a = 1$. We can further express the light L_i' coming into point $\mathbf{p}$ in the direction of $\mathbf{v}$ as a function of the light from the hidden scene L_h, the light from the rest of the world L_w, and the known binary visibility function v of the occluder:

$$L_i'(\mathbf{p}, \mathbf{v}) = v(\mathbf{p}, \mathbf{v}) L_h(\mathbf{p}, \widehat{\mathbf{v}}) + L_w(\mathbf{p}, \widehat{\mathbf{v}}). \tag{2.2}$$

We note that the parameterization of L_h depends on the information we are interested in recovering. In Equation 2.2, we parameterize the light from the hidden scene by the point $\mathbf{p}$ and direction $\widehat{\mathbf{v}}$ at which it hits the observation plane, if no occluders are present. However, different occluders are better suited for recovering different kinds of information from the hidden scene. As such, we adjust our parameterization of L_h accordingly.

For instance, the visibility function of a pinhole at location $\mathbf{q}$ can expressed as

$$v(\mathbf{p}, \widehat{\mathbf{v}}) = \begin{cases} 1, & \widehat{\mathbf{v}} = \frac{\mathbf{p}-\mathbf{q}}{|\mathbf{p}-\mathbf{q}|} \\ 0, & \text{otherwise,} \end{cases} \tag{2.3}$$

Then our observations $L_o(\mathbf{p}) \propto L_h\big(\mathbf{p}, \frac{\mathbf{p}-\mathbf{q}}{|\mathbf{p}-\mathbf{q}|}\big)$ allow us to directly recover a 2D image of the hidden scene. For this occluder, we could parameterize L_h to be centered about $\mathbf{q}$, solely as a function of the direction at which the light hits the observation plane: $L_h(\widehat{\mathbf{v}})$.

In this thesis, we employ different parameterizations, depending on the information the occluder allows us to recover. In Chapter 3, we examine a scenario in which the vertical edge of a wall corner allows us to recover the 1D angular projection of the hidden scene behind the corner. In this case, we parameterize our hidden scene to be centered about the vertical edge, as a function of the angle from wall θ (see Figure 2.1(A)). In Chapter 5, a doorway occluder allows us to recover a two-dimensional snapshot of the scene. We can therefore parameterize the light from the hidden scene

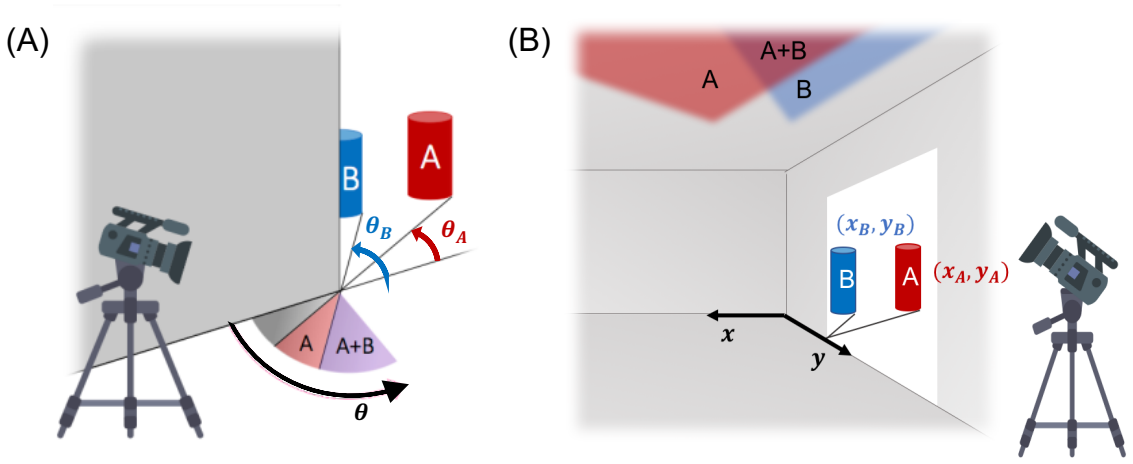

Figure 2-1: We show the various ways the light from the hidden scene can be parameterized, depending on the occluder we use. In (A), we show a system in which the wall corner allows us to recover the 1D angular projection of the scene behind the corner. We parameterize the hidden scene in 1D, as a function of the angle from the wall. In (B), we show a system in which the doorway occluder allows us to recover two dimensions of information. We parameterize the hidden scene in 2D, as a function of position in a plane in the scene.

as coming from a plane, as a function of 2D position, (x, y) (see Figure 2.1(B)). In Chapter 6, complex occluders such as plants allow us to recover 4D light fields of a hidden scene. We therefore parameterize the hidden light with a two-plane light field parameterization [15].

In all our scenarios, our observations and computations are discrete, and so we now present our model in these terms. We observe the light from the observation plane as M pixels $\mathbf{y}$, and we express the light from the hidden scene for some parameterization, discretely with N terms $\mathbf{x}$. We express the visibility function as an $M \times N$ matrix $\mathbf{V}$. More specifically, $\mathbf{V}_{ij}$ is zero if the occluder blocks the light of $\mathbf{x}_j$ from reaching observed pixel $\mathbf{y}_i$, and 1 if not. Then we have

$$\mathbf{y}_i = \sum_j \mathbf{V}_{ij} c_{ij} \mathbf{x}_j \tag{2.4}$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{2.5}$$

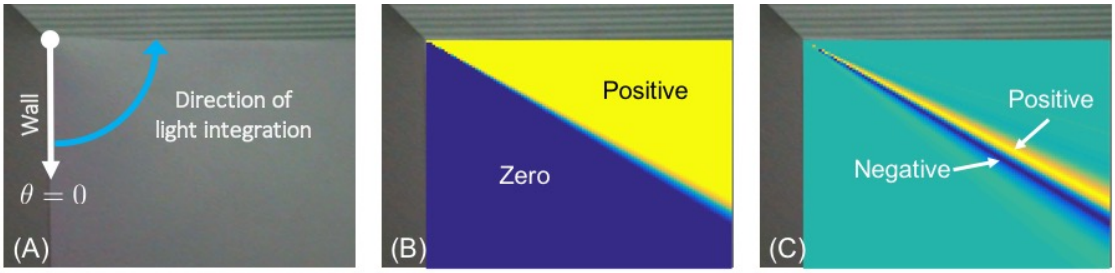

Figure 2-2: (A) We show an example observation for the edge camera introduced in 1. In this case, the occluder is the wall edge, which focuses light from the hidden scene behind the corner onto the observation plane at the base of the wall edge. In (B), we show the response of the system to a single point light source in the middle of our scene. The impulse responses of the system for all such desired reconstruction pixels form the columns of the light transport matrix $\mathbf{A}$. In (C) we visualize the row of the estimation gain matrix that recovers the light source of (B) from the observations.

where c_{ij} is some factor that takes into account the incident angle of light and various emission effects. In Equation 2.5, we combine the emission effects and the occlusion effects into the matrix $\mathbf{A}$, which we call the light transport matrix.

Every element $\mathbf{A}_{ij}$ indicates how the j-th term of the scene, $\mathbf{x}_j$ contributes to the irradiance observed at pixel i of the observation, $\mathbf{y}_i$. The columns of $\mathbf{A}$ can be then thought of as the impulse responses of the system; the column $\mathbf{A}_j$ is the shadow created on the observation plane when an isotropic light placed at the j-th element of the scene $\mathbf{x}_j$ is illuminated. In Figure 2.1(B), we visualize a column of the light transport matrix that corresponds with the system show in 2.1(A).

2.2 Inference

We seek a maximum a posteriori (MAP) estimate of the hidden scene parameters $\mathbf{x}$, given the M pixels $\mathbf{y}$ observed by the camera. To do so, we impose a prior distribution on $\mathbf{x}$. Note that the prior we choose will depend on how we parameterize the hidden scene, but we enforce that our prior be Gaussian. If we assume i.i.d. Gaussian noise, then our likelihood and prior models are:

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x}, \lambda\mathbf{I}) \tag{2.6}$$

$$p(\mathbf{x}) = \mathcal{N}(\mu, \mathbf{Q}). \tag{2.7}$$

Our posterior distribution is then

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\widehat{\mathbf{x}}, \mathbf{\Sigma}) \tag{2.8}$$

$$\mathbf{\Sigma} = \left[\lambda^{-1}\mathbf{A}^T\mathbf{A} + \mathbf{Q}\right]^{-1} \tag{2.9}$$

$$\widehat{\mathbf{x}} = \lambda^{-1}\mathbf{\Sigma}\mathbf{A}\mathbf{y}. \tag{2.10}$$

To gain intuition about the information we use to infer each element of $\mathbf{x}$, we consider the matrix $\mathbf{K}$, which we call the *estimation gain matrix*:

$$\mathbf{K} = \lambda^{-1}\left[\lambda^{-1}\mathbf{A}^T\mathbf{A} + \mathbf{Q}\right]^{-1}\mathbf{A} \tag{2.11}$$

$$\widehat{\mathbf{x}} = \mathbf{K}\mathbf{y}. \tag{2.12}$$

In Figure 2.1(C), we visualize one row i of the estimation gain matrix that is used to reconstruct that element i of $\mathbf{x}$ from $\mathbf{y}$, for the system we show in 2.1(A). This visualization can be thought of as a pixel-wise map of how we use parts of the observation $\mathbf{y}$ to estimate the element $\mathbf{x}_i$. In this case, the Figure 2.1(C) shows the reconstruction operation is essentially performing an angular derivative on the observations.

Chapter 3

Corner Cameras: Reconstructing Motion in One Dimension

We first explore the case of the imaging system created by a vertical-edge occluder, like a wall corner, which we call the *corner camera* [4]. Since vertical edges are ubiquitous, naturally occurring corner cameras can be easily found all around us, and we explore it most extensively in this thesis.

In this system, the wall corner only filters light in one dimension, along the angular projection about the corner, and so we seek a one-dimensional reconstruction over time of the scene behind the corner. As such, the images we present are not images of the scene itself, but are rather space-time trajectories of elements of the scene moving in time. We show that our method reconstructs clear trajectories of moving people of several sizes, including children.

We note that in the formulation of our general model in Chapter 2, we make a few assumptions about the albedo and texture of the observation plane. However, we show in this chapter that our that the corner camera can still recover clear trajectories even when these assumptions are not met. For instance, our system operates for observation planes of several textures and surface reflectances, such as tile, carpet, concrete, brick, and wood. Our method also handles several lighting conditions of the hidden and visible scenes. Finally, we show that our method can detect the signal using cameras of varying image quality and pixel noise levels.

3.1 Model

We present our model of the imaging system in terms of the framework we introduced in Chapter 2. We parameterize the light from the hidden scene as a function of the angle θ from the corner (see Figure 3.1). The 1-D image $L_h(\theta)$ we wish to recover is defined over the continuous space of angular coordinates $\theta = [0, \pi/2]$. Rather than simply assuming a discretized image of point sources for reconstruction, we instead parameterize the continuous $L_h(\theta)$ with N terms, $\mathbf{x}^{(t)}$. We approximate $L_h(\theta)$ as a series of scaled and shifted triangle functions with width $\frac{\pi}{N}$. This representation linearly interpolates between specified centers, and the contribution of each parameter $\mathbf{x}$ to each $\mathbf{y}$ can be calculated as

$$L_h(\theta) \approx \sum_{n=0}^{N-1} \mathbf{x}[n] \max\left(1 - \frac{2N}{\pi}\left|\frac{\pi(2n+1)}{4N} - \theta\right|, 0\right) \tag{3.1}$$

Once we have parameterized the hidden scene, we can easily derive the linear relation between the observed pixels $\mathbf{y}$ and the hidden scene parameters $\mathbf{x}$. In Figure 3.1, we show how we build the light transport matrix $\mathbf{A}$ for this system. Each column i of the transport matrix $\mathbf{A}$ is the observed response of the ground irradiance to an isotropic light placed at the i-th element of $\mathbf{x}$, at the corresponding angle θ_i. That is, for each column i of $\mathbf{A}$, we light up the i-th element of x at the angle θ_i, leaving the other terms dark. This transport matrix then transforms the parameters $\mathbf{x}$ of an arbitrary scene into the observed $\mathbf{y}$.

3.2 Method

Using a video recording of the observation plane, we generate a 1-D video indicating the changes in a hidden scene over time. These 1-D angular projections of the hidden scene viewed over many time-steps reveal the trajectory of a moving object behind the occluding edge. We present our method within the framework introduced in Chapter 2.

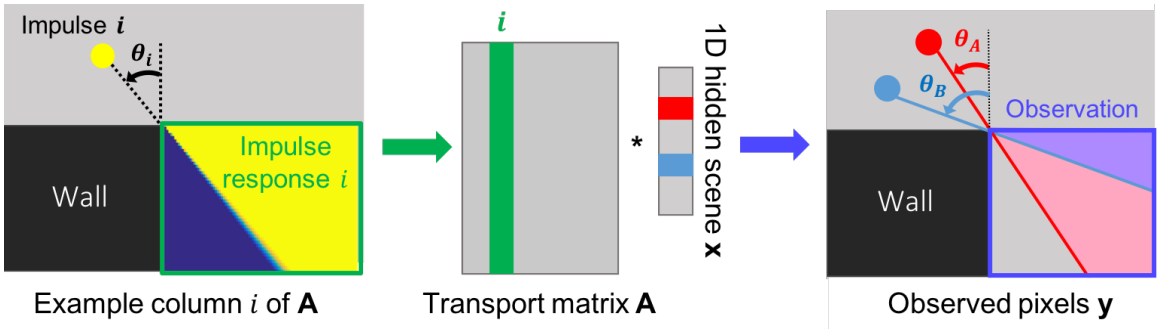

Figure 3-1: We show our linear model of the corner camera. On the left, we show an example column of the light transport matrix $\mathbf{A}$, which is built from the impulse responses of the system for each element of $\mathbf{x}$. The resulting $\mathbf{A}$ transforms the scene parameters $\mathbf{x}$ into the observed pixels $\mathbf{y}$.

Likelihood: At each time t, we relate the observed M-pixels on the projection plane, $\mathbf{y}^{(t)}$, to the 1-D angular projection of the hidden scene, $L_h^{(t)}(\phi)$. We formulate a discrete approximation to our edge camera system by describing the continuous image $L_h^{(t)}(\phi)$ using N terms, $\mathbf{x}^{(t)}$. Then assuming i.i.d. Gaussian noise, the relation between $\mathbf{y}^{(t)}$ and $\mathbf{x}^{(t)}$ can be written as

$$\mathbf{y}^{(t)} = L_v^{(t)} + \mathbf{A}\mathbf{x}^{(t)} + \mathbf{w}^{(t)}, \qquad \mathbf{w}^{(t)} \sim \mathcal{N}(0, \lambda\mathbf{I}), \tag{3.2}$$

where $\mathbf{A}$ is the light transport matrix specific to this system's geometry. We show an example of a row of $\mathbf{A}$ for the corner camera setup in Figure 2.1(B).

