Occluded Imaging with Time-of-Flight Sensors

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

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1145/2836164">http://dx.doi.org/10.1145/2836164</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Association for Computing Machinery (ACM)</td>
</tr>
<tr>
<td>Version</td>
<td>Author’s final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Dec 14 04:23:10 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/110555">http://hdl.handle.net/1721.1/110555</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
1. INTRODUCTION

With the emergence of low cost, yet powerful Time of Flight (ToF) range cameras (such as the new Microsoft Kinect) a question that is raised is whether such off the shelf devices can be used to look around corners and through diffusers.

The looking around corners problem (Figure 6a) is an important area of study. It is a theoretically challenging problem in image formation where the image sensor receives photons that scatter from line-of-sight (LOS) objects, non-line-of-sight (NLOS) objects, and background illumination. Separating these reflections has been a substantial challenge. The current state of the art is the solution from [Velten et al. 2012] where they demonstrate looking around corners in a controlled laboratory setup. Although this result does not use scene priors, it is dependent on ultrafast optical hardware, which comes at a high price tag and significant limitations. Recent work has demonstrated a solution that may drastically lower the cost of such a system [Heide et al. 2014]. However, this solution exploits scene-dependent priors in the reconstruction and uses customized hardware to acquire “transient images”.

Following the problem’s introduction in [Velten et al. 2012], several fundamental questions remain unanswered. These include:

--- How high does the camera modulation frequency need to be? We show that the camera modulation frequency has an approximately linear relationship with desired resolution.
--- How “diffuse” or “shiny” does the wall have to be so we can look around the corner? We show that the width of the specular lobe has a nonlinear relationship to recovery.
--- Do we need only amplitude or both phase and amplitude? In practice, amplitude-only localization is susceptible to noise; we compare the two strategies.
--- Under what conditions is recovery possible? It depends on the physical constraints and computational choices we make.

It turns out that all of these questions can be addressed through a unified forward model that we propose in the paper. We call this model a Virtual Sensor Array (VSA) as it connects ToF range measurements with array signal processing. We recognize that problems such as looking around corners (also denoted as “corners” problem in the following text, for simplicity) are specialized. However, the VSA model generalizes to handle imaging through diffusers as well.

In summary, our key contribution is:

--- A unifying theoretical framework for occluded imaging with time of flight cameras; we use array signal processing to formulate limits on recoverability and add specularity to the formulation.

Secondary technical contributions:

--- Demonstration of occluded imaging in the context of the corners problem.
--- Generalizing the theory to handle imaging through diffusers, and a practical demonstration of this case.

Benefits As compared to prior art [Velten et al. 2012; Heide et al. 2014], our model is rooted in array signal processing theory. In addition, while comparison work exploits customized hardware for “transient imaging”, our hardware uses standard data from a ToF camera (i.e. only phase/amplitude at each pixel). A key benefit—beyond reproducibility and low cost—is real-time acquisition. To our knowledge this paper is the first to provide detailed bounds on recovery for looking around corners. This may be used as a blueprint for future camera designs.

Limitations In this paper, we propose a theoretical framework to understand the problem. Due to the extensive customization of transient imaging hardware, we validate our theory on off-the-shelf devices. This has the benefit of reproducibility, low cost, and real-time potential but a drawback in the perceived quality of results. Specific engineering challenges include a lack of customized illumination and the use of a single modulation frequency that restrict our demonstrations to relatively simple scenes. Nevertheless, to validate the theory, our scenes convey important information, such as the resolution between two objects.

2. RELATED WORK

Holography literature contains a closely related work that overlaps with our theoretical model. Specifically, [Rivenson et al. 2013], provide reconstruction guarantees for compressive holography using discrete spatial sampling to map a sparse set of 3D points to intensity and phase measurements of a 2D surface. Reconstruction guarantees are provided using the Gram matrix of a Fresnel sensing operator. In contrast, our paper is entirely in the realm of ray-based optics and maps intensity and phase of TOF measurements to a surface.
Time profile imaging represents an increasingly popular research area where captured photons are parameterized by both space and time. The “Femtophotometry” technique by [Velten et al. 2013] uses a laboratory grade optical setup to capture visualizations of light in flight, however the technique is expensive (half million dollars). Recently, [Heide et al. 2013] and [Kadambi et al. 2013] have repurposed low cost 3-D, time of flight sensors to achieve some of the capabilities of Velten’s system. However, beyond customized hardware, both of these techniques require the acquisition of a time-frequency shift matrix [Heide et al. 2013; Lin et al. 2014] or a time shift vector [Kadambi et al. 2013], which mitigates the real-time advantage that time of flight sensors usually enjoy. Time profile imaging has several applications, and in particular, [Velten et al. 2012] and [Heide et al. 2014] use such data to address the corners problem. Multipath interference correction is closely related to this paper. While this paper exploits information in scattered light, related techniques in light transport (cf. [O’Toole et al. 2014; Gupta et al. 2014]), or time-frequency analysis [Bhandari et al. 2014] correct for such interference. A light transport analysis of transient imaging can be found in [Velten et al. 2012; Wu et al. 2014], and a comprehensive review of transient technique was collected by [Masia 2014].

