2019-09-09

Occlusion-based computational periscopy with consumer cameras

Occlusion-based computational periscopy with consumer cameras

John Murray-Bruce, Charles Saunders, and Vivek K Goyal

Department of Electrical and Computer Engineering, Boston University
Boston, MA 02215 USA

ABSTRACT

The ability to form images of scenes hidden from direct view would be advantageous in many applications – from improved motion planning and collision avoidance in autonomous navigation to enhanced danger anticipation for first-responders in search-and-rescue missions. Recent techniques for imaging around corners have mostly relied on time-of-flight measurements of light propagation, necessitating the use of expensive, specialized optical systems. In this work, we demonstrate how to form images of hidden scenes from intensity-only measurements of the light reaching a visible surface from the hidden scene. Our approach exploits the penumbra cast by an opaque occluding object onto a visible surface. Specifically, we present a physical model that relates the measured photograph to the radiosity of the hidden scene and the visibility function due to the opaque occluder. For a given scene–occluder setup, we characterize the parts of the hidden region for which the physical model is well-conditioned for inversion – i.e., the *computational field of view* (CFOV) of the imaging system. This concept of CFOV is further verified through the Cramér–Rao bound of the hidden-scene estimation problem. Finally, we present a two-step computational method for recovering the occluder and the scene behind it. We demonstrate the effectiveness of the proposed method using both synthetic and experimentally measured data.

Keywords: Computational periscopy, non-line-of-sight imaging, computational photography, computer vision, remote sensing

1. INTRODUCTION

The goal of non-line-of-sight (NLOS) imaging systems is to form images of scenes hidden from direct view. In general, this has been achieved using measurements of properties of the light scattered onto visible surfaces from the hidden scene. For example, these methods have relied on measurements of time-of-flight,[1–17] coherence,[18–20] or even intensity-only[21–26] information. NLOS imaging using acoustic[27] and long-wave infrared[28] waves have also been recent lines of work.

Most active NLOS imaging methods are based on locating hidden reflective surfaces through multi-lateration from measurements of transient light transport, as first introduced by Kirmani *et al.*[1] Velten *et al.*[3] used a femtosecond laser and a streak camera with 2 picosecond resolution to realize a 3D NLOS imaging system; a comparison of several reconstruction algorithms for this setup was provided by Gupta *et al.*[4] Comparatively cheaper transient imaging-based NLOS systems have been enabled by the use of homodyne time-of-flight sensors[5,7,11] or single-photon avalanche diode (SPAD) detectors with time-correlated single photon counting (TC-SPC) modules.[6,9,10,12–14,29,30] Additionally, NLOS imaging has been achieved with inexpensive continuous-wave amplitude modulated cameras (e.g., Kinect).[7,11] Using SPAD-based systems, NLOS imaging beyond roughly one-meter round-trip distances for object estimation[8] and around fifty meters for human localization[12] have been achieved. O'Toole *et al.*[14] demonstrated how confocal scanning yields a deconvolution problem, which is efficiently solved with reduced memory usage and computational complexity. More recently, inverse rendering techniques from computer graphics[31] and a theory of Fermat paths[32] have been used to recover the hidden scene. These methods often demand the use of powerful and complex equipment, have higher data acquisition times, and may be less stealthy than passive alternatives.

Further author information (E-mail): J.M.-B. (johnmb@bu.edu), C.P.S. (cs13@bu.edu), V.K.G. (goyal@bu.edu).
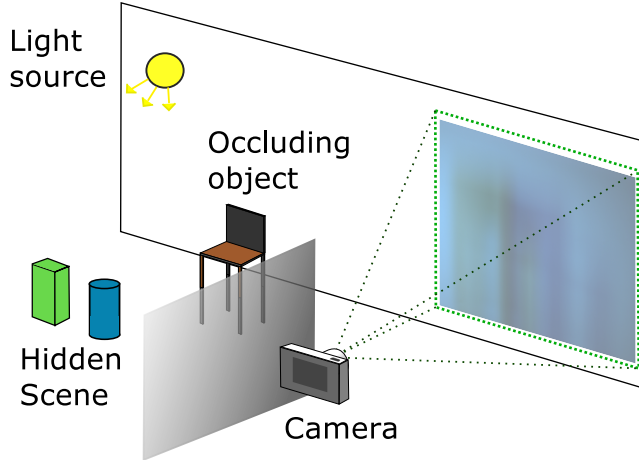
Figure 1: Experimental setup for computational periscopy.