Let $\widetilde{\mathbf{A}}$ be the column augmented matrix $[\mathbf{1} \;\; \mathbf{A}]$. We can then express the likelihood of an observation given $\mathbf{x}^{(t)}$ and $L_v^{(t)}$ as:

$$p(\mathbf{y}^{(t)}|\mathbf{x}^{(t)}, L_v^{(t)}) = \mathcal{N}\left(\widetilde{\mathbf{A}}\left[L_v^{(t)} \; \mathbf{x}^{(t)T}\right]^T, \lambda\mathbf{I}\right). \tag{3.3}$$

Prior: The signal we are trying to extract is very small relative to the total light intensity on the observation plane. To improve the quality of results, we enforce spatial smoothness of $\mathbf{x}^{(t)}$. We use a simple L2 smoothness regularization over adjacent

parameters in $\mathbf{x}^{(t)}$. This corresponds, for a gradient matrix $\mathbf{G}$, to using the prior

$$\begin{aligned} p(\mathbf{x}^{(t)}) &\propto \prod_{n=1}^{N-1} \exp\left[-\frac{1}{2\sigma^2}\|\mathbf{x}^{(t)}[n] - \mathbf{x}^{(t)}[n-1]\|_2^2\right] \\ &= \mathcal{N}(0, \sigma^2(\mathbf{G}^T\mathbf{G})^{-1}). \end{aligned} \tag{3.4}$$

We note that in Equations 5.1 and 5.2, we separate out the $L_v^{(t)}$, the changing light from the visible region of the scene, not obscured from the observation plane. This light has magnitude much greater than our signal, and has an under-determined relationship with the light we observe on the ground. We encode this lack of relationship in the final posterior distribution below.

Posterior: We seek a maximum a posteriori (MAP) estimate of the hidden image coefficients, $\mathbf{x}^{(t)}$, given M observations, $\mathbf{y}^{(t)}$, measured by the camera. By combining the defined Gaussian likelihood and prior distributions, we obtain a Gaussian posterior distribution of $\mathbf{x}^{(t)}$ and $L_v^{(t)}$,

$$\begin{aligned} p(\mathbf{x}^{(t)}, L_v^{(t)}|\mathbf{y}^{(t)}) &= \mathcal{N}\left(\left[\hat{L}_v^{(t)}\ \hat{\mathbf{x}}^{(t)T}\right]^T, \Sigma^{(t)}\right) \\ \Sigma^{(t)} &= \left[\lambda^{-1}\widetilde{\mathbf{A}}^T\widetilde{\mathbf{A}} + \sigma^{-2}\begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T\mathbf{G}\end{pmatrix}\right]^{-1} \\ \left[\hat{L}_v^{(t)}\ \hat{\mathbf{x}}^{(t)T}\right]^T &= \Sigma^{(t)}\lambda^{-1}\widetilde{\mathbf{A}}^T\mathbf{y}^{(t)} \end{aligned} \tag{3.5}$$

where the maximum a posteriori estimate is given by $\hat{\mathbf{x}}^{(t)}$. We show the operation we perform on $\mathbf{y}$ to recover an element i of $\mathbf{x}$, in Figure 2.1(C). This is one row of the estimation gain matrix, $\mathbf{K} = \Sigma^{(t)}\lambda^{-1}\widetilde{\mathbf{A}}^T$, visualized to show the pixel-wise contributions of $\mathbf{y}$ to the estimated x_i. As expected for this geometry, the operation performed resembles an angular derivative about the corner.

3.3 Implementation Details

Background Subtraction: Since we are interested in identifying temporal differences in a hidden scene due to a moving subject, we must remove the effect of the scene's background illumination. Although this could be accomplished by first subtracting a background frame, L_o^0, taken without the subject, we avoid requiring the availability of such a frame. Instead, we assume the subject's motion is roughly uniform over the video, and use the video's mean image in lieu of a true background frame.

We note that our above model assumes a uniform, Lambertian ground plane, which is not always the case for our following experiments. However, when we subtract out this background image, the motion of objects behind the corner can still be discerned. The effects of a nonuniform ground albedo, as well as a non-Lambertian BRDF, are largely encompassed in the background illumination we subtract.

Temporal Averaging: In addition to spatial smoothness, we can also impose temporal smoothness on our MAP estimate. $\hat{\mathbf{x}}^{(t)}$. This helps to further regularize our result, at the cost of some temporal blurring. However, to emphasize the coherence among results, we do not impose this additional constraint. Instead, we batch and average every ten frames to increase the signal to noise ratio (which still results in an effective frame rate much faster than the motion behind the corner). We then compute each 1-D image, $\mathbf{x}^{(t)}$, independently of each other. Results obtained with temporal smoothness constraints are shown in Appendix A.2.

Rectification: All of our analysis thus far has assumed we are observing the floor with an orthographic view, parallel to the plane's normal vector. However, in most situations, the camera observes the ground plane at an angle. In order to make the construction of the matrix $\mathbf{A}$ easier, we begin by rectifying all of our images to the orthographic view using a homography. In these results, we assume the ground is perpendicular to the occluding edge, and estimate the homography using either a calibration grid or regular patterns, such as tiles, that naturally appear on the ground.

Alternatively, a known camera calibration could be used. Furthermore, we manually identify the region of interest on the ground, although several methods can automate this process.

Parameter Selection: The noise parameter λ^2 is set for each video as the median variance of estimated sensor noise. The spatial smoothness parameter, σ, is set to 0.1 for all results.

3.4 Experiments and Results

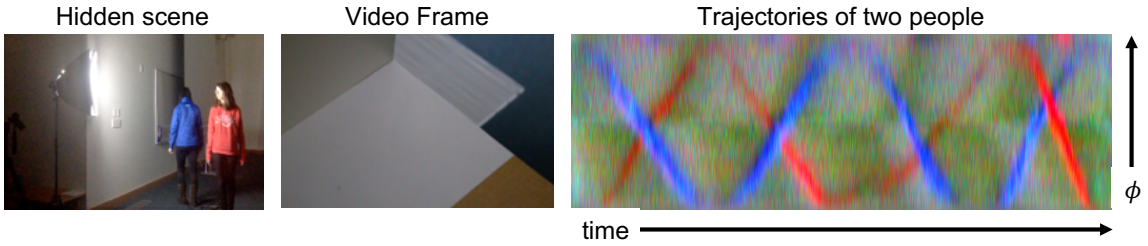

Figure 3-2: We show a result in which two people, one red and one black, walk around, illuminated by a large diffuse light, in an otherwise dark room. The resulting trajectories of their motion therefore show the colors of their clothing.

Our algorithm reconstructs a 1-D video of a hidden scene from behind an occluding edge, allowing users to track the motions of obscured, moving objects. In all results shown, the subject was not visible to an observer at the camera.

We present results using space-time images. These images contain curves that indicate the angular trajectories of moving people. All results, unless specified otherwise, were generated from standard, compressed video taken with a SLR camera.

3.4.1 Environments

We show several applications of our algorithm in various indoor and outdoor environments. For each environment, we show the reconstructions obtained when one or two people were moving in the hidden scene.

Indoor: In Fig. 3.4, we show a result obtained from a video recorded in a mostly dark room. A large diffuse light illuminated two hidden subjects wearing red and blue clothing. As the subjects walked around the room, their clothing reflected light, allowing us to reconstruct a 1-D video of colored trajectories. As correctly reflected in our constructed video, the subject in blue occludes the subject in red three times before the subject in red becomes the occluder. We note that our following results do not show the colors of the hidden people, and were taken in environments with more natural lighting.

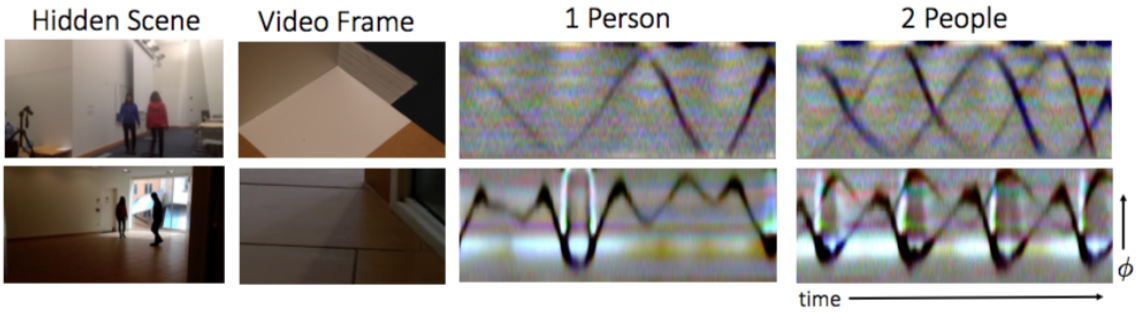

Figure 3-3: One-dimensional reconstructed videos of indoor, hidden scenes. Results are shown as space-time images for sequences where one or two people were walking behind the corner. In these reconstructions, the angular position of a person, as well as the number of people, can be clearly identified. Bright line artifacts are caused by additional shadows appearing on the penumbra.

Fig. 3-3 shows additional examples of 1-D videos recovered from indoor edge cameras. In these sequences, the environment was well-lit. The subjects occluded the bright ambient light, resulting in the reconstruction's dark trajectory. Note that in all the reconstructions, it is possible to count the number of people in the hidden scene, and to recover important information such as their angular size and speed, and the characteristics of their motion.

Outdoor: In Fig. 3-4 we show the results of a number of videos taken at a common outdoor location, but in different weather conditions. The top sequences were recorded during a sunny day, while the bottom two sequences were recorded while it was cloudy. Additionally, in the bottom sequence, raindrops appeared on the ground

during recording, while in the middle sequence the ground was fully saturated with water. Although the raindrops cause artifacts in the reconstructed space-time images, you can still discern the trajectory of people hidden behind the wall.

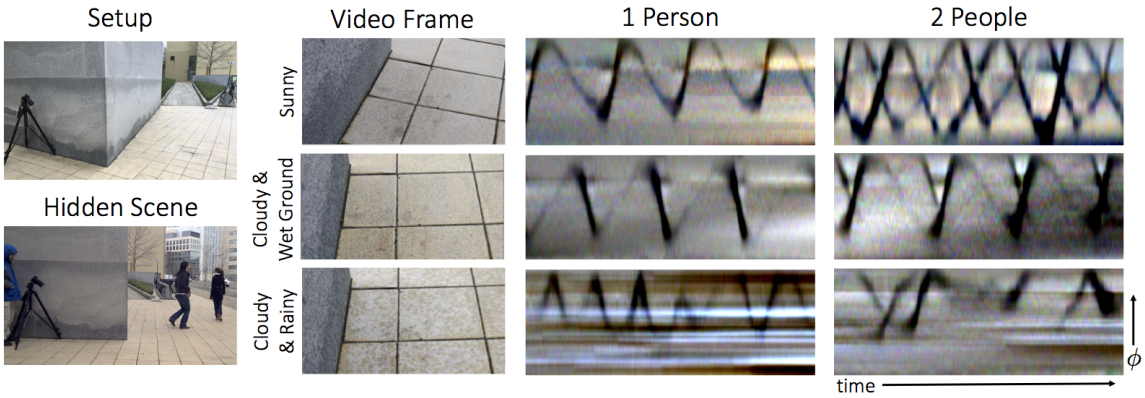

Figure 3-4: 1-D reconstructed videos of a common outdoor, hidden scene under various weather conditions. Results are shown as space-time images. The last row shows results from sequences taken while it was beginning to rain. Although artifacts appear due to the appearing raindrops, motion trajectories can be identified in all reconstructions.

3.4.2 Video Quality

In all experiments shown thus far we have used standard, compressed video captured using a SLR camera. However, video compression can create large, correlated noise that may affect our signal. We have explored the effect video quality has on results. To do this, we filmed a common scene using 3 different cameras: an iPhone 5s, a Sony Alpha 7s SLR, and a uncompressed RGB Point Grey. Fig. 3-5 shows the results of this experiment assuming different levels of i.i.d. noise. Each resulting 1-D image was reconstructed from a single frame. The cell phone camera's compressed videos resulted in the noisiest reconstructions, but even those results still capture key features of the subject's path.

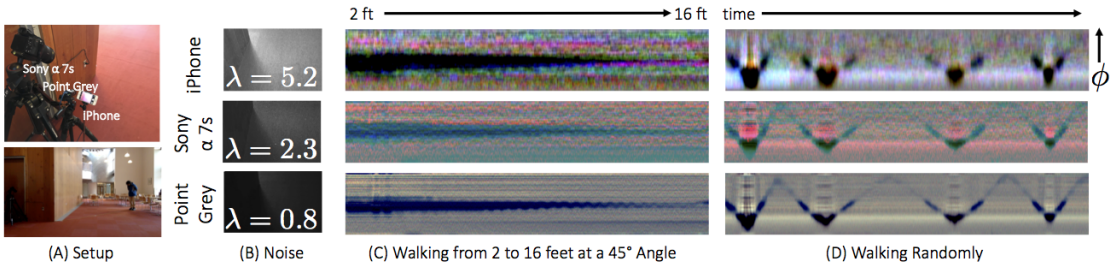

Figure 3-5: The result of using different cameras on the reconstruction of the same sequence in an indoor setting. Three different 8-bit cameras (an iPhone 5s, a Sony Alpha 7s SLR, and an uncompressed RGB Point Grey) simultaneously recorded the carpeted floor. Each camera introduced a different level of per-pixel sensor noise. The estimated standard deviation of sensor noise, λ, is shown in (B). We compare the quality of two sequences in (C) and (D). In (C), we have reconstructed a video from a sequence of a single person walking directly away from the corner from 2 to 16 feet at a 45 degree angle from the occluded wall. This experiment helps to illustrate how signal strength varies with distance from the corner. In (D), we have done a reconstruction of a single person walking in a random pattern.

3.4.3 Estimated Signal Strength

In all of our presented reconstructions we show images with an intensity range of 0.1. As these results were obtained from 8-bit videos, our target signal is less than 0.1% of the video's original pixel intensities. To better understand the signal measurement requirements, we consider a simple model of the edge camera system that both explains experimental performance and enables the study of asymptotic limits. We provide a more detailed formulation in Appendix A.1.

We consider three sources of emitted or reflected light, show in Figure 3-6: a target cylinder (which acts as a rough model of a person), a hemisphere of ambient light (from the surrounding environment), and the occluding wall (which we assume to be much taller than the cylinder). We assume that all surfaces are Lambertian. The strength of the signal due to the cylinder, relative to the strength of the ambient illumination and light bouncing off the wall, can be computed using this model.