Non-line-of-sight target localization is a classic inverse problem that has been studied in a variety of domains. A prominent example is multi-path radar system. For example in [Sen and Nehorai 2011], a Doppler radar system is equipped with spatial diversity (i.e. detectors at different spatial locations) to allow the system to obtain multiple “looks” of a target and resolve NLOS objects in motion. In [Sume et al. 2011], they demonstrate a radar system designed to track sources around a corner. In [Adib et al. 2014], radio waves are used to track humans: however, this technique works for NLOS only when the medium is transparent to radio waves, and makes the limiting assumption that the target is in motion. Optically localizing a source is challenging as diffuse scattering needs to be taken into account. Within computational imaging, various techniques have used indirect reflections to infer various scene properties [Reshetouski and Ihrke 2013; Reshetouski et al. 2011; Naik et al. 2011].

Phased array source localization can be described as follows: given phase and amplitude measurements from multiple sensors located near a source signal, how can we localize the source? The computational technique of choice may depend on whether the source is near or far field. The classic technique is time shifted beamforming which was introduced in [Carter 1981]. Holography methods have also been used for decades to achieve near-field acoustic source localization [Maynard et al. 1985]. In discrete approaches, such as [Malioutov et al. 2005] and [Cevher et al. 2008], a coordinate grid system of N voxels is drawn in the search space. Assuming that K targets are located on grid, and since K ≪ N, by coupling grid source localization with sparse priors, it is possible to resolve very closely spaced sources. More recent work in signal processing often leverages model-based algorithms to enhance recovery [Boufounos et al. 2011; Hegde et al. 2014].

Sparse approximation refers to the problem of estimating sparse vector that satisfies a linear system of equations. Concretely, given a measurement vector y and a dictionary matrix D, the goal is to solve for x where y = Dx and x is known to be sparse. To solve the linear system and enforce sparsity on x, popular solutions include iterative approaches that use an ℓ1 regularization penalty

3. TIME OF FLIGHT LOCALIZATION

We begin by recasting time of flight 3-D imaging into the realm of array signal processing.

Time of Flight imaging We use the term time of flight (ToF) to refer to the time it takes for photons to travel through a medium. Although there are several devices to measure ToF, we restrict subsequent technical discussion to amplitude modulated continuous wave (AMCW) time of flight cameras.

To obtain depth, a light source strobes in a periodic pattern and photons are captured with a lock-in CMOS sensor. The carrier signal is the optical signal and the modulation envelope is the strobing pattern with modulation frequency f_{M}. The phase difference between the received and emitted modulation codes, φ_{M}, encodes the propagation distance via the following linear relation:

\begin{equation}
    z = \frac{c\varphi_M}{2\pi f_M}, \quad c \approx 3 \times 10^8 \text{ m/s}.
\end{equation}

Here, z is the propagation (in meters) of the optical path. A common value of f_{M} is 30 MHz, which corresponds to a λ of 10 meters. The camera also measures the amplitude of the reflected light, denoted as A. In summary, a ToF camera is unique in that the pair of phase and amplitude is measured at each pixel.

Source localization The 2-D source localization problem is as follows. Consider a set of M sensors spaced evenly on a horizontal axis, u. There are K transmitting sources located on the 2-D space parametrized by u and w axis. Denote the signal time delay from k-th source to m-th sensor as σ_{k,m}. Then, in frequency domain, the m-th sensor receives Y_{m}(2\pi f_{M}) =
Triangulation with lensless measurements.

and phase can be expressed as a measurement phasor $A\zeta$ where $c$ is an $M$-dimensional measurement vector defined over the coordinates $(u, v)$ on the wall. The wall serves as a virtual sensor array (VSA). By taking a picture of the virtual sensor we obtain the measurement $M$. The goal of Section 4.1 is to extract $M$ from the ToF camera measurements. (b) Given such measurements, a reasonable next step is to use backprojection to localize the source. This second step connects to lensless transient imaging, as introduced in [Wu et al. 2014].