Instead of transmitting radiation into the hidden scene, passive NLOS imaging methods exploit light illumination already in the hidden scene. These methods benefit from occluding structures separating the light paths reaching a visible surface from the hidden scene.[22, 23, 26, 33–35] Vertical edges were first exploited as occluders to reconstruct a video of 1D projections of the moving parts of the hidden scene, from a video of the floor on the visible side.[22] This approach was extended by Seidel *et al.*[35] for stationary scenes and arbitrary floor patterns, using only a single photograph. Baradad *et al.*[23] exploited the presence of visible, precalibrated, complex occluding structures in their recovery. In a subsequent work, the hidden scene and complex occluder structure were recovered from a video sequence.[26]

Our work, which first appeared earlier this year,[24] recovers the position of the occluder (assumed to have a known shape) and a 2D color photograph of the hidden scene from a single photograph of a visible wall. In this paper, we give an overview of that approach for NLOS imaging along with new results that relate the computational field of view (CFOV) concept to the Cramér–Rao bound (CRB), new simulations, and new experimental results. We first present a physical model for the light transport in this scenario (Section 2). Given the non-linear model, which describes the interaction between light from the hidden scene and the unknown occluder, we derive the CRB for scene-only recovery as a proxy for studying the conditioning of the NLOS problem (Section 3); this supports the concept of CFOV[24] (Figures 3 and 4), which is a subset of the hidden scene that is well-conditioned for recovery. In Section 4, a computational method for estimating the occluder and scene is presented. Finally, this proposed approach is verified on both synthetic and real, experimentally measured data (Section 5).

2. FORWARD MODEL AND INVERSE PROBLEM FORMULATION

Consider an NLOS imaging scenario shown in Figure 1, where the visible portion of a *white imaging wall* is illuminated by both ambient illumination and light from the hidden scene after a non-linear interaction with the opaque, non-light-emitting *occluder*. This interaction generates a penumbra – regions of partial shadows – on the visible surface. In this work, we demonstrate how this non-linear interaction between the hidden scene of interest and the occluding object can be exploited to simultaneously estimate the hidden occluder and scene of interest, from intensity-only measurements of the light on the visible surface.

To this end, we begin with a forward model of light propagation for describing the intensity distribution of light reaching the visible wall from the hidden scene and ambient illumination. Let $f(\mathbf{x})$ denote the hidden-scene radiosity at the position $\mathbf{x} \in \mathbb{R}^3$, and let $\boldsymbol{\theta}_{\mathrm{occ}}$ be a suitable parameterization[*] of the occluding object. Then the

---

[*]If the occluding object is a thin opaque square card, for instance, a possible parameterization would be the dimension $r \in \mathbb{R}$ and position $\boldsymbol{p}_{\mathrm{o}} \in \mathbb{R}^3$ of the occluder in 3D space; i.e., $\boldsymbol{\theta}_{\mathrm{occ}} = (r, \boldsymbol{p}_{\mathrm{o}}) \in \mathbb{R}^4$. In this work, for computational simplicity, we will focus only on recovering $\boldsymbol{p}_{\mathrm{o}}$ and assume that the shape (and dimensions) of the occluder is known *a priori*.

irradiance $i(\boldsymbol{p}_{\mathrm{w}})$ at a point $\boldsymbol{p}_{\mathrm{w}}$ is given by

$$i(\boldsymbol{p}_{\mathrm{w}}) = \int_{\mathbf{x} \in \mathcal{S}} \frac{g(\mathbf{x}, \boldsymbol{p}_{\mathrm{w}})}{\|\mathbf{x} - \boldsymbol{p}_{\mathrm{w}}\|_2^2} v(\mathbf{x}, \boldsymbol{p}_{\mathrm{w}}, \boldsymbol{\theta}_{\mathrm{occ}}) f(\mathbf{x}) \, \mathrm{d}\mathbf{x} + b(\boldsymbol{p}_{\mathrm{w}}), \tag{1}$$