The observation plane can be split into three regions of interest, as shown the figure. Region 1 receives no light from the target cylinder. Region 2 receives increasing partial light from the cylinder as we move around the corner. Finally, region 3 sees

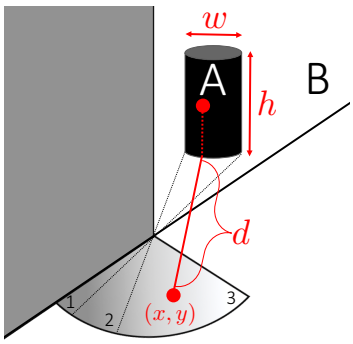

Figure 3-6: Calculation of the contribution to the brightness of the target. Region 1 is the umbra region of the observation plane; none of the cylinder is visible from region 1, and thus its intensities do not change after introducing the target to the scene. Region 2 is the penumbra region; in region 2, part of the cylinder is visible. In region 3, the entire cylinder is visible.

the entire cylinder. We assume that all three regions receive equal contributions of light from the environment and the wall. We use this model to gain insight toward the contribution of the cylinder to the total brightness observed at each of these three regions on the observation plane.

The intensity in region 1 is simply $I_1 = A + B$. To determine the change of intensity in region 3, we compute the difference between the reflected light with and without the cylinder. We can express the impact of the cylinder's presence on a patch at a point (x, y) in region 3 as follows:

$$I_3 = (A - B)\frac{w}{d}\int_0^{\arctan(\frac{h}{d})} \sin(\theta)d\theta \tag{3.6}$$

$$= (A - B)\frac{w}{d}\frac{d}{\sqrt{h^2 + d^2}} \tag{3.7}$$

where w, h, and d are defined as shown in Figure 3-6.

For region 2, the same calculation applies, but is additionally multiplied by the fraction of the cylinder that is visible from location (x, y). More specifically, γ_1 to be the angular boundary between region 1 and region 2 and γ_2 to be the angular boundary between region 2 and 3. Then the brightness we observe at location (x, y) in region 2 is given by

$$I_2(x,y) = \frac{\tan^{-1}(y/x) - \gamma_0}{\gamma_1 - \gamma_0} \; I_3. \tag{3.8}$$

For reasonable assumed brightnesses of the cylinder, hemisphere, and half-plane (150, 300, and 100, respectively, in arbitrary linear units), the brightness change on the observation plane due to the cylinder will be an extremum of -1.7 out of a background of 1070 units. This is commensurate with our experimental observations of $\sim 0.1\%$ change of brightness over the penumbra region.

This model shows the asymptotic behavior of the edge camera. Namely, at large distances from the corner, brightness changes in the penumbra decrease faster than would otherwise be expected from a 1-D camera. This is because the arrival angle of the rays from a distant cylinder are close to grazing with the ground, lessening their influence on the penumbra. However, within 10 meters of the corner, such effects are small.

Chapter 4

Extensions of the Corner Camera

4.1 Reconstruction with a moving observer

The corner camera is a relevant contribution especially because wall corners are ubiquitous. Specifically, it could potentially be used in automotive collision avoidance systems. In order explore this possibility, we adapt the imaging system to detect motion behind the corner in real time, from a moving camera observer. This new problem setting introduces a few new challenges, which we begin to address in this chapter. First, the moving camera observer introduces deviations from the original light transport model, and complicates the recovery of the signal. Second, the system must be implemented efficiently to run in real-time. We address each of these problems in the following sections, and offer insight to future directions to take to further this potential application.

4.1.1 Problems of a Moving Camera

The problem of stabilizing video from a moving camera has been extensively treated in vision and in robotics [10, 23, 28]. In this work, we make a few assumptions about the observed video frames. Namely, we assume that observer is moving slower than the frame rate of the camera, and will not encounter motion blur in the resulting video frames. Second, we assume that the corner of interest and its corresponding

observation plane is in every video frame we process. Moreover, we assume a comparable number of observed pixels from the moving camera observer to the stationary camera observer videos we process with the original formulation in Chapter 3.

The moving camera observer introduces two major problems when recovering our desired signal. First is the problem of video stabilization. Because the signal-to-noise ratio of the system is so low, the added noise from imprecise video stabilization adds artifacts to the recovered motion. Second is the changing viewing angle of the moving camera, which introduces unmodeled changes in the irradiance we observe from the ground.

Video Stabilization: Traditional methods of video stabilization involve detecting and matching features in consecutive frames, and calculating the perspective transform between the frames [10, 23, 28]. We approached this problem in the same way: we detected SIFT features in each frame, matched them using nearest-neighbor techniques, and used RANSAC to compute the homography between the frames. We also experimented with methods to track the distinctive wall corner and wall base structures, such as line segment detection and tracking. However, we found that the first approach was the least noisy and most reliable.

In Figure 4.1.1, we show results obtained on video that was stabilized using the classic feature detection and rectification approach. However, we note that homography estimation on smoother, less textured floors (such as the smooth tile floor at the top of the figure), is more error prone. As such, we notice large horizontal artifacts in the space-time images, due to misalignment with the background image. As a result, sharp features and large gradients in the background image interrupt and occasionally obscure the signal. This effect is especially noticeable against the small signal we try to recover.

Deviations from Light Transport Model Even if video alignment were perfect, the moving camera observer exacerbates certain deviations from our light transport model. In our original formulation of the corner camera, we can successfully re-

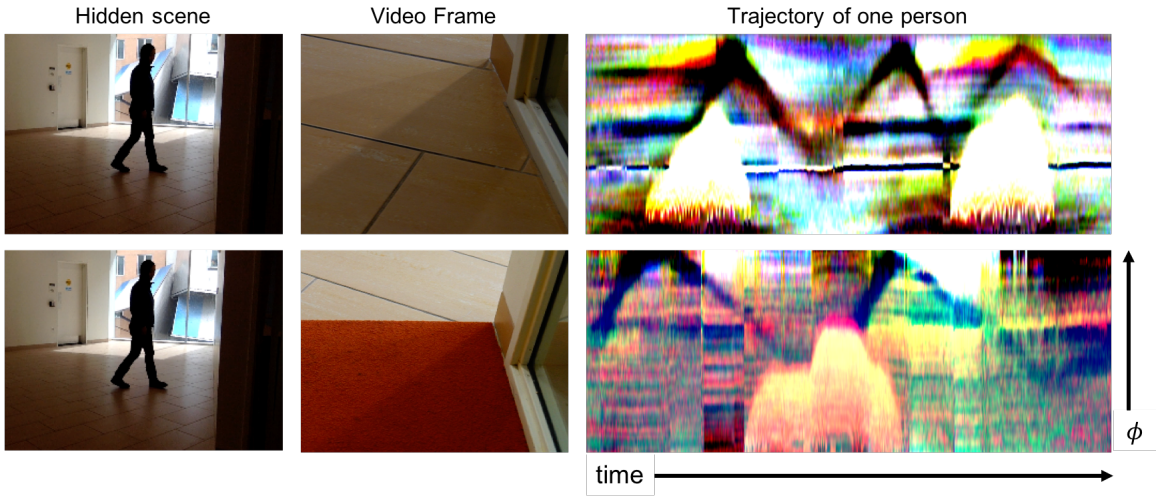

Figure 4-1: We show the reconstructed trajectories of the hidden scene, for a moving camera observer. We show results for two scenarios. In the top panel, the observation plane is a tiled surface; in the bottom, it is carpeted. The tile surface is more reflective, and shows noticeable changes in the brightness of the reconstruction as the camera viewing angle changes. The carpet shows a less dramatic effect. Horizontal line artifacts in both reconstructions are due to misalignment between the current frame and the mean frame.

cover our desired trajectories because the background irradiance was fairly constant throughout the input video. However, in the case of a moving camera, the viewing angle of the ground changes. For a non-Lambertian surface, such as reflective tile or wood, this causes our standard background subtraction to fail.

To address this problem, we subtracted an exponentially decaying moving average frame from each input frame. In this way, we can approximately capture the most recent background irradiance. Despite this step, we still see shifts in color and brightness in the trajectories shown in Figure 4.1.1. Notice that for reflective tile, the changing viewing angle introduces dramatic brightness changes in the space-time images, while for a carpeted ground, the changes in brightness are weaker.

4.1.2 Real-time system

We implement the real-time interactive corner camera system in C++, using OpenCV [10]. In Figure 4.1.2, we show a screen shot of the system usage. The user can use either a

USB web camera or a PointGrey camera as a live input stream to the system; more work must be done to enable the system to process live video for other commercially available video recorders. The code and implementation are publicly available [1].

The user first selects the corner they are interested in, as well as the region at the base of the corner to observe (see Figure 4.1.2 middle). The system can then immediately begin processing the observations live. We show snapshots of real-time output at three different times in Figure 4.1.2.

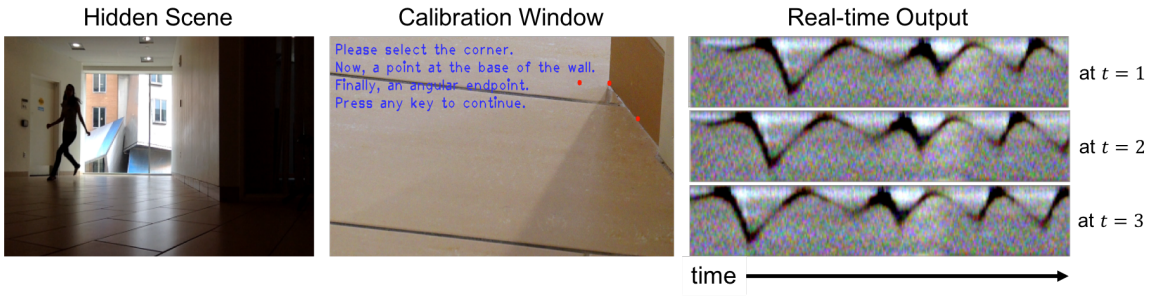

Figure 4-2: Here we show an example of the usage of the real-time corner camera. In the left panel, we show the hidden scene of a girl behind a corner. In the middle panel, we show the interactive calibration window, in which the user selects the corner and observation region on the ground. In the right panel, we show snapshots of the real-time output at three different times. In each frame, we see a later part of the trajectory being added to the right of the space-time image.

For a stationary camera observer, the inference algorithm for recovering the hidden motion behind the corner is very lightweight. Once the light transport matrix $\mathbf{A}$ and resulting estimation gain matrix $\mathbf{K}$ is computed, the per-frame operation essentially becomes a matrix multiplication operation. Currently the system can process live video streams of frame rates up to 60fps.

We have begun adding video stabilization features to this system to allow for real-time moving video input, but this slows down our processing algorithm to only handle video streams with frame rates up to 25fps. Moreover, this system remains very sensitive, because of the remaining problems in extracting signal from a moving camera.

[1] https://github.com/vye16/corner-camera

4.2 Stereo Tracking with Edge Cameras

In a hidden scene behind a doorway, the pair of vertical doorway edges yield a pair of corner cameras that inform us about the hidden scene. By treating the observation plane at the base of each edge as a camera, we can obtain stereo 1-D images that we can then use to triangulate the absolute position of a subject over time.

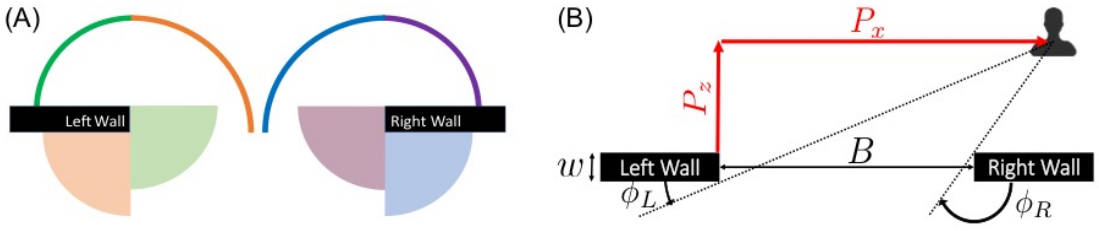

Figure 4-3: (A) The four edges of a doorway contain penumbras that can be used to reconstruct a 180° view of a hidden scene. Parallax occurs in the reconstructions from the left and right wall. (B) A hidden person will introduce an intensity change on the left and right wall penumbras at angles $\phi_L^{(t)}$ and $\phi_R^{(t)}$. We can recover the hidden person's two-dimensional location using Eq. 4.2.

4.2.1 Recovering Depth with Adjacent Corner Cameras

A single corner camera allows us to reconstruct a 90° angular image of an occluded scene. We now consider a system composed of four edge cameras, such as an open doorway, as illustrated in Fig. 4-3(A). Each side of the doorway contains two adjacent edge cameras, whose reconstructions together create a 180° view of the hidden scene.

The two sides of the doorway provide two views of the same hidden scene, but from different positions. This causes an offset in the projected angular position of the same person (see Fig. 4-3(B)). Our aim is to use this angular parallax to triangulate the location of a hidden person over time. Assume we are observing the base of a doorway, with walls of width w separated by a distance B. A hidden person will introduce an intensity change on the left and right wall penumbras at angles of $\phi_L^{(t)}$ and $\phi_R^{(t)}$, respectively. From this correspondence, we can triangulate their 2-D location.

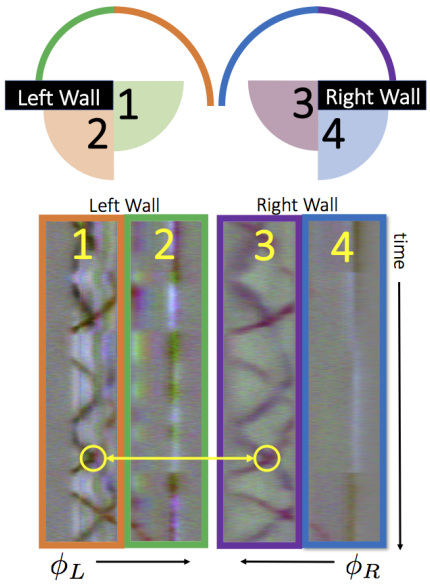

Figure 4-4: The four edges of a doorway contain penumbras that can be used to reconstruct a 180° view of a hidden scene. The top diagram indicates the penumbras and the corresponding region they describe. Parallax occurs in the reconstructions from the left and right wall. This can be seen in the bottom reconstruction of two people hidden behind a doorway. Numbers/colors indicate the penumbras used for each 90° space-time image.

$$P_z^{(t)} = \frac{B - \eta^{(t)}}{\cot \phi_L^{(t)} + \cot \phi_R^{(t)}} \tag{4.1}$$

$$P_x^{(t)} = P_z^{(t)} \cot \phi_L^{(t)} \tag{4.2}$$

$$\eta^{(t)} = \begin{cases} w \cot(\phi_R) & P_x \leq 0 \\ 0 & 0 \leq P_x \leq B \\ w \cot(\phi_L) & P_x \geq B \end{cases} \tag{4.3}$$

where (P_x, P_z) are the x- and z-coordinate of the person. We define the top corner of the left doorway, corner 1 in Fig. 4-3(A), as $(P_x, P_z) = (0, 0)$.