$$\sum_{k=1}^{K} A_k \exp (-j2\pi f_M \tau_{k,m})$$

Substitution using $z = c\tau_{k,m}$ and Equation 1 yields:

$$Y_{\ell}(2\pi f_M) = \sum_{k=1}^{K} A_k \exp (-j\varphi_{k,m}).$$

The superscript $M$ on $\varphi$ emphasizes that this is the phase associated with modulation frequency $f_M$, i.e., for a fixed $z$, $\varphi^M \neq \varphi^J$ for $\mathcal{I} \neq \mathcal{J}$. As most of the analysis is concerned with narrowband scenarios, we drop the superscript. Therefore, the observation model is written as:

$$\bar{y} \triangleq \left[ Y_1 (2\pi f_M), \ldots, Y_M (2\pi f_M) \right]^T$$

$$= \left[ \sum_{k=1}^{K} A_k \exp (-j\varphi_{k,1}), \ldots, \sum_{k=1}^{K} A_k \exp (-j\varphi_{k,K}) \right]^T,$$ (2)

where $\bar{y}$ is an $M$-dimensional measurement vector defined over the complex field. We summarize the key intuition: each entry of $\bar{y}$ represents the measured amplitude and phase at a single sensor.

4. VIRTUAL SENSOR ARRAY FOR RECONSTRUCTION

Recall that a sensor array is an array of $M$ sensors each measuring phase and amplitude. A virtual sensor array (VSA) probes the idea of turning ordinary surfaces into a sensor array. Consider the toy problem in Figure 2a: a point source emitter is hidden around the corner and the goal is to recover its location and amplitude from VSA measurements. Taken together, Sections 4.1 and 4.2 introduce the virtual sensor array in the context of this toy problem, following which Sections 4.3 and 4.4 generalize the model to broader scenes.

4.1 Virtual Sensor Array

In a time of flight camera, intensities are parameterized both spatially and temporally as

$$c(u, v, t) = A(u, v) \sin (2\pi f_M t + \varphi(u, v)) + \zeta(u, v),$$ (3)

where $c(u, v, t)$ is the correlation waveform with amplitude $A(u, v)$ and phase $\varphi(u, v)$. The quantity $\zeta(u, v)$ is an offset term that represents ambient lighting. Note that $\zeta(u, v), \varphi(u, v),$ and $A(u, v)$ are not parameterized in time — we assume these quantities are constant over a short integration time. Then, the amplitude and phase can be expressed as a measurement phasor

$$\mathcal{M}(u, v) \triangleq A_0 \exp (j\varphi(u, v)),$$ (4)

where the addition of subscript $\mathcal{M}$ to amplitude and phase links the two quantities with the phasor $\mathcal{M}$. Note that the DC offset from Equation 3 is not captured in the phasor notation of Equation 4. This is perfectly fine, as the offset is not useful (it is uncontrolled, ambient light).

Consider the case in Figure 2a, where a single omni-directional point source emits rays of light onto a wall. Localizing the source is trivial when the wall is mirrored, which allows the point source to be observed directly by the camera. This section is concerned with the more general scenario of localizing the point source when the wall is modeled as a Lambertian surface. The key insight is to represent the wall itself as a virtual, lensless imaging sensor in the $(u, v)$ plane. We are interested in obtaining this “lensless image” formed on the virtual sensor.

Using ray optics we begin by analyzing the complex domain light transport of a unit amplitude strobing signal. As illustrated in Figure 2a, the transport phasor from source to wall is represented as $L(u, v, \theta)$, where $u$ and $v$ are the coordinates of the wall that the ray strikes at an angle of $\theta$ to the normal. Therefore, the phasor that models transport from light source to diffuse wall is written as

$$L(u, v) = \frac{\cos \theta}{\varphi_L(u, v)^2} e^{j\varphi_L(u, v)},$$ (5)

Amplitude Decay

Here, note that $L$ is parametrized by only $u$ and $v$ because, assuming the geometry in Figure 2a, the angle $\theta$ is a function of $(u, v)$. The amplitude of the transport phasor is designed to represent an amplitude decay term. A similar transport phasor can be formulated from rays emitted from the wall to the camera. As illustrated in Figure 2a, an outgoing ray makes an angle $\psi$ with the normal vector of the wall. The corresponding phase is $\varphi_C(u, v)$. Therefore, the transport phasor is

$$C(u, v, \psi) = \rho(u, v) \frac{\cos \psi}{\varphi_C(u, v)^2} e^{j\varphi_C(u, v)},$$ (6)

Amplitude Decay

where now $\rho(u, v)$ represents the Lambertian albedo of the wall at coordinates $(u, v)$. Using the two transport phasors as well as the original amplitude of the strobing signal, $A_0$, the combined phasor transport from source to camera is a phasor multiplication:

$$\mathcal{M}(u, v, \psi) = A_0 L(u, v) C(u, v, \psi).$$ (7)

ACM Transactions on Graphics, Vol. , No. , Article , Publication date: .
Surfaces are point reflectors (b) can be thought of as emitters; and (b) over the voxel grid, the blue surfaces can be thought of as multiple point reflectors (green).