where $g(\mathbf{x}, \boldsymbol{p}_{\mathrm{w}}) \stackrel{\text{def}}{=} \cos\left(\measuredangle(\boldsymbol{p}_{\mathrm{w}} - \mathbf{x}, \mathbf{n}_{\mathbf{x}})\right) \cos\left(\measuredangle(\mathbf{x} - \boldsymbol{p}_{\mathrm{w}}, \mathbf{n}_{\boldsymbol{p}_{\mathrm{w}}})\right)$ is the Lambertian bidirectional reflectance distribution function (BRDF), with $\measuredangle(\cdot, \cdot)$ denoting the angle between its pair of vector arguments and $\|\cdot\|_2$ representing the Euclidean norm of a vector. $\mathcal{S} \subset \mathbb{R}^3$ denotes the portion of the hidden region that is of interest; in this work, we assume that we intend to form an image of a rectangular portion of a plane in the hidden scene at a distance D from the visible wall, so that $\mathcal{S} = \{(x, D, z) : (x, z) \in [L_1, L_2]^2\}$. Furthermore, $v(\mathbf{x}, \boldsymbol{p}_{\mathrm{w}}, \boldsymbol{\theta}_{\mathrm{occ}})$, is a Boolean-valued visibility function that equals 1 when the path from $\mathbf{x}$ to $\boldsymbol{p}_{\mathrm{w}}$ is unoccluded by the opaque object parameterized by $\boldsymbol{\theta}_{\mathrm{occ}}$ and 0 otherwise. The final term $b(\boldsymbol{p}_{\mathrm{w}})$ represents the light reaching the visible wall from all sources outside the desired hidden scene region $\mathcal{S}$. Equation (1) is the rendering equation from computer graphics adapted to our setting.[36] From a single photograph of the irradiance distribution $i(\boldsymbol{p}_{\mathrm{w}})$ on the visible wall obtained using an ordinary digital camera, we wish to simultaneously recover the occluder parameterization $\boldsymbol{\theta}_{\mathrm{occ}}$ and form a 2D photograph of the hidden scene $f(\mathbf{x})$. Thus, in the rest of this section, we show how to discretize the linear part of (1) to obtain an affine system that relates the hidden scene and ambient light contributions to the discrete camera measurements. We then discuss the conditioning of the resulting inverse problem in Section 3.

2.1 Discretization

Assume that the visible wall is a white planar surface with diffuse Lambertian reflectivity, and the digital photograph of the visible wall obtained by the camera is a discretization of (1) with $\boldsymbol{p}_{\mathrm{w}}$ evaluated at a number of discrete points – corresponding to the resolution of the camera – on the visible surface within the camera's field of view (FOV).

Let the hidden scene plane of interest, $\mathcal{S} = \{(x, D, z) : (x, z) \in [L_1, L_2] \times [L_1, L_2]\}$, be divided into N scene pixels, $\{\mathcal{S}_n\}_{n=1}^N$, of equal area a_{scene}, and define entry n of the discretized scene $\mathbf{f}$ as the mean of $f(\mathbf{x})$ for $\mathbf{x}$ in scene pixel $\mathcal{S}_n$. Mathematically,

$$f_n = \frac{1}{a_{\mathrm{scene}}} \int_{\boldsymbol{x} \in \mathcal{S}_n} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \tag{2}$$

and $\mathbf{f} = [f_1, f_2, \ldots, f_N]^{\mathsf{T}}$. Thus, $f(\mathbf{x})$ is approximated by a piecewise-constant function of the form

$$f(\mathbf{x}) \approx \sum_{n=1}^N f_n \mathbf{1}_{\mathcal{S}_n}(\mathbf{x}), \tag{3}$$

where

$$\mathbf{1}_{\mathcal{S}_n}(\mathbf{x}) = \begin{cases} 1, & \text{for } \mathbf{x} \in \mathcal{S}_n \\ 0, & \text{for } \mathbf{x} \notin \mathcal{S}_n \end{cases}$$

denotes the indicator function.

Remark 1. The value of N denotes the resolution of the image we attempt to compute of the hidden scene plane. Generally, a higher value for N corresponds to higher resolution image, because the piecewise constant model for the hidden scene (3) becomes more accurate. This, however, comes at a cost: The CRB computed for estimating an N-pixel image of the hidden scene increases with the resolution N (Section 3.1).

Substituting (3) into (1) gives

$$i(\boldsymbol{p}_{\mathrm{w}}) = \sum_{n=1}^N i_n(\boldsymbol{p}_{\mathrm{w}}) + b(\boldsymbol{p}_{\mathrm{w}}), \tag{4}$$

where

$$i_n(\boldsymbol{p}_{\mathrm{w}}) = f_n \int_{\mathbf{x} \in \mathcal{S}_n} \frac{g(\mathbf{x}, \boldsymbol{p}_{\mathrm{w}})}{\|\mathbf{x} - \boldsymbol{p}_{\mathrm{w}}\|_2^2} v(\mathbf{x}, \boldsymbol{p}_{\mathrm{w}}, \boldsymbol{\theta}_{\mathrm{occ}}) \, \mathrm{d}\mathbf{x} \tag{5}$$

is the irradiance at visible wall location $\boldsymbol{p}_\mathrm{w}$ due to scene patch $\mathcal{S}_n$. Similarly, we divide the camera FOV into M camera FOV pixels[†] of equal area a_camera, and we treat $i_n(\boldsymbol{p}_\mathrm{w})$ and $b(\boldsymbol{p}_\mathrm{w})$ as being approximately constant over any camera FOV pixel m. Ignoring camera imperfections, the measurement of camera pixel m is proportional to $i(\boldsymbol{p}_\mathrm{w})$:

$$y_m \propto a_\mathrm{camera} \left[\sum_{n=1}^{N} f_n r_n(\boldsymbol{p}_\mathrm{w}) + b(\boldsymbol{p}_\mathrm{w}) \right],$$

where

$$r_n(\boldsymbol{p}_\mathrm{w}) \stackrel{\mathrm{def}}{=} \int_{\mathbf{x} \in S_n} \frac{g(\mathbf{x}, \boldsymbol{p}_\mathrm{w})}{\|\mathbf{x} - \boldsymbol{p}_\mathrm{w}\|_2^2} v(\mathbf{x}, \boldsymbol{p}_\mathrm{w}, \boldsymbol{\theta}_\mathrm{occ}) \, \mathrm{d}\mathbf{x}, \tag{6}$$