Assuming the wall is sufficiently thin compared to the depth of moving objects in the hidden scene, the $\eta^{(t)}$ term can be ignored. In this case the relative position of the person can be reconstructed without any knowledge of the absolute geometry

of the doorway (e.g. B or w). In all results shown in this paper, we have made this assumption.

While automatic contour tracing methods exist [12], for simplicity, in our stereo results, we identify the trajectories of objects in the hidden scene manually by tracing a path on the reconstructed space-time images.

4.2.2 Experiments and Results

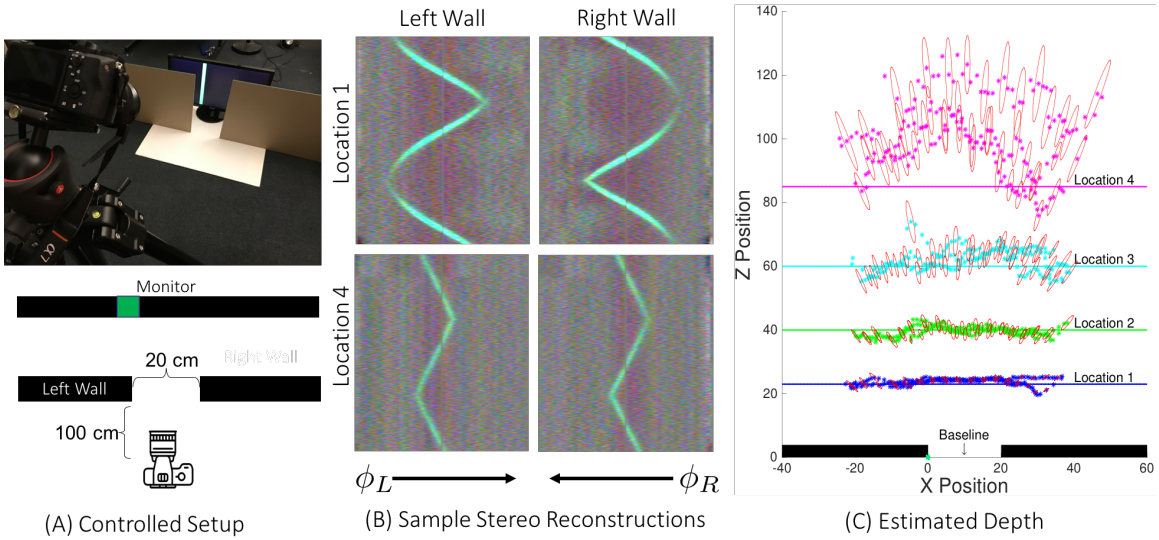

Figure 4-5: We performed controlled experiments to explore our method to infer depth from stereo edge cameras. A monitor displaying a moving green line was placed behind an artificial doorway (A) at four locations: 23, 40, 60, and 84 cm, respectively. (B) shows reconstructions done of the edge cameras for the left and right wall when the monitor was placed at 23 and 84 cm. Using tracks obtained from these reconstructions, the 2-D position of the green line in each sequence was estimated over time (C). The inferred position is plotted with empirically computed error ellipses (indicating one standard deviation of noise).

We show that this method can estimate the 2D position of a hidden object using four edge cameras, such as in a doorway. We present a series of experiments in both controlled and uncontrolled settings.

Controlled Environment: We first conducted a controlled experiment in which a monitor displaying a slowly moving green line was placed behind two walls, separated

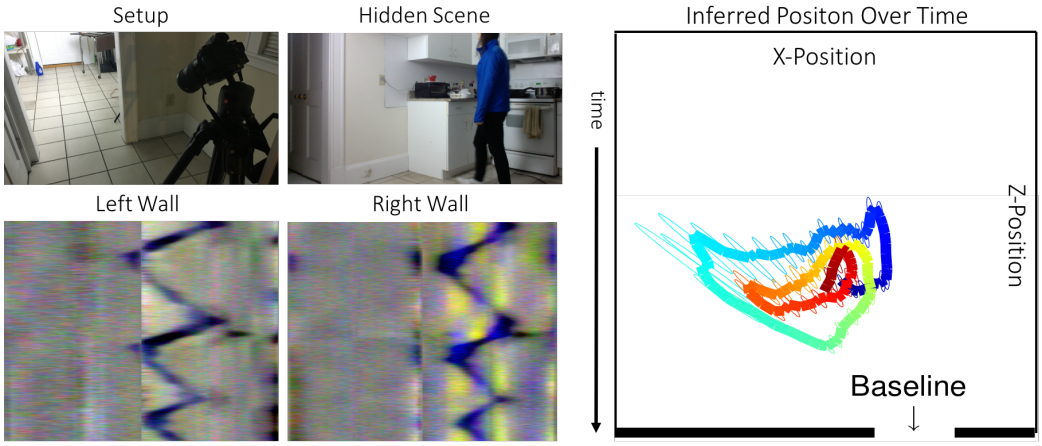

Figure 4-6: The results of our stereo experiments in a natural setting. A single person walked in a roughly circular pattern behind a doorway. The 2-D inferred locations over time are shown as a line from blue to red. Error bars indicating one standard deviation of error have been drawn around a subset of estimates. Our inferred depths capture the hidden subject's cyclic motion, but are currently subject to large error.

by a baseline of 20 cm, at a distance of roughly 23, 40, 60, and 84 cm. Fig. 4-5B shows sample space-time reconstructions of each 180° edge camera. The depth of the green line was then estimated from trajectories obtained from these space-time images. Empirically estimated error ellipses are shown in red for a subset of the depth estimates.

Natural Environment: Fig. 4-6 shows the results of estimating 2-D positions from doorways in natural environments. The hidden scene consists of a single person walking in a circular pattern behind the doorway. Although our reconstructions capture the cyclic nature of the subject's movements, they are sensitive to error in the estimated trajectories. Ellipses indicating empirically estimated error have been drawn around a subset of the points.

We note that there are multiple sources of error that can introduce biases into location estimates. Namely, because P_z scales inversely with $cot(\phi_L) + cot(\phi_R)$, small errors in the estimated projected angles of the person in the left and right may cause large errors in the estimated position of the hidden person, particularly at larger depths. Misidentifying the corner of each occluding edge will also introduce systematic error to the estimated 2-D position.

To better understand how to mitigate these sources of error, we propose to conduct more controlled experiments, in the spirit of the controlled monitor experiment, to understand when and how our method breaks down.

We also propose to directly model the doorway system, as a way to recover two-dimensional information. The depth and angular location are encoded in the geometry of the doorway, and by directly modeling this, we may be able to more precisely reconstruct two-dimensional location. In Chapter 5, we explore ways to model and reconstruct two dimensional scenes in a similar way.

Chapter 5

Reconstructing in Two Dimensions

Because of their ubiquity, we have thus far focused on wall edges and corners as occluders in naturally-occurring camera systems. However, the vertical edge can only integrate light from the hidden scene in vertical slices; all of the light in a single vertical slice is squashed together in our observation. As such our reconstructions of what happens behind a corner are in one spatial dimension. In this chapter, we expand our scope of accidental camera systems to recover 2D projections of a hidden scene.

In Section 5.1, we develop the imaging model for a spherical occluder. We find that because of the geometry of the occluder, reconstructed pixels are warped and ambiguous. In Section 5.2, we explore rectangular occluders, which suffer less from this ambiguity. Because of the sharp corners of rectangles, adjacent pixels at the edges can more clearly reconstructed than for the spherical occluder.

5.1 Spherical Occluders

Spheres are also commonly occurring, in doorknobs and balls, and are not constrained to recover scenes in a single dimension. In Figure 5.1, we show a schematic of the imaging system created by a sphere placed on the floor. We show in 5.1(A) a sphere in between the wall and two cylinders. The wall is our observation plane, upon which we can observe the faint shadow cast by the sphere. The camera sees observation

plane behind the sphere, but cannot see the two cylinders. We project the hidden scene onto an imagined 2D plane, shown in Figure 5.1(B). We place this plane at a depth of our choosing.

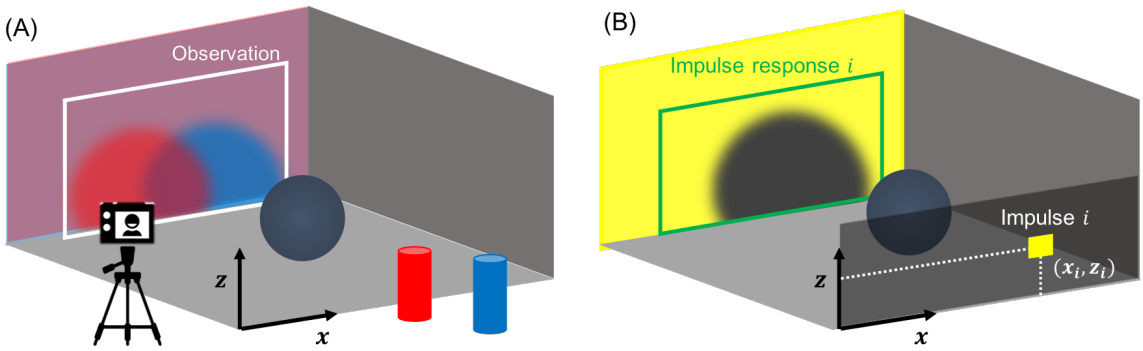

Figure 5-1: We show a schematic of an imaging setup used for the spherical occluder. In (A), we show a sphere in between the observation plane and two colored cylinders. The camera can see the observation plane, but the two cylinders are out of sight. In (B) we show the discrete parameterization of the light from the hidden scene. We choose a grid parallel to the observation plane, placed at a known depth. The i-th column of the transport matrix will be the observed shadow from a point source placed at the i-th location on this grid.

We use this imagined 2D plane to model the light from the hidden scene as imagined light sources on this plane, $L_h(x, y)$. We assume a discretized image of $N = R \times C$ evenly spaced point sources on this plane for reconstruction. While we could take the approach of 3.1 and linearly interpolate between sources to better approximate a continuous $L_h(x, y)$, we take this approach for simplicity.

With this parameterization, we can easily derive the linear relation between the M observed pixels $\mathbf{y}$ and the N hidden scene parameters $\mathbf{x}$. In Figure 5.1(B), we show an impulse response for the i-th point source of $\mathbf{x}$; as before, this rasterized response forms the i-th column of $\mathbf{A}$. Each column i of the transport matrix $\mathbf{A}$ is the observed response of the ground irradiance to an isotropic light placed at the i-th element of $\mathbf{x}$, at the corresponding angle θ_i. We show in Figure 5.1(B) an example of a column of our transfer matrix imposed on a picture of an experimental observation. The reshaped column we show is the impulse response of the middle pixel in our scene.

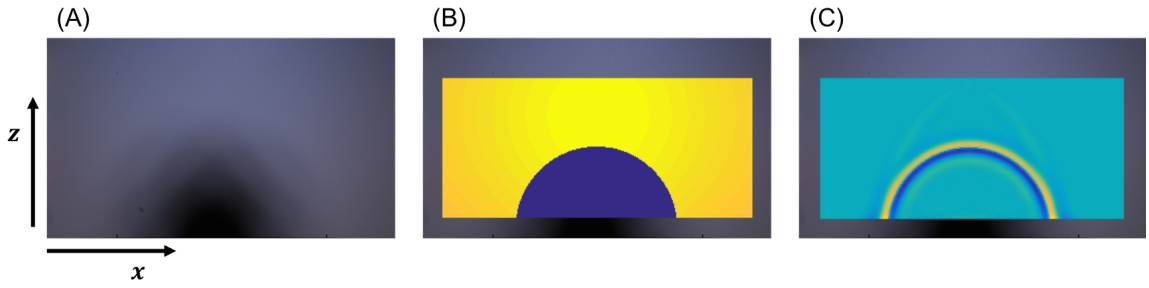

Figure 5-2: (A) An image of the observation plane behind a sphere and scene. The sphere creates a shadow on the observation plane that contains information about the scene behind it. (B) The response of the system to a single point light source in the middle of our scene. These impulse responses form the columns of our transfer matrix $\mathbf{A}$. (C) The estimation gain image of middle pixel in our scene. Namely, the i-th estimation gain image shows what information in the observation is used to reconstruct the i-th pixel in the scene.

5.1.1 Inference

In this work, we recover 2D static images, but this method can easily be extended to videos to recover motion.

Likelihood: We relate the observed M-pixels on the projection plane, $\mathbf{y}$, to the 2D projection of the hidden scene, $L_h(x, y)$, which we discretely approximate with a rasterized grid of $N = R \times C$ terms, $\mathbf{x}$. Assuming i.i.d. Gaussian noise, the relation between $\mathbf{y}^{(t)}$ and $\mathbf{x}^{(t)}$ can be written as

$$\mathbf{y}^{(t)} = L_v + \mathbf{A}\mathbf{x} + \mathbf{w}, \qquad \mathbf{w} \sim \mathcal{N}(0, \lambda\mathbf{I}), \tag{5.1}$$

where L_v is the unmodeled light from the visible scene.

Again, defining $\widetilde{\mathbf{A}}$ be the column augmented matrix $[\mathbf{1}\ \mathbf{A}]$, we can express the likelihood of an observation given $\mathbf{x}$ and L_v as:

$$p(\mathbf{y}|\mathbf{x}, L_v) = \mathcal{N}\left(\widetilde{\mathbf{A}}\left[L_v\ \mathbf{x}^T\right]^T, \lambda\mathbf{I}\right). \tag{5.2}$$

Prior: We enforce spatial smoothness in both dimensions of the hidden image $\mathbf{x}$. That is, we again use an L2 smoothness regularization over the gradients in x and

y. We define the $RC \times RC$ matrix $\mathbf{G}^{\mathbf{x}}_{\mathbf{R},\mathbf{C}}$ as the operator returning the horizontal gradients of the rasterized $R \times C$ image. We define $\mathbf{G}^{\mathbf{y}}_{\mathbf{R},\mathbf{C}}$ similarly. Then, our prior can be expressed as

$$p(\mathbf{x}) = \mathcal{N}\left(0, \sigma^2\left({\mathbf{G}^{\mathbf{x}}_{\mathbf{R},\mathbf{C}}}^T \mathbf{G}^{\mathbf{x}}_{\mathbf{R},\mathbf{C}} + {\mathbf{G}^{\mathbf{y}}_{\mathbf{R},\mathbf{C}}}^T \mathbf{G}^{\mathbf{y}}_{\mathbf{R},\mathbf{C}}\right)^{-1}\right).$$

As before, we note that the visible light L_v has an under-determined relationship with the light on the observation plane and encode this in the posterior below.