Since the camera is perfectly focused on the wall, Equation 7 can be written as

\[
\mathbb{M}(u, v) = A_0 \mathbb{L}(u, v) \int C(u, v, \psi) d\psi,
\]

where the emitting angle has been integrated out. Embedded within Equation 8 is the phasor \( \mathbb{L}(u, v) \), which is the projection of the source onto the wall or the virtual imaging plane. If this phasor could be isolated, then these measurements could be used to formulate a phased array, source localization problem. Because the forward problem is phasor multiplication, it is simple to isolate \( \mathbb{L}(u, v) \) as

\[
\mathbb{L}(u, v) = \left( \frac{1}{A_0} \right) \left( \mathbb{M}(u, v) \right) \mathbb{C}(u, v, \psi).
\]

Although \( \mathbb{C}(u, v) \) is an unknown, simply computing a depth map of the wall provides the phase \( \varphi_C \). Using just \( \varphi_C \) and unit amplitude as a proxy for \( \mathbb{C}(u, v) \) we obtain

\[
\hat{\mathbb{L}}(u, v) = \left( \frac{1}{A_0} \right) \left( \mathbb{M}(u, v) \right) e^{j \varphi_C(u, v)}
\]

\[
= \left( \frac{\cos \theta}{\mathbb{L}(u, v)} \right) \int \rho(u, v) \frac{\cos \psi}{\varphi_C(u, v)} d\psi, e^{j \varphi(u, v)}
\]

\[
= A_0(u, v) e^{j \varphi(u, v)}.
\]

Here, \( \hat{\mathbb{L}}(u, v) \) is an estimate of \( \mathbb{L}(u, v) \) with the correct phase and different amplitude. In order to treat the amplitude as a uniform difference in scaling (across all array elements), we assume that the reflectance profile is uniform, i.e., \( \rho(u, v) \) is the same for all \( u \) and \( v \). In summary, the key measurements are \( \hat{\varphi}_L(u, v) \) and \( \hat{\varphi}_A(u, v) \). These represent the projection of phases and amplitudes of the light source onto the virtual sensor which can be used in the context of source localization.

4.2 Reconstruction

The problem has now been abstracted to 3-D source localization with a 2-D array of sensors, parameterized by \((u, v)\). Each virtual sensor element gives phase and amplitude measurement \( \hat{\varphi}_L(u, v) \) and \( \hat{\varphi}_A(u, v) \). The target is a point source whose real-world location is parameterized in 3-D spatial coordinates \((X, Y, Z)\). Our goal is to find the points \((\hat{u}, \hat{v}, \hat{w})\) which correspond to the target coordinates with respect to the wall and camera sensor.

Without loss of generality we will consider 2-D source localization using a 1-D slice of measurements, i.e., to obtain \((\hat{u}, \hat{v})\) by using sensor measurements only along the horizontal \( u \)-axis. The measurement vector is of the form

\[
\hat{y} = \left[ \hat{\mathbb{L}}(u_1) \hat{\mathbb{L}}(u_2) \cdots \hat{\mathbb{L}}(u_M) \right]^T.
\]
Emitters to reflectors  Recall that the goal of the toy problem in Figure 2a was to localize an active light source around the corner. Of course this is not a realistic scenario: it is unlikely that the hidden object is an active light source. Figure 3a illustrates a more common scenario of localizing point reflector(s) are around the corner. To solve this problem the position of the light source must be known a priori. Then the path to each possible voxel location is known and a valid dictionary for the space can be constructed. We must mention that the reflectors do not have to be Lambertian. More precisely, the directivity of the object is equivalent to having a directional array, which actually facilitates recovery (see Section 5).

On-grid to off-grid  To this point we have assumed that point targets, e.g., a point reflector or point source, lies on a search voxel. In realistic scenarios points are not guaranteed to lie on-grid and “off-grid” localization must be performed. Fortunately, the source localization community has developed powerful tools to address this very scenario. Approaches are in the style of iterative multiresolution methods which upsamples the dictionary [Maliioutov et al. 2005] or Continuous Basis Pursuit which interpolates the dictionary [Ekanadham et al. 2011].

Points to surfaces  As illustrated in Figure 3b, the object of interest is usually a continuous surface. In this case, the object can be modelled as many closely spaced point emitters. In this case, the recovered surface would be a convolution of the surface with the beampattern of the single point source. This would allow us to recover a blurred version of the occluded surface, where the degree of blur depends on the width of the beampattern of a single point source.

Illumination position  The goal is to build a camera that can look around corners without any gadgets in the line-of-sight. Therefore it is desirable to have the source on the same side as the camera. Achieving this turns out to be an engineering challenge as opposed to a theoretical one. Following from Figure 3a, as long as the position of the light source is known—whether it is next to the camera or not—a dictionary can be constructed for the space. The engineering challenges are:

—Saturation from an area source. ToF cameras are designed to illuminate an area, and thus if an area source is aimed at a wall the direct reflections will saturate the sensor.
—Very little light comes back: light has to bounce off the wall twice and the object once before returning to the camera.