Posterior: By combining the defined Gaussian likelihood and prior distributions, we obtain a Gaussian posterior distribution of $\mathbf{x}$ and L_v,

$$\begin{aligned} p(\mathbf{x}, L_v|\mathbf{y}) &= \mathcal{N}\left(\left[\hat{L}_v\ \hat{\mathbf{x}}^T\right]^T, \Sigma\right) \\ \Sigma &= \left[\lambda^{-1}\widetilde{\mathbf{A}}^T\widetilde{\mathbf{A}} + \sigma^{-2}\begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & {\mathbf{G}^{\mathbf{x}}_{\mathbf{R},\mathbf{C}}}^T \mathbf{G}^{\mathbf{x}}_{\mathbf{R},\mathbf{C}}\end{pmatrix} + \sigma^{-2}\begin{pmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & {\mathbf{G}^{\mathbf{y}}_{\mathbf{R},\mathbf{C}}}^T \mathbf{G}^{\mathbf{y}}_{\mathbf{R},\mathbf{C}}\end{pmatrix}\right]^{-1} \\ \left[\hat{L}_v\ \hat{\mathbf{x}}^T\right]^T &= \mathbf{K}\mathbf{y} = \lambda^{-1}\Sigma\widetilde{\mathbf{A}}^T\mathbf{y} \end{aligned} \tag{5.3}$$

where the maximum a posteriori estimate is given by $\hat{\mathbf{x}}$.

We show the operation we perform on $\mathbf{y}$ to recover an element i of $\mathbf{x}$, in Figure 5.1(C). This is one row of the estimation gain matrix, $\mathbf{K}$, visualized to show the pixel-wise contributions of $\mathbf{y}$ to the estimated x_i. We see that for this geometry, we perform a derivative across the edge of the i-th impulse response shown in Figure 5.1(B).

5.1.2 Experimental Results

We test our method with two experimental setups. In the first, shown in Figure 5.1.2(A), we project a hidden scene from a monitor toward a white wall, four feet away. A standard 9 inch soccer ball sits halfway between the monitor and observed wall. The scene is not visible from the camera. The room contains other

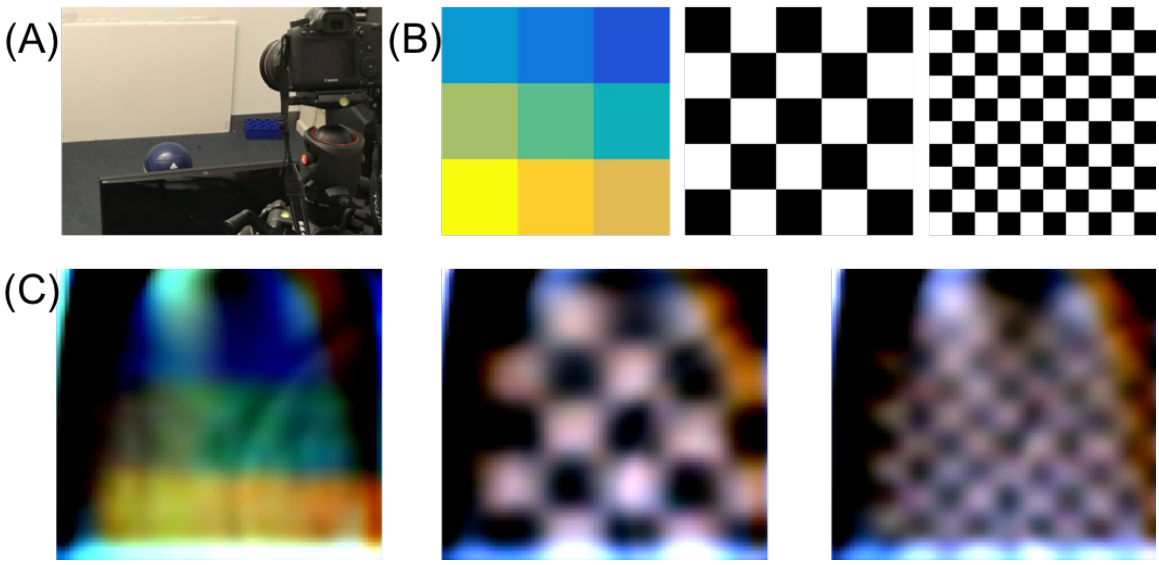

Figure 5-3: We show results for our first experimental scenario. In (A), we show the setup: we project the hidden scene from a monitor onto a blank observation wall, four feet away. A 9 inch soccer ball sits halfway between the two. In (B), we show the hidden scenes we project from the monitor, and in (C), we show the resulting reconstructions.

objects, as well as ambient lighting.

For our reconstructions, we place our 2D projection plane directly at the location of the monitor, four feet away from the wall. Using these location measurements, we then analytically compute the transport matrix $\mathbf{A}$. We note that this method of obtaining $\mathbf{A}$ is error-prone, which creates artifacts in the reconstructions.

We then project the simple scenes in 5.1.2(B) from the monitor, and obtain the static 2D reconstructions shown in Figure 5.1.2(C). We note that our method causes some artifacts, resulting from several sources of error. The first is from the error in the measurements we use to construct $\mathbf{A}$. This error can be combated with a more precise calibration method, which involves capturing each impulse response for the 2D projected grid we wish to reconstruct. We develop and employ this calibration method to reconstruct 4D light fields in Chapter 6.

In our second setup, shown in Figure 5.1.2, we reconstruct two simple objects made of paper, 1 foot tall, illuminated by a large diffuse light. All else remains the same; we continue reconstructing the 2D projection four feet away from the wall, we keep the soccer ball halfway between the two planes, and the room contains ambient

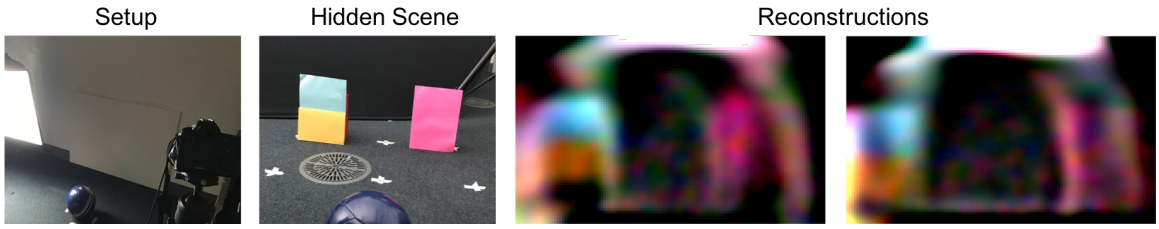

Figure 5-4: We show results for our second experimental setup. On the left, we show the camera capturing a white wall, with a soccer ball placed 2 feet away from it. The hidden scene consists of two simple paper objects, and is illuminated by a large diffuse light, as well as ambient lighting in the room. We show reconstructions of the two hidden objects placed at two depths. On the left, the two objects are placed 5 feet away from the observation wall, and on the right, they are placed 4.5 feet away from the wall.

lighting. We place the pieces of paper at various locations behind our projection plane. We show reconstructions at two locations in Figure 5.1.2. On the left, the two objects were both placed a foot away from our projection plane, 5 feet away from the wall. On the right, the two objects were placed half a foot away from our projection plane, 4.5 feet away from the wall.

In both experimental setups, we notice warping and artifacts at the top of the reconstructions. This may be due to the geometry of the occluder, exacerbated by calibration errors. For a pixel at the top of the reconstruction, the shadow the point source subtends on the wall will have less curvature. This therefore causes ambiguity between that pixel, and the pixel immediately to the right or left of it. We next explore an occluder that does not suffer from these effects: rectangular occluders.

5.2 Rectangular Occlusions

For rectangular occluders, we are again interested in reconstructing 2D projections of a hidden scene. We parameterize the hidden scene in the same way we do in Section 5.1, as a 2D plane of $N = R \times C$ discrete point sources. We also note that a naturally occurring rectangular occlusion is a doorway. As shown in Figure 5.2, the doorway casts structured shadows from light outside a room, onto the ceiling of the room. As such, an observer inside the room can still reconstruct a 2D projection of

the hidden scene outside from the ceiling. For this doorway scenario, we place our projection plane above the scene and reconstruct an aerial view. We note that other placements and orientations of the plane could be better suited to this system.

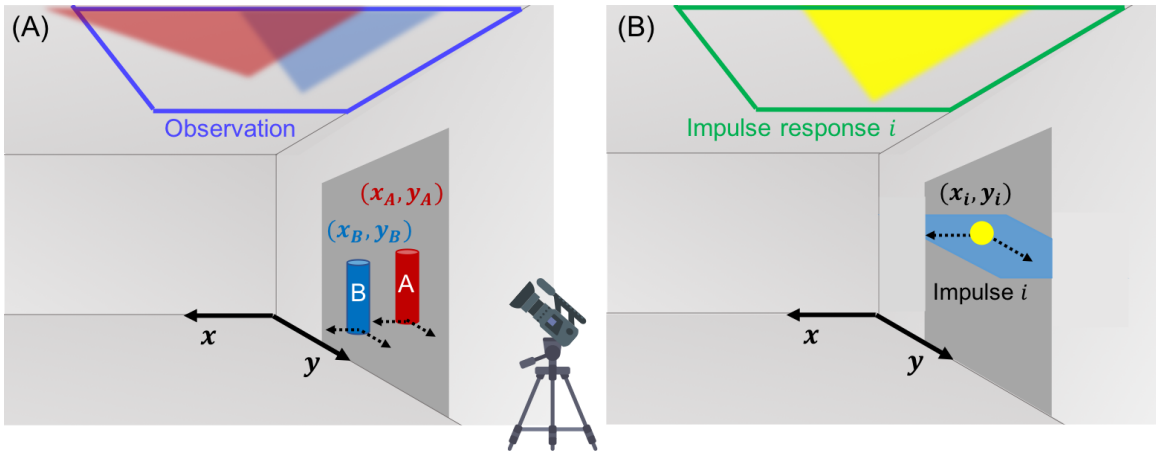

Figure 5-5: We show the imaging setup for imaging a hidden scene through a doorway, from the shadow on the ceiling. In (A), we show two colored cylinders outside a room, casting structured shadows onto the ceiling. The camera sees the ceiling, but cannot see outside the room. In (B), we choose to represent the hidden scene as grid of point sources on a 2D plane placed above the scene at a known height, parallel to the ceiling. The i-th column of the transport matrix will be the observed shadow from a point source at the i-th location on this grid.

5.2.1 Experimental Results

Figure 5-6: We show the experimental results for our first scenario, shown in the leftmost panel. A rectangular occluder is placed 2ft way from the observation plane; a hidden object shown in the second panel is placed at various distances away from it. We show reconstructions of the object placed 5ft away from the wall (third panel) and 6ft away from the wall (right-most panel).

We test our method with two experimental setups. In the first, shown in Figure 5.2.1, a rectangular occluder sits 2 feet away from the wall. As before, we choose

our 2D projection plane to be 4 feet away from the wall. We reconstruct a foot-tall hidden object made of colored paper, placed at various depths away from the wall. The object is illuminated by a diffuse light, and there is ambient lighting in the room. As before, we compute the light transport matrix $\mathbf{A}$ from the measurements and dimensions of the system.

On the left, the hidden object is 5 feet away from the wall, 1 foot away from the projection plane. On the right, the hidden object is 6 feet away from the wall, 2 feet away from the projection plane. We note that we can recover faithful reconstructions of the hidden object even at large distances away from the wall.

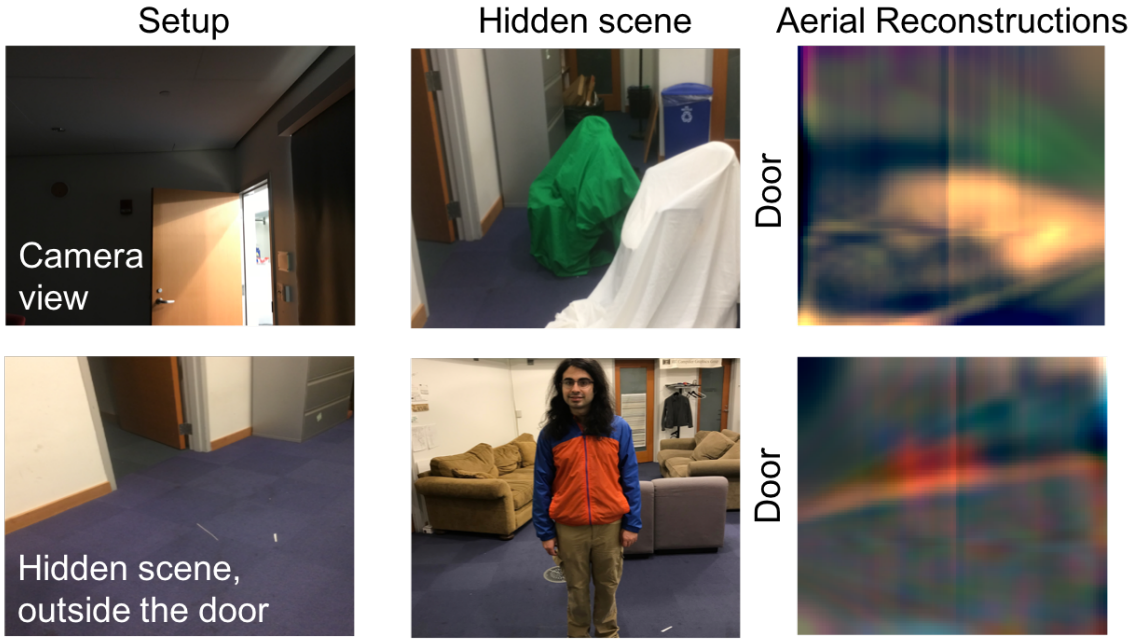

Figure 5-7: We show the experimental results for our second scenario: imaging a hidden scene outside a doorway. The camera inside the dark room sees the ceiling, and cannot see the scene outside the door. We reconstruct two scenes. The first scene is of two covered chairs, one green and one white, placed a few feet from the door. The second is of a man wearing orange, standing in the same location (the hidden scene is viewed from through the doorway).

In our second setup, shown in Figure 5.2.1, we take observations of light from outside a room on the ceiling of a dark room. We reconstruct the hidden scene outside the door. For this scenario, we take measurements of the door and room to construct $\mathbf{A}$. The reconstructions of two hidden scenes are shown in Figure 5.2.1.

The first is of two covered chairs, one white and one green, placed outside the door. the second is of a man wearing orange, standing outside the door. We note that the aerial reconstructions of the two scenes contain spurious artifacts, but capture the colors, relative shapes, and locations of the hidden objects.

We note that in order to compute the transfer matrix $\mathbf{A}$, we need several measurements about the geometry of the system. In the experiments presented, we compute the transport matrix from measurements and dimensions of the system. However, if we have more complex occluders, we need a more precise way to compute the transport matrix. We explore this problem in order to recover 4D light fields in Chapter 6.

Chapter 6

Recovering 4D Light Fields

In this chapter, we employ this approach of using the 2D shadows cast by known occluders to recover the 4D light field of a hidden scene [2]. The recovered light field would allow us to view the hidden scene from multiple perspectives and thereby recover depth information within the scene.

The first step toward recovering 4D light fields from 2D shadows is formulating the linear light transport model, much in the vein of the rest of this thesis. The light transport matrix **A** for an arbitrary light field is depicted schematically in 6-1.

As the light transport matrix **A** is sparse and row-deficient, our model in this scenario is extremely underdetermined. Unlike in the scenarios described in previous chapters, we must dramatically reduce the dimensionality of the inverse problem. As such, we must choose a prior distribution on light fields that turns our original ill-posed problem into one that is well-posed and computationally feasible. Light field sampling theory [6] and light field priors [14] inform how we translate mild assumptions about the hidden scene light field into a reasonable prior.

In this thesis, we primarily discuss the first aspect of this imaging problem, as it relates to the work discussed thus far. [2] contains further discussion and derivation of the inference process; in this thesis, we will cite the equations derived from that work.

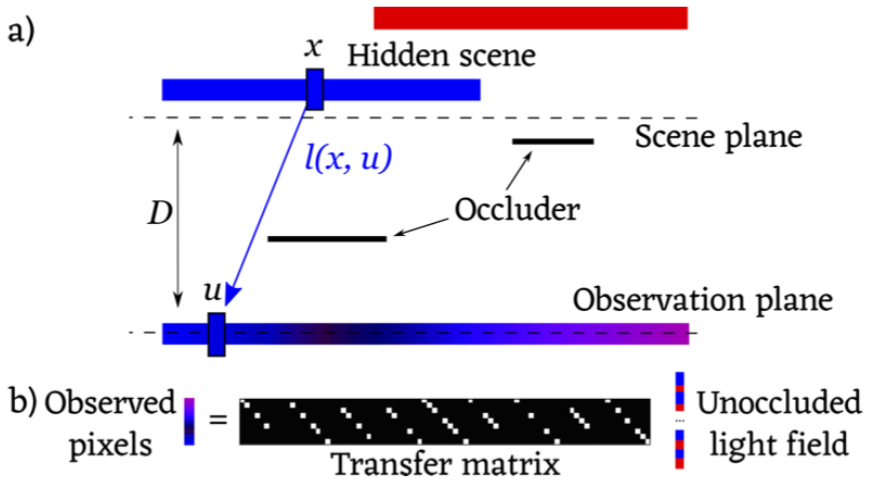

Figure 6-1: a) The simplified 2D scenario. The scenario depicts all the elements of the scene (occluder, hidden scene and observation plane) and the parametrization planes for the light field (dashed lines). In (b) we show the discretized version of the scenario, with the light field and the observation 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.

6.1 Method

6.1.1 Light Field Model

We parameterize the 2D light field $L_h(x, u)$ using the two-plane parametrization [15], as shown in Figure 6-1. 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. We discretize the unoccluded light field $L_h(x, u)$ as the rasterized vector $\mathbf{x}$, using M uniform samples from the scene and N uniform samples from the observation plane. We note that with this parameterization, the element $\mathbf{x}_{Ni+j}$ is the light ray between the i-th sample of the scene and the j-th sample of the observation. 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. 6-1b):

$$\mathbf{A}_{k,Ni+j} = \begin{cases} c_{Ni+j} \cdot v_{Ni+j} & k = j \\ 0 & \text{else} \end{cases}, \tag{6.1}$$

where c_{Ni+j} takes into account distance attenuation and other near-field effects.

As before, each column of $\mathbf{A}$ can be thought of the observation we see for an impulse placed at the $(Ni+j)$-th element of $\mathbf{x}$. In this case, we imagine an impulse as a single laser beam, and the resulting observation in each column either a single point of light, if the $(Ni+j)$-th light ray of $\mathbf{x}$ is not occluded, or completely dark, if it is.

6.1.2 Light Field Inference

As before, we seek a MAP estimate of the hidden scene parameters $\mathbf{x}$, given the M pixels $\mathbf{y}$ observed by the camera. However, we now formulate the problem in the Fourier domain of the light field, in terms of spectral variables. As derived in [2], assuming i.i.d. Gaussian noise with variance λ, the likelihood of the observations $\mathbf{y}$ given the spectrum of $\mathbf{x}$, $\mathbf{f}_\mathbf{x}$, can be expressed using the $N \times N$ DFT matrix $\mathbf{F}$:

$$p(\mathbf{y}|\mathbf{f}_\mathbf{x}) = \mathcal{N}(\mathbf{B}\mathbf{f}_\mathbf{x}, \lambda\mathbf{I}) \tag{6.2}$$

$$\mathbf{B} = \mathbf{A}\mathbf{F}^{-1}, \tag{6.3}$$

where $\mathbf{f}_\mathbf{x} = \mathbf{F}\mathbf{x}$. For a Gaussian prior with mean 0 and covariance $\mathbf{C}_{\mathbf{f}_\mathbf{x}}$ on the spectrum $\mathbf{f}_\mathbf{x}$, our MAP estimate of $\mathbf{f}_\mathbf{x}$ is then

$$\hat{\mathbf{f}}_\mathbf{x} = \left[\lambda^{-1}\mathbf{B}^H\mathbf{B} + \mathbf{C}_{\mathbf{f}_\mathbf{x}}^H\mathbf{C}_{\mathbf{f}_\mathbf{x}}\right]^{-1}\mathbf{B}^H\mathbf{y}. \tag{6.4}$$

We choose in a prior that assumes the scene is composed of planar and diffuse elements with no occlusions between elements whose textures are independent of their position. These assumptions can be formulated in the Fourier domain; we provide the derivation of the final prior distribution in Appendix A.3.

6.1.3 Occluder Calibration

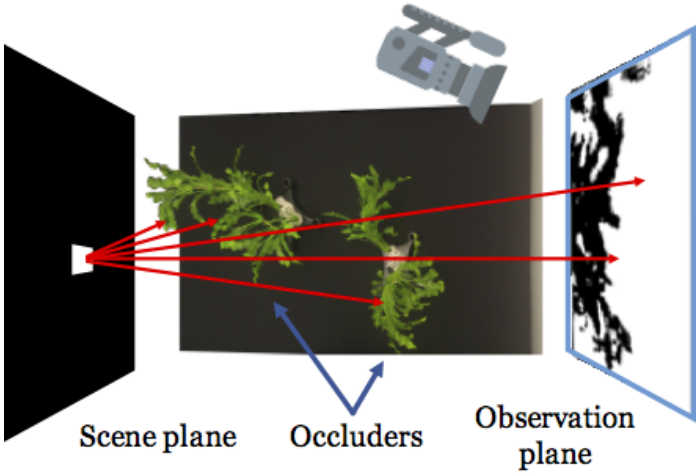

Figure 6-2: Experimental setup used to calibrate complex occluders. An impulse at a single position on an LCD screen placed at scene plane casts a shadow. This shadow is used to compute the visibility function of that scene position and all observation positions. This process is done for each scene position.

In our experiments, we reconstruct a hidden scene in the depth range between $a = 0$ and $a = 0.5$; that is, we assume the elements of interest are within a factor of 1.5 of the minimum depth D (for D as defined in 6-1). This assumption allows for reasonable reconstructions given the finite samples we take. We take 150×225 samples of the observation plane, and 14×23 samples of the scene plane.

For these experiments, we use two desk plants as occluders, shown in Figure 6-2. 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 capture the observation obtained by lighting up each location on the scene plane (in practice this is done with an LCD monitor at the scene plane). We then convert these observations into binary masks by thresholding pixel intensity. These masks approximately encode which light rays from the scene plane reach the observation plane. The transfer matrix is derived from the visibility function, accounting for near-field effects. This process is schematically shown in Figure 6-2.

We note that it is possible to directly use the impulse responses in the transfer

matrix. However, doing so would encode calibration conditions, such as the LCD monitor BRDF, that may not hold true in the experiment.

6.2 Experimental Reconstructions

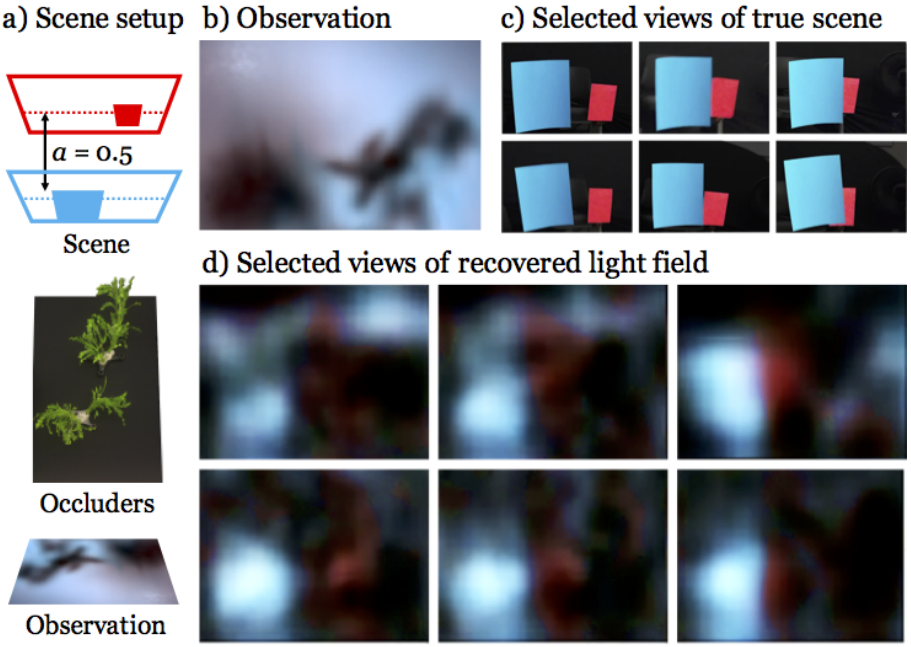

Figure 6-3: Reconstructions of an experimental scene with two rectangles, one blue at $a = 0$ and one red at $a = 0.5$. (a) Schematic of the system setup. (b) Observation plane after background subtraction. (c) Six views of the true scene, shown to demonstrate of what the true light field would look like. These are taken with a standard camera from equivalent positions on the observation wall. The "camera" moves right as we move left to right on the grid, and down as we move top to bottom on the grid. (d) These selected views of the reconstructed light field.

In Figures 6-3 and 6-4, we show six selected views of recovered light fields of different target scenes, with different levels of spatial (texture-related) and angular (parallax-related) in the complexity of the hidden scene. Both are obtained using the same prior, without adapting the depth range to recover the particulars of the scene. The results demonstrate faithful reconstruction of both texture and albedo information, as well as parallax consistency with the expected light fields. In Figure 6-3, we see the blue and red squares of the true scene. Moreover, we see the red square

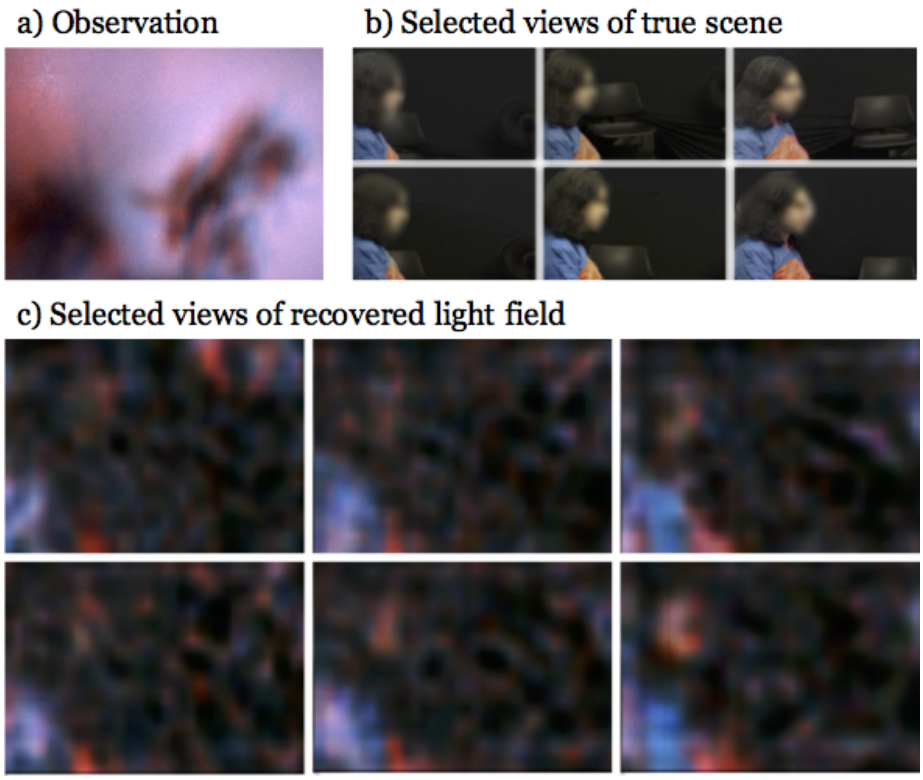

Figure 6-4: Reconstructions of an experimental scene with a single subject, seated at the scene plane. (a) Observation plane after background subtraction. (b) Six views of the true scene, captured in the same manner as in Fig. 6-3 (c) Reconstructions of the light field for the same views as in (b).

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 Figure 6-4, 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.

Chapter 7

Conclusion

We show how everyday objects can act as cameras for the surrounding environment. We first show how to turn corners into cameras, exploiting a common, but overlooked, visual signal. The vertical edge of a corner selectively blocks light to let the ground nearby display an angular integral of light from around the corner. The resulting penumbras from people and objects are invisible to the eye – typical contrasts are 0.1% above background – but are easy to measure using consumer-grade cameras. We produce 1-D videos of activity around the corner, measured indoors, outdoors, in both sunlight and shade, from brick, tile, wood, and asphalt floors. The resulting 1-D videos reveal the number of people moving around the corner, their angular sizes and speeds, and a temporal summary of activity. Since nearly every corner now offers a 1-D view around the corner, this opens potential applications for automotive pedestrian safety, search and rescue, and public safety.

We then explore extensions of this corner camera toward two applications. We show how to extend the corner camera from a static camera observer to a moving camera observer. We address the challenges of a moving camera and changing viewing angle, and offer suggestions for continuing to adapt the method in this way. We also implement a real-time, publicly available corner camera. We then show how open doorways with two vertical edges offer stereo views inside a room, visible even away from the doorway. Directly modeling more complex occluders also allow us to recover more dimensions of information from the hidden scene. We show the imaging

capabilities of spherical and rectangular occluders to recover 2D projections of hidden scenes. In particular, the structured light filtered through a doorway onto the ceiling can allow us to recover a 2D view of the scene outside the door.