A solution to the first problem is to use a collimated beam, either from a laser or by blocking the area source. A solution to the latter problem is to use a more powerful light source than what is stock on time of flight cameras. The simulation shown in Figure 4 verifies that the VSA model holds when the illumination is on the same side as the occluder. Refer to the figure caption for details.

Shininess of the wall and reflectors  An interesting link between computer graphics and array signal processing exists between the bidirectional reflectance distribution function (BRDF) and antenna directivity. In the signal processing community, when sensors show directional preference (as opposed to being omnidirectional), resolving targets within the aperture becomes much easier. Indeed, in the corners problem the directionality of the reflectance is critical. Consider two opposing cases: (i) the wall is a mirror and thus the BRDF has strong directional preference, and (ii) the wall is purely diffuse corresponding to a constant BRDF. This confirms the physical intuition where we expect high resolvability of targets with specular BRDFs (e.g. mirrors) and low resolvability with diffuse BRDFs. We must also mention that the directionality of the VSA (determined by the BRDF of the wall) is dual to the directivity of the reflectors. Practitioners should note that to generate D, an estimate of the BRDF should be obtained for optimal results. Sections 5 and 6 probe further into the reflectivity of the wall (i.e. shininess).

4.4 Imaging through Diffusers  To this point we have proposed the VSA model and shown its utility for looking around corners. We now show that it can also be used to image through diffuse media. Specifically, we consider a transmissive toy problem: localization of a source through a diffuser, shown in Figure 5. Here, the key idea is that the scattered paths are deflected and thus have a fraction longer time duration to some of the sensors.

Note that the problem is almost identical to the looking around corners problem in Figure 2a. In the corners problem, the virtual
sensor array was the wall itself. In the diffuser problem the virtual sensor array is the visible surface of the diffuser. Similar parallels exist for light transport analysis. Since the corners problem is using reflected measurements, the BRDF controls the directionality of sensors. In contrast, since the diffuser problem uses transmission measurements, sensor directionality is determined by the bidirectional transmittance distribution function (BTDF). Recall that the trivial case for the corners problem was a mirror, which has a peaked BRDF. The analogous trivial case for the diffuser problems occurs for a clear object, which has a peaked BTDF. We can therefore address the corresponding questions for the transmissive system, such as “how opaque or clear does the diffuser have to be”?

In crux, Equations 1 to 18 all apply to the case of imaging through diffusers.

5. ANALYSIS OF RECOVERABILITY

In this section we provide numerical guarantees to quantify when the VSA model can recover the occluded image and when it can’t. To provide guarantees, parallel work in holography [Rivenson et al. 2013] proposes the use of mutual coherence as a key metric. Using our notation, we write the mutual coherence \( \mu(D) \) as:

\[
\mu(D) = \max_{i \neq j} |G_{ij}|, \quad G = |D^H D|,
\]

where \( |D_n| \) is 1 for \( n = 1, \ldots, N \). Intuitively, the mutual coherence encodes the similarity between the columns of \( D \). For robust recovery, it is important to reduce the coherence, which is achieved through the choice of physical parameters. In the following paragraph we show that the specularity of the wall has an inverse relationship to the FWHM, and since all our functions are Gaussian, the relationship holds for mutual coherence, providing us a bound on target resolution. For more details on using mutual coherence to provide reconstruction guarantees we refer to the reader to [Rivenson et al. 2013] and our supplementary material.

**Recovery guarantees for specular surfaces:** When the wall is non-Lambertian, the virtual sensors are no longer omnidirectional. To model the directionality of virtual sensors, we first define the **beam pattern** as a row of the Gramian matrix \( G \).\(^1\) Like in optics, the FWHM of the beam pattern specifies how far apart two targets must be to resolve both of them. As one can imagine, this has been explored in array signal processing. For instance, in [Van Trees 2004] a derivation is provided for the FWHM of omnidirectional sensors, which takes the form of the familiar Rayleigh limit:

\[
\text{FWHM}^\ell = \arcsin \left( \frac{\lambda}{D} \right),
\]

where \( D \) is the diameter of the sensor array (in meters) and FWHM\(^\ell \) is the angular resolution (in radians). This equation provides the resolution to which targets can be resolved. A 300 MHz camera has a \( \lambda \) of approximately 1 meter and typically our virtual sensor array is about \( D = 1 \) meter wide. Then, for omnidirectional sensors, the resolution to which one can distinguish targets is approximately 1 meter, which is poor. Fortunately, if the virtual sensors were directional (e.g. if the wall is shiny), then the resolution limit improves.