Finally, we show how we can use even more complex occluders, such as plants, to recover 4D light fields from shadows. 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. This is the first method known 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], [17], and [22] focus on extracting the size and shape of objects. Our methods and experimental results demonstrate the fact that the hidden patterns of shadows on walls can, if they are properly interpreted, reveal a rich world of previously invisible visual information.

Appendix A

Additional Derivations

A.1 Simple Perturbation Model

This section analyzes the signal on the observation plane for the corner camera system. We model three sources of light in the image, shown in Fig. A-1: a hemisphere of external illumination of uniform brightness A; a wall of brightness B; and an imaging subject of brightness C. The wall is assumed to be infinitely tall (relative to the scale of the imaged region of the ground), and the height, width, and location of the cylindrical subject are as shown in the figure. We assume that all surfaces are Lambertian.

While this model is quite simple, it allows us to examine the signal strength–the change of ground brightness due to the presence of the subject–for many cases of interest. We can calculate

- the signal as a function of a person's (the subject's) distance away from the corner, and position relative to the corner;
- the signal as a function of position on the observation plane;
- the approximate effects of more or less sunlight.
- the effect of the wall's brightness on the corner camera.

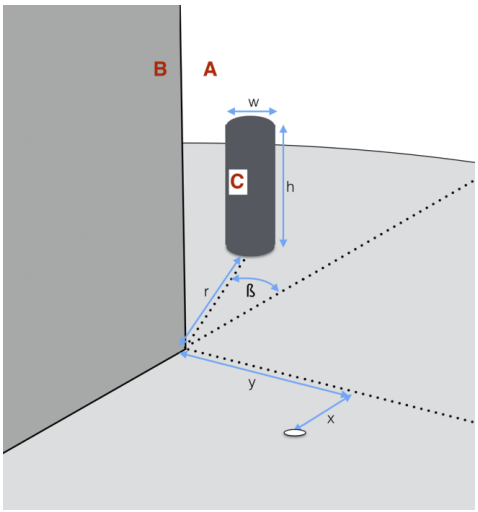

Figure A-1: Simple model of a corner camera. We seek formulas for the image intensities at a point (x, y) in the observation plane. We have a wall of brightness B, a subject of brightness C, and a uniform background of brightness A.

A.1.1 Photometric signal of the corner camera

We assume a moving subject. We want to calculate the ground plane intensities as a function of position, both with and without the subject present. The difference in the brightnesses observed on the ground plane will then tell us the requirements for our photometric measurements to enable imaging from a corner camera.

Intensity contributions of the wall of brightness B

Consider a small patch on the wall at position x', z'. To calculate its brightness contribution to the brightness on the ground at position (x, y), we need to calculate the normal vector from the ground position to the wall patch, $\hat{N}_{gw}$. The dot product of that normal vector with the ground's normal controls the brightness contribution from the point (x', z').

But if the wall is Lambertian, and the angle at which we view it doesn't change its brightness, B, then we can just as easily, and more simply, perform the integration from the observation plane point (x, y). There we just have to integrate the brightness B over all directions from (x, y) where we see the wall. Let θ be the angle of a ray to the wall with respect to the ground plane normal, where the z-axis has $\theta = \frac{\pi}{2}$. Let ϕ

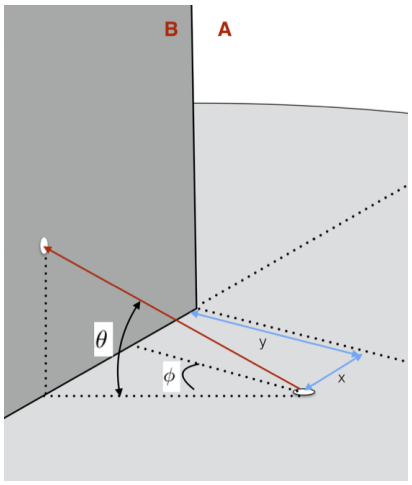

Figure A-2: Brightness contribution of one patch of the wall to the brightness of the ground plane depends on the angle between the red arrow and the ground's surface normal.

be the azimuthal angle to the point on the wall, where the y-axis direction has $\phi = 0$.

Then the brightness contribution from the entire wall, I_B, is the following double-integral

$$I_B = B \int_{-\frac{\pi}{2}}^{\arctan(\frac{x}{y})} \mathrm{d}\phi \int_0^{\frac{\pi}{2}} \sin(\theta)\mathrm{d}\theta \tag{A.1}$$

$$= B \int_{-\frac{\pi}{2}}^{\arctan(\frac{x}{y})} \mathrm{d}\phi \tag{A.2}$$

$$= B \int_{-\frac{\pi}{2}}^{\arctan(\frac{x}{y})} \mathrm{d}\phi \tag{A.3}$$

$$= B \left(\arctan\left(\frac{x}{y}\right) + \frac{\pi}{2} \right) \tag{A.4}$$

Intensity contributions of the non-wall environment of brightness A

Similarly, we want to integrate the background brightness A over all angles where we see A and not B from the observation plane patch at (x, y). Calling this brightness I_A, this is the same integral as above, except we are integrating the constant A, not

B, and over the complement of the azimuthal angles as before. So the result is

$$I_A = A\left(2\pi - \left(\arctan\left(\frac{x}{y}\right) + \frac{\pi}{2}\right)\right) \quad \text{(A.5)}$$

$$= A\left(\frac{3}{2}\pi - \arctan\left(\frac{x}{y}\right)\right) \quad \text{(A.6)}$$

$$\text{(A.7)}$$

A.1.2 Intensity contributions of the subject of brightness C

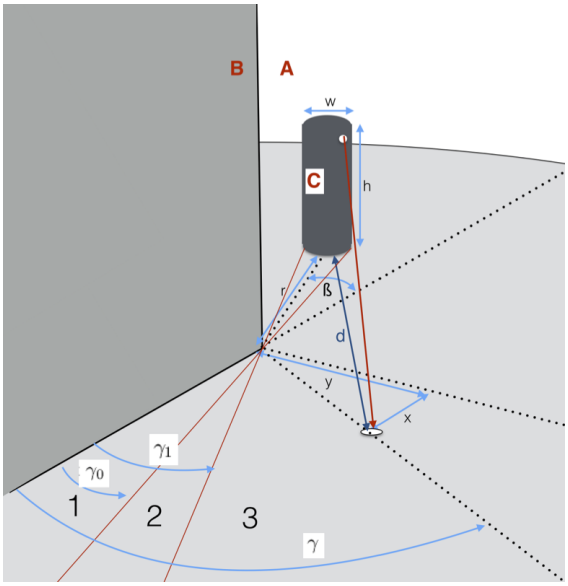

Figure A-3: Calculation of the contribution to the brightness of the subject.

The geometry of the edge and the subject divides the observation plane into three regions, with different intensity functions in each region. Let $\gamma = \arctan(\frac{y}{x})$. Let $\gamma_0 = \beta - \frac{w}{r}$ and $\gamma_1 = \beta + \frac{w}{r}$. Points of the observation plane for $\gamma < \gamma_0$ (using the small angle approximation for the angular extent of the cylinder) don't see the cylinder and form region 1 of the ground plane. Points where $\gamma > \gamma_1$ see a full view of the subject cylinder and form region 3 of the ground plane. Points (x, y) for $\gamma_0 < \gamma(x, y) < \gamma_1$ form region 2 of the ground plane. We derive the formulas for the intensities within each region.

I_{C1}, for ground plane region 1

Points on the ground plane in region 1 never see the subject and so their intensities are governed by the function $I_{C1}(x, y) = I(x, y) = I_A(x, y) + I_B(x, y)$.

I_{C3}, for ground plane region 3

Points in region 3 have a full view of the subject, not clipped by the wall edge. The subject has brightness C over its cross-section, and occludes the background brightness A over the same cross-section. So its brightness contribution, I_C, will be the brightness difference $C - A$, integrated over the cross-section of the subject. Then the total brightness at the point (x, y), with the subject in the scene, $I_T(x, y)$, will be

$$I_T(x, y) = I_A + I_B + I_{C3} \tag{A.8}$$

Then the total brightness at the point (x, y), without the subject present, $I(x, y)$, will be

$$I(x, y) = I_A + I_B \tag{A.9}$$

Since it is also assumed to be Lambertian, it will be easier to perform this integration from the point of view of the ground plane patch at (x, y). We first want to calculate the distance, d, from the ground plane patch at (x, y) to the subject.

The x-coordinate difference from the subject to the patch, d_x is

$$d_x = r\cos(\beta) + x \tag{A.10}$$

The y-coordinate difference from the subject to the patch, d_y is

$$d_y = r\sin(\beta) + y \tag{A.11}$$

Then we have, for the distance d from the subject to the patch,

$$d = \sqrt{d_x^2 + d_y^2} \tag{A.12}$$

$$= \sqrt{(r\cos(\beta) + x)^2 + (r\sin(\beta) + y)^2} \tag{A.13}$$

Now, we integrate the brightness difference, $C - A$, over the cylinder of width and height w and h at a distance d from the patch at (x, y). We simplify the cylinder to be a w by h rectangle, as seen from the patch at (x, y). If we assume the ground plane is Lambertian, then intensities arriving at the patch are modulated by a factor of $\sin(\theta)$:

$$I_{C3} = (C - A)\int_0^{\arctan(\frac{h}{d})} \sin(\theta)\mathrm{d}\theta \int_k^{k+\frac{w}{d}} \mathrm{d}\phi \tag{A.14}$$

$$= (C - A)\frac{w}{d}\left(1 - \cos\left(\arctan\left(\frac{h}{d}\right)\right)\right) \tag{A.15}$$

$$= (C - A)\frac{w}{d}\left(1 - \frac{d}{\sqrt{d^2 + h^2}}\right) \tag{A.16}$$

where, for simplicity in the integration limit for ϕ, we assume that $w \ll d$, and k an offset in ϕ that doesn't matter in the result. And we used $\cos(\arctan(x)) = \frac{1}{\sqrt{1+x^2}}$.

I_{C2}, for ground plane region 2

Points in region 2 only see a partial image of the subject. The integral over ϕ in Eq. (A.16) has a different bottom limit depending on the point's angular position within region 2.

The partial view of the subject will lead to this simple behavior of the intensities in region 2: at a constant radius away from the edge (and therefore a constant distance d from the subject), the intensities in region 2 will linearly interpolate between the intensity values at the boundaries with regions 1 and 3.

For a given (x,y) in region 2:

1. Find $\gamma(x, y)$.
2. Find $\rho(x, y) = \sqrt{x^2 + y^2}$

3. Use $I_{C2}(\gamma, \rho) = \alpha I_{C2}(\gamma_0, \rho) + (1 - \alpha) I_{C2}(\gamma_1, \rho)$, where

4. $\alpha = \frac{\gamma - \gamma_0}{\gamma_1 - \gamma_0}$

The linear behavior of the limits of integration over ϕ in Eq. (A.16) allow this linear interpolation to hold.

A.1.3 Final Model

Brightness at (x,y) with subject present in scene, $I_T(x, y)$

$$
\begin{aligned}
I_T &= I_A + I_B + I_C && \text{(A.17)} \\
&= A\left(\frac{3}{2}\pi - \arctan\left(\frac{x}{y}\right)\right) + && \text{(A.18)} \\
&\quad B\left(\arctan\left(\frac{x}{y}\right) + \frac{\pi}{2}\right) + && \text{(A.19)} \\
&\quad I_C(x, y) && \text{(A.20)}
\end{aligned}
$$

Brightness at (x,y) with no subject in scene, $I(x, y)$

$$
\begin{aligned}
I &= I_A + I_B && \text{(A.21)} \\
&= A\left(\frac{3}{2}\pi - \arctan\left(\frac{x}{y}\right)\right) + && \text{(A.22)} \\
&\quad B\left(\arctan\left(\frac{x}{y}\right) + \frac{\pi}{2}\right) && \text{(A.23)}
\end{aligned}
$$

Fractional difference in brightness at (x,y) due to subject, $\frac{I_T(x,y)}{I_T(x,y)}$

$$
\begin{aligned}
\frac{I_T(x, y)}{I(x, y)} &= \frac{I_A + I_B + I_C}{I_A + I_B} && \text{(A.24)} \\
& && \text{(A.25)}
\end{aligned}
$$

A.1.4 Examples

Typical brightness change induced by a person 10 feet from corner

Let's examine values we consider to be typical for an outdoor, cloudy day scene. Let's say the surroundings have brightness $A = 300$, the wall has brightness $B = 100$, and the person has brightness $C = 150$ (arbitrary units, but roughly mapping to pixel intensities from a photograph). The person is $h = 5.5$ feet tall, and $w = 1.5$ feet wide, positioned at angle $\beta = \frac{\pi}{4}$, and at distance from the corner $r = 10$ feet.

Let's gaze upon the observation plane a distance $\rho = \sqrt{x^2 + y^2}$ from the corner. At the boundary between ground plane regions 1 and 2, we have, for the ground plane image intensity,

$$
\begin{aligned}
I_{C1}(\rho, \gamma_0) &= I_A(x, y) + I_B(x, y) && \text{(A.26)} \\
&= A\left(\frac{3}{2}\pi - \arctan(\frac{x}{y})\right) + B\left(\arctan\left(\frac{x}{y}\right) + \frac{\pi}{2}\right) && \text{(A.27)} \\
&= 300\left(\frac{3}{2}\pi - (\pi - \gamma_0)\right) + 100\left((\pi - \gamma_0) + \frac{\pi}{2}\right) && \text{(A.28)} \\
&= 300\left(\frac{\pi}{2} + \gamma_0\right) + 100\left(\frac{3\pi}{2} - \gamma_0\right) && \text{(A.29)} \\
& && \text{(A.30)}
\end{aligned}
$$

Note this intensity doesn't depend on the distance away from the wall, at a constant angle $\gamma = \gamma_0$. The wall subtends a constant solid angle for all ground plane points along this straight line from the corner. Note that $\arctan(\frac{x}{y}) = \pi - \gamma_0$.