We now derive the FWHM for a directional sensor system (such as a specular wall). Let \( \gamma^\ell \) denote the FWHM of the directional response function of an individual virtual sensor, with units in radians. In the corners problem, the directional response function is the BRDF of the wall. Therefore, we use simplifications from computer graphics [Ramamoorthi and Hanrahan 2001; Han et al. 2007] to approximate the specular lobe of the BRDF by a Gaussian. From Chapter 3, of [Van Trees 2004], the FWHM for a directional sensor system is a composite of the omnidirectional FWHM (Equation 20) with the FWHM of the individual sensor response (\( \gamma^\ell \)). Following this recipe, we obtain the FWHM for the directional system:

\[
\text{FWHM}^\ell = \arcsin \left( \left( \frac{\lambda \gamma^\ell}{\lambda + D \gamma^\ell} \right) \right)\quad (21)
\]

As a sanity check, it can be verified that if the sensor is omnidirectional, then Equation 21 simplifies to Equation 20.\(^2\) One can also verify that a low value of \( \gamma^\ell \), i.e., a specular BRDF, corresponds to a narrower FWHM for the system. We arrived at Equation 21 using angular quantities for FWHM, but in this paper, we are also interested in spatial resolution of two targets (i.e. how many meters apart do they have to be). The relation between angular resolution and spatial resolution is written as

\[
\text{FWHM}^d = d \text{FWHM}^\ell,\quad (22)
\]

where \( d \) is the depth of the object from the array (in meters). We use the superscripts to denote the units of a scalar variable; \( \ell \) for length (meters), \( \ell \) for angular quantities in radians, and \( ^\ell \) to denote angular quantities in degrees.

**General recovery guarantees based on rank and span constraints:** General guarantees can be obtained using rank and span constraints. For example, \( D \) needs to have sufficient linear independence to uniquely recover sources. This can be expressed as a **rank-constraint**, where \( \text{rank}(D) = 2 \) encodes an upper bound on the dimensionality of the convex hull of targets (cf. pages 81-97 of [Gower 1985]). A complementary, **span constraint** characterizes appropriate arrangements of virtual sensors (e.g., the wall geometry) that avoid degenerate solutions. Using the general frameworks of rank, span, and mutual coherence, the supplemental material explores how other parameters—beyond specularity—influence reconstruction. This includes gridding, modulation frequency, aperture size, non-planar walls, and choice of reconstruction algorithm.

6. RESULTS

For all experiments, the time of flight camera used is the Mesa Swissranger SR4050 lock-in module. It can be purchased directly from MESA Imaging.\(^3\) in Zurich, Switzerland. This time of flight camera has a decoupled light source and operates at a modulation frequency of 30 MHz.

---

\(^1\)Equivalently, beam pattern is a column of \( G \) since \( G \) is Hermitian.

\(^2\)Specifically, \( \lim_{\gamma^\ell \rightarrow \infty} \frac{\lambda \gamma^\ell}{\lambda + D \gamma^\ell} = \frac{\lambda}{D} \)

\(^3\)http://www.mesa-imaging.ch/
6.1 Qualitative Results

Real-time occluded imaging  As illustrated in Figure 6, a single point reflector is occluded from the camera’s line-of-sight. The reflector is placed in motion and can be localized in real-time using the backprojection algorithm. The supplement includes a video of this demonstration.

Imaging around the corner  In Figure 7 we replace the moving ping pong ball from the real-time result with a small, “T” shaped object. By using pseudoinverse beamforming, we are able to recover the hidden image. The height represents confidence in a given voxel.

4The size of the object is 20 by 20 centimeters
paper. To keep the framework scene-independent, priors are not placed on the reconstruction—implementing total variation or edge constraints would improve the reconstruction for some scenes.

**Imaging through scattering media** In Section 4.4, the VSA model has been analytically shown to generalize to the diffuser problem. Here, an experimental result is demonstrated in Figure 8 for imaging through scattering media. We place a tank of milky water in between the camera and two ping pong balls. For this problem the BTDF is the dual of BRDF, which allows us to use the same computation that was used in the corner experiments.

First we add relatively little milk and show that it is possible to localize the ball (Figure 8a). Then we add a much greater quantity of milk and show it is still possible to localize the ball (Figure 8b). Of course, in the latter case the coherence is much greater, and therefore the two peaks are not distinctly separated. To distinctly separate the peaks, a scene-dependent prior, such as sparsity can be used (Figure 8b). Imaging through scattering media is such a well established area that we must emphasize that our results are very preliminary results that need to be compared against other methods (e.g. structured light, phase conjugation, etc.). However, these experiments are sufficient to show the generality of the VSA model.

### 6.2 Quantitative Assessment

In this section we will perform real and simulated experiments to quantify the conditions for successful recovery.