Now let's examine the brightness at the region 2/region 3 boundary, at $\gamma = \gamma_1$. There we have

$$
\begin{aligned}
I_T &= I_A(x, y) + I_B(x, y) + I_{C3}(x, y) && \text{(A.31)} \\
&= I_{C1}(\rho, \gamma_0) + I_{C3}(x, y) && \text{(A.32)} \\
&= 300\left(\frac{\pi}{2} + \gamma_0\right) + 100\left(\frac{3\pi}{2} - \gamma_0\right) + && \text{(A.33)} \\
&= (C - A)\frac{w}{d}\left(1 - \frac{d}{\sqrt{d^2 + h^2}}\right) && \text{(A.34)}
\end{aligned}
$$

Remembering that $\gamma_0 = \frac{\pi}{4} - \frac{w}{r}$ and $w = 1.5$, $h = 5.5$, $r = 10$, $C = 150$, $B = 100$, $A = 300$, we have

$$\begin{aligned} I_T &= 300\left(\frac{\pi}{2} + \frac{\pi}{4} - \frac{1.5}{10}\right) + && \text{(A.35)} \\ & \quad 100\left(\frac{3\pi}{2} - \left(\frac{\pi}{4} - \frac{1.5}{10}\right)\right) + && \text{(A.36)} \\ & \quad (-150)\frac{1.5}{d}\left(1 - \frac{d}{\sqrt{d^2 + 5.5^2}}\right) && \text{(A.37)} \end{aligned}$$

Remember d is the distance from (x, y) to the cylinder, or $10 + \sqrt{x^2 + y^2}$. Let's look at $d = 12$. For this distance, we have

$$\begin{aligned} I_T &= 300(2.356 - 0.15) + && \text{(A.38)} \\ & \quad 100(4.7124 - 0.6354) + && \text{(A.39)} \\ & \quad (-150)\frac{1.5}{12}\left(1 - \frac{12}{\sqrt{12^2 + 5.5^2}}\right) && \text{(A.40)} \\ &= 662 + 408 - 1.7 && \text{(A.41)} \end{aligned}$$

That's in the right ballpark: The signal without the subject has brightness 1070. The signal with the subject has brightness 1068.3 requiring a measurement dynamic range of $\frac{1.7}{1070}$. Thus, we seek a signal change that is 0.16% of the DC signal. This is consistent (within a factor of two) with what we find to be typical for real images, $\frac{1}{1000}$,

Looking at the subject's signal as a function of ground plane distance away from the edge, ρ, we have $I_{C3}(\rho) = \frac{412.5}{(10+\rho)\sqrt{(10+\rho)^2+5.5^2}}$. For $\rho = 0$, we have $I_{C3}(\rho) = 3.6$ (a higher subject signal). For $\rho = 4$, we have $I_{C3}(\rho) = 1.96$ (a lower subject signal).

A.1.5 Rate of signal decay as person moves away from corner

As the person walks away from the corner, we lose signal strength for two independent, reinforcing reasons. One, from perspective projection, they subtend a smaller solid angle, with linear dimensions going as $\frac{1}{d}$ and areal dimensions contributing to brightness going down as $\frac{1}{d^2}$, where d is the distance from the corner. Second, as the

person moves away from the corner, the light from the person comes in at a shallower and shallower angle, further reducing the signal intensity. The grazing angle to the horizontal will scale as $\frac{1}{d}$, but the brightness contribution will be $1 - \cos(\theta)$, which scales as $\frac{1}{d^2}$. Thus, the delta in brightness, from a person walking away from the corner, scales as $\frac{1}{d^4}$.

This is also apparent from applying a 2nd order Taylor series expansion to the square roots with d in them in Eq. (A.37).

A.2 Temporal Smoothing for the Corner Camera

In addition to spatial smoothness we can also reduce noise temporally by imposing smoothness on our MAP estimate. $\hat{\mathbf{x}}^{(t)}$, or averaging adjacent frames in time. This helps to reduce noise, at the cost of some temporal blurring. However, to emphasize the coherence among results, we do not impose this additional constraint in results shown in Chapter 3. Each 1-D slice, $\mathbf{x}$, that we show is independently computed.

In Figure A-4, we show some results obtained with temporal smoothing. We either computed the 1-D video on averaged video frames, or applied Kalman filter and smoothing to temporally smooth the result.

Kalman Filtering and Smoothing We model our scene as a linear dynamic system.

$$x_t = Fx_{t-1} + w_k \tag{A.42}$$

$$y_t = Ax_t + n_t \tag{A.43}$$

$$w_k \sim \mathcal{N}(0, R) \tag{A.44}$$

$$n_t \sim \mathcal{N}(0, \lambda \mathbf{I}) \tag{A.45}$$

The matrix F can be set as the identity matrix so changes over time are penalized. The marginals $p(x_t|y_1, ..., y_n)$ of this HMM can then be solved using the forward

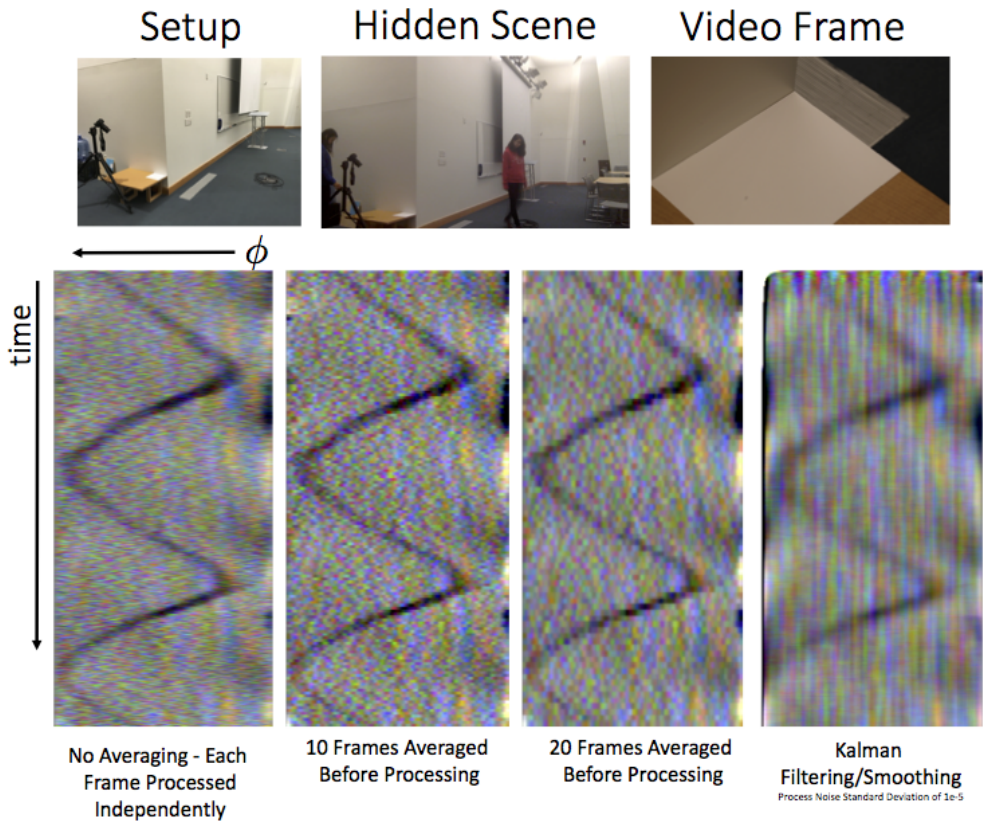

Figure A-4: The result of imposing temporal smoothness or averaging adjacent frames in time to help in reducing noise.

backward algorithm - or equivalently Kalman filtering and smoothing.

Note that the Kalman filtering and smoothing causes the noise to blur out into vertical lines. Frame averaging results appear noisier only because there are fewer 1D reconstructed frames being shown.

A.3 Derivation of the Light Field Prior

We provide the derivation and final prior distribution used in recovering light fields from shadows.

The prior we propose in [2] is derived using the following assumptions: the scene is composed of planar and diffuse elements with no occlusions between elements whose textures are produced by a random process that is independent of their position.

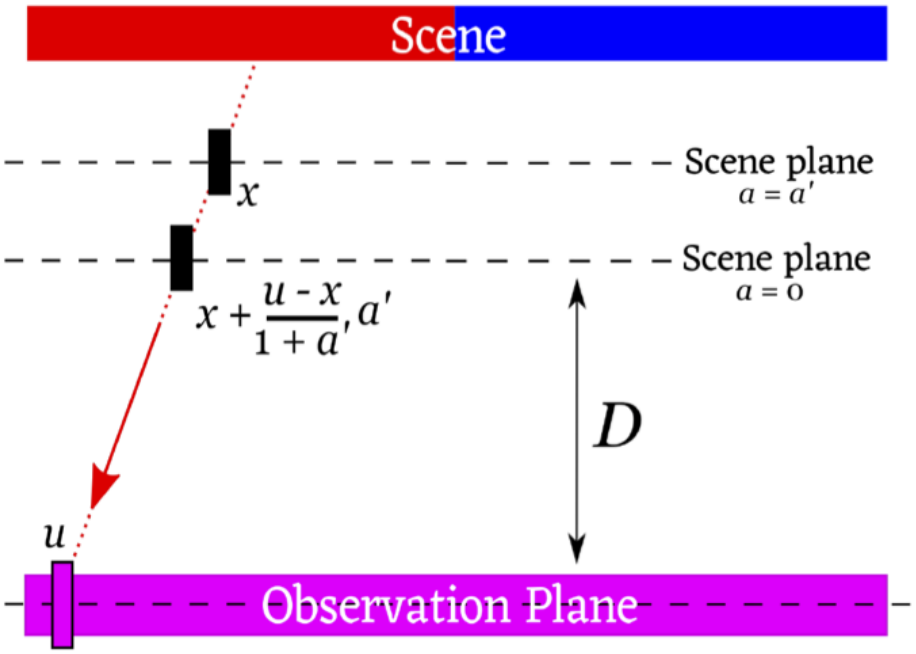

Figure A-5: Light field reparameterization by a displacement of the scene plane. For any u, x, and a, $l_a(x, u) = l(x + \frac{u-x}{1+a}a, u)$.

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 $\mathbf{f}_\mathbf{x}^i$ is also Gaussian and independent, following:

$$P(\mathbf{f}_\mathbf{x}^i) \propto P_m(\mathbf{f}_\mathbf{x}^i) P_t(\mathbf{f}_\mathbf{x}^i) \tag{A.46}$$

A.3.1 Reparameterization of light fields

To characterize our assumptions, it is necessary to model light field reparameterization of the scene plane. Given a light field $l(x, u)$, we define the reparameterized 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 A-5. The relation between l_a and $l(x, u)$ is:

$$l_a(x,u) = l\left(x + \frac{u-x}{1+a}a, u\right) \tag{A.47}$$

This reparameterization 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 [**?**]), following:

$$L_a(f_x, f_u) = (1+a)L((1+a)f_x, f_u - f_x a) \tag{A.48}$$

A.3.2 Dimensionality gap of 4D light fields

For a single diffuse 2D observation at depth $a = a'$, its spectrum $L_{a'}$ has non-zero power only for $f_u = 0$. Following the light field reparameterization derived in A.48, the spectrum for the light field of this same planar texture but at depth $a = 0$, is a sheared (and scaled) version of $L_{a'}$. The spectrum L for that 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 reparameterization L is proportional to a' in both dimensions, and follows:

$$\begin{aligned} f_x &= a' f_u \\ f_y &= a' f_v \end{aligned} \tag{A.49}$$

Considering only planar and parallel textures at any a, the equations of A.49 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 the equations of A.49.

A.3.3 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 [5], 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:

$$\left\{ \begin{array}{ll} 1/f_x^{\gamma} & \text{if } f_u = 0 \\ 0 & \text{else} \end{array} \right\} \tag{A.50}$$

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

$$P(L_a(f_x, 0)) = N(0, 1/f_x^{\gamma}) \tag{A.51}$$

This equation expresses the fact that, if we had a light field created by a single planar element at depth $a = a'$ and we placed the scene plane at depth $a = a'$, 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 reparameterization of light fields formulated in A.48 and the relationship between a, f_x and f_u ($a = \frac{f_x}{f_u}$), we extend the principle of Eq. A.52 to the whole light field, assuming no mutual occlusions between parallel textures occur (even though the real scene may have mutual occlusions).

$$L(f_x, f_u) = \frac{1}{1+a} L_a \left(\frac{1}{1+a} f_x, 0 \right)$$
$$P_t(L(f_x, f_u)) = N \left(0, \left(\frac{1}{1+\frac{f_x}{f_u}} \right)^{1-\gamma} f_x^{-\gamma} \right) \tag{A.52}$$

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

$$P_t(L(f_x, f_y, f_u, f_v)) =$$
$$N \left(0, \left(\frac{1}{1+\frac{f_x}{f_u}} \right)^{1-\gamma} (f_x^2 + f_y^2)^{-\gamma/2} \right) \tag{A.53}$$

Bibliography

[1] F. Adib and D. Katabi. See through walls with wifi! *ACM*, 43(4):75–86, 2013.

[2] M. Baradad, V. Ye, A. Yedidia, G.W. Wornell, W.T. Freeman, F. Durand, and A. Torralba. Inferring light fields from shadows. In *In Proc. IEEE CVPR*, pages 1–8, 2018.

[3] P.V. Borges, A. Tews, and D. Haddon. Pedestrian detection in industrial environments: Seeing around corners. 2012.

[4] Katie Bouman, Vickie Ye, Adam Yedidia, Fredo Durand, Gregory Wornell, Antonio Torralba, and William T. Freeman. Turning corners into cameras: Principles and methods. *International Conference on Computer Vision*, 2017.

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

[6] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung 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.

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

[8] R. Fergus, A. Torralba, and W.T. Freeman. Random lens imaging. 2006.

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

[10] Itseez. Open source computer vision library, 2015.

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

[12] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

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

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

[15] Marc Levoy and Pat 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.

[16] K. Nishino and S.K. Nayar. Corneal imaging system: Environment from eyes. *International Journal of Computer Vision*, 70(1):23–40, 2006.

[17] R. Pandharkar, A. Velten, A. Bardagjy, Bawendi M. Lawson, E., and R. Raskar. Estimating motion and size of moving non-line-of-sight objects in cluttered environments, 2011. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on (pp. 265-272).

[18] D. Shin, A. Kirmani, V.K. Goyal, and J.H. Shapiro. Computational 3d and reflectivity imaging with high photon efficiency. *Image Processing (ICIP), 2014 IEEE International Conference*, 2014.

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

[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.

[21] A. Torralba and W. T. Freeman. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. *Computer Vision and Pattern Recognition (CVPR). IEEE.*, pages 374–381, 2012.

[22] A Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M.G. 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.

[23] Etienne Vincent and Robert Laganiére. Detecting planar homographies in an image pair. In *Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on*, pages 182–187. IEEE, 2001.

[24] H.Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W.T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *IEEE Signal Processing Letters*, 2012.

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

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

[27] Z. Zhang, P. Isola, and E.H. Adelson. Sparklevision: Seeing the world through random specular microfacets. 2014.

[28] Marco Zuliani, Charles S Kenney, and BS Manjunath. The multiransac algorithm and its application to detect planar homographies. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III–153. IEEE, 2005.