#### 6.2.1 Quantitative Physical Experiments

**Directionality of the virtual sensor array** Creating a camera that can look around corners requires an understanding of the material properties of the virtual sensor array. To form the wall for the corners problem we collect four materials in increasing order of specularity: (i) posterboard; (ii) photo paper; (iii) metal; and (iv) a mirror. Figure 11 illustrates the measured directionality of the first three materials (we assume the mirror has a delta function for specularity).

In Table II, we list quantitative reflectance parameters, where \( \rho_s \) and \( \alpha \) measure the specular intensity and surface roughness, as defined in the Ward BRDF model [Ward 1992].

We draw specific conclusions from our empirical study, the resolution could improve by an order of magnitude using more sophisticated solvers.

**Resolving multiple point sources** Our end goal is to image around the corner; therefore, a critical performance metric is how close two point reflectors can be localized (without relying on sparsity assumptions). As this is a material dependent property, Figure 12 illustrates localization of two point sources for different materials. For the posterboard (Figure 12a) using pseudoinverse back-projection we are able to localize the two point sources that are 10 centimeters apart. We then plot the beampattern, which is one row of the Gram matrix \( \mathbf{G} \). Note how, because the beampattern is very wide, the Gram matrix is very coherent. Figures 12b, 12c and 12d show results for the photo paper, metal, and mirrored objects. Observe that the beampattern narrows as the material changes from the posterboard to the mirror. In particular, for the mirror, the beampattern is a Dirac and the mutual coherence reaches the minimum value of 0.

Define the minimum resolvable distance as the minimum separation between two ping pong balls that can be detected. In our experimental results we found that for the posterboard it was 10 cm, for the photo paper 3 cm, and for the metal 2 cm.

A practical implication of this result is that even when using the relatively diffuse posterboard it is possible to obtain an image of the object around the corner (if the objects are large enough). For a fixed reflectance, using our theory, it is clear that increasing the modulation frequency scales linearly with expected resolution.

**What is angular response like in the wild?** If the virtual sensor array was omnidirectional, recovery of \( \hat{\mathbf{x}} \) is challenging and—in the context of phased array processing—not possible with today’s time of flight cameras. If the Lambertian assumption can be relaxed then the technique would be more readily applicable today. We use the Mitsubishi Electric Research Labs (MERL) BRDF database [Matusik et al. 2003] to evaluate the directionality of various real materials. We use the fitted Ward parameters in [Ngan et al. 2005] to calculate FWHM and FWHM‘ in the same manner as Table II. The results of all 100 BRDFs are shown in Figure 13. From this we can conclude that:

- Highly specular materials (like metals) have a very small FWHM and thus high localization (about 2 cm).
- Materials with a dominant specular lobe (relative to the diffuse lobe) have a reasonably small FWHM and acceptable localization (about 5 cm).

It is interesting to note that 38 of the 100 materials in the database could produce a beamforming resolution of smaller than 10 cm and 29 materials can make it smaller than 5 cm. We can therefore expect beamforming to be feasible for many real-world materials, and, from our empirical study, the resolution could improve by an order of magnitude using more sophisticated solvers.

**Superresolution via sparsity** A great deal of research in signal processing centers around techniques for solving linear inverse problems. Here, we probe this idea by using a simulated array where the expected FWHM of the beampattern is approximately 2 meters. Since simple beamforming cannot resolve targets that are spaced closer than 2 meters apart, the interesting question that follows, is whether using more sophisticated solvers will allow for better resolution.

---

5 The Ward lobe is rendered with the BRDF Explorer (http://www.disneyanimation.com/technology/brdf.html).

6 The directionality of the sensor is determined by the BRDF of the surface. Details on measuring and fitting the directionality are in the supplement.
Fig. 9: Sparse priors allow superresolution of reflectors, and hence, potentially higher resolution images of the occluded scene. Ground truth reflectors are at 20cm and 80cm, which is closer than the Rayleigh limit for this scene (Equation 21). Here, Basis Pursuit Denoising (the convex relaxation) is the only technique that resolves target positions even though they are within the Rayleigh limit. Note the varying orders of magnitude of the y-axis due to sensitivity of some of these techniques to additive noise (e.g., pseudoinverse) or coherence (e.g. CoSaMP).

Table I. : Different solvers for reconstruction and their objective functions.

<table>
<thead>
<tr>
<th>Backprojection Solvers</th>
<th>Objective Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic Beamforming</td>
<td>$\hat{x} = D^{H} \hat{y}$</td>
</tr>
<tr>
<td>Pseudoinverse</td>
<td>$\hat{x} = \arg\min_{x} |D\hat{x} - \hat{y}|_2^2$</td>
</tr>
<tr>
<td>CoSaMP</td>
<td>$\hat{x} = \arg\min_{x} |x|_0$ s.t. $D\hat{x} = \hat{y}$</td>
</tr>
<tr>
<td>Basis Pursuit Denoising</td>
<td>$\hat{x} = \arg\min_{x} |x|_1$ s.t. $|D\hat{x} - \hat{y}|_2^2 \leq \epsilon$</td>
</tr>
</tbody>
</table>

Figure 9 compares the four techniques shown in Table I. Both CoSaMP and Basis Pursuit Denoising are solvers that enforce sparsity on $\hat{x}$. However, the former is a greedy algorithm while the latter is a convex relaxation. For a well-defined optimization program, a convex relaxation is guaranteed to find the correct solution. No such guarantees exist for greedy methods, and therefore the latter are considered superior for recovery.\(^7\) Our results are consistent with this intuition. As illustrated in Figure 9 we observe that both beamforming and the pseudoinverse are unable to resolve the targets, CoSaMP converges to a poor solution, and Basis Pursuit Denoising is able to superresolve the targets.

7. DISCUSSION

Do we approach the bounds? A natural question is whether the bounds we have proposed are meaningful in practice. We will restrict ourselves to the bound provided in Equation 21 as this is the most general bound: it is invariant to any model assumptions on $D$ or $\hat{x}$. Another way to describe Equation 21 is that it provides a lower bound for the width of the beampattern. In the last column of Table II we list the computed bound based upon the acquired directionality of the materials. The width of the experimental beampatterns collected for different materials (Figure 12) approaches, but does not violate the bound. Any discrepancy from the bound is due to experimental error either in the measurements of $\gamma^c$ used to calculate the bound or in obtaining the beampattern.\(^8\) In practice, the bound is perhaps useful when comparing materials with distinct properties. For example, the slack in the bound for the photo paper is about 2 cm, but the difference in beampattern width between the photo paper and posterboard is about 60 cm.

We must also mention that the bound provided in Equation 21 guarantees success when using the most basic solver (i.e. $\hat{x} = D^{H} \hat{y}$). Empirically we observe that we can often obtain a resolution close to an order of magnitude better using a stronger solver, such as the pseudoinverse. For example, in Figure 12a although the beampattern is approximately 1 meter wide, we are able to resolve targets 10 cm apart. There are many ways to derive sharper guar-

\(^7\)Greedy methods are not without merit; they are simpler and well suited for model-based approaches.

\(^8\)Errors inherent to ToF 3-D depth estimation can disrupt calibration between the VSA and camera.
Fig. 11: Rendered directionality for three different walls from Figure 12 along with 1-D plots of the directional response in both degrees (black) and centimeters (red). Only the specular lobe is shown (the diffuse lobe is omitted). Quantitative parameters derived from these plots can be found in Table II.

Fig. 12: How close can we resolve two point sources around the corner without using any prior assumptions? Consistent with the text, it depends on the material properties of the wall. Across each row the images represent: (i) a photograph of the wall, (ii) pseudoinverse backprojection \((D^\dagger\hat{y})\), (iii) the beampattern, and (iv) the matrix \(G\). Note that as expected, for specular objects, \(G\) is sharply diagonal and algebraically incoherent, while for more diffuse objects the Gram matrix is more coherent.
Occluded Imaging with Time of Flight Sensors

...are able to show a real-time scene capture and (ii) we are able to provide bounds on recovery.

Future work This paper has proposed a unifying framework for occluded imaging via off the shelf, unmodified time of flight cameras. The next step is to apply the theory to a heavily customizedToF camera.

Although we have made the connection between concepts such as time of flight and scattering, this paper only shows that the virtual sensor array model is valid for localization through scattering. Future work would go into depth into what is a classic problem in optics and would also provide comparisons to known techniques, such as structured light and phase conjugation.

Conclusion In summary, we reveal the link between looking around corners and phased array processing. We have proposed the first model for occluded imaging with standard time of flight cameras. Such cameras are increasing in popularity, both for their applications to 3-D imaging as well as novelty in computational photography applications.

Looking around the corner is a complex problem with many variables. We conclude that today's time of flight cameras are able to image around the corner with low spatial resolution by exploiting the property that standard walls are not purely Lambertian. We hope that this paper is a step toward having commodity cameras that can look around corners.

Acknowledgments

We thank the anonymous reviewers for their detailed comments, and for pointing out the mathematical connection to compressive holography. We thank Ayush Bhandari for technical comments. Aachita Kadambi was funded by the Charles S. Draper Doctoral Fellowship and Boxin Shi was partially supported the Singapore MOE Academic Research Fund MOE2013-T2-1-159 and the...
REFERENCES


Elad, M. 2010. Sparse and redundant representations: from theory to applications in signal and image processing. Springer.