$\boldsymbol{p}_\mathrm{w}$ is any representative of camera FOV pixel m, and the proportionality includes camera efficiency parameters and exposure time. Omitting the a_camera factor and proportionality constant by absorbing them with other scalings within the camera that lead to dimensionless (camera) pixel values, we have that

$$y_m \propto \left(\sum_{n=1}^{N} f_n r_n(\boldsymbol{p}_\mathrm{w}) \right) + b(\boldsymbol{p}_\mathrm{w}).$$

With access to M measurements $\mathbf{y} = [y_1, y_2, \ldots, y_M]^\mathsf{T}$ of the irradiance distribution at points $\{\boldsymbol{p}_{\mathrm{w},m}\}$ on the visible wall, writing $\mathbf{b} = [b_1, b_2, \ldots, b_M]^\mathsf{T}$ gives:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} r_1(\boldsymbol{p}_{\mathrm{w},1}) & r_2(\boldsymbol{p}_{\mathrm{w},1}) & \cdots & r_N(\boldsymbol{p}_{\mathrm{w},1}) \\ r_1(\boldsymbol{p}_{\mathrm{w},2}) & r_2(\boldsymbol{p}_{\mathrm{w},2}) & \cdots & r_N(\boldsymbol{p}_{\mathrm{w},2}) \\ \vdots & \vdots & \ddots & \vdots \\ r_1(\boldsymbol{p}_{\mathrm{w},M}) & r_2(\boldsymbol{p}_{\mathrm{w},M}) & \cdots & r_N(\boldsymbol{p}_{\mathrm{w},M}) \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} + \mathbf{b}, \tag{7}$$

and the desired matrix-vector form $\mathbf{y} = \mathbf{A}(\boldsymbol{\theta}_\mathrm{occ})\mathbf{f} + \mathbf{b}$. Hence, for each of the red-green-blue (RGB) camera color channels, we have the affine model above, with $\mathbf{A}(\boldsymbol{\theta}_\mathrm{occ}) \in \mathbb{R}^{M \times N}$. Forming an image of the hidden scene amounts to inverting the resulting linear system for each color channel. As we will see in Section 3, the visibility function – equivalently, the presence of the occluder – is central to the conditioning of the inversion. Without an occluder, the columns of the matrix $\mathbf{A}(\boldsymbol{\theta}_\mathrm{occ})$ are too similar for a well-conditioned recovery of $\mathbf{f}$.[2, 3, 23, 26, 33]

3. COMPUTATIONAL FIELD OF VIEW

The presence of an occluder makes certain hidden-scene light contributions computationally separable because they influence the irradiance in the camera FOV in different ways. A *computational field of view* is a set of points in the hidden scene such that shadows cast by the occluder into the camera FOV by light sources at those points would be distinct, nonempty, and not encompass the entire camera FOV. Through this definition, we find that solving an inverse problem to form an image of the CFOV does not rely only the weakly-varying $g(\mathbf{x}, \boldsymbol{p}_\mathrm{w})/\|\mathbf{x} - \boldsymbol{p}_\mathrm{w}\|_2^2$ factor of (6), but also on the visibility function $v(\mathbf{x}, \boldsymbol{p}_\mathrm{w}, \boldsymbol{\theta}_\mathrm{occ})$ differing for distinct CFOV points $\mathbf{x}$, provided all points $\boldsymbol{p}_\mathrm{w}$ in the camera FOV are considered together.

In this paper, we consider only planar CFOVs. For any fixed camera FOV and occluder parameterization $\boldsymbol{\theta}_\mathrm{occ}$, one is generally interested in the largest possible CFOV. Candidate CFOVs induced by a small rectangular occluder are illustrated in Figure 2. Light sources at points S1 and S2 create very different shadows within the camera FOV; thus, light from these locations is computationally separable. In contrast, a light source at point S3 does not create any shadow within the camera FOV, so light from S3 is not computationally separable from ambient light. With a point light source very close to the occluder, the shadow cast by the occluder could encompass the entire camera FOV, meaning light from that point would not be measured by the camera at all and thus not be estimable; this case is not illustrated.

[†]In general, the maximum value for M is set by the resolution of the digital photograph. It can be reduced by downsampling the photograph.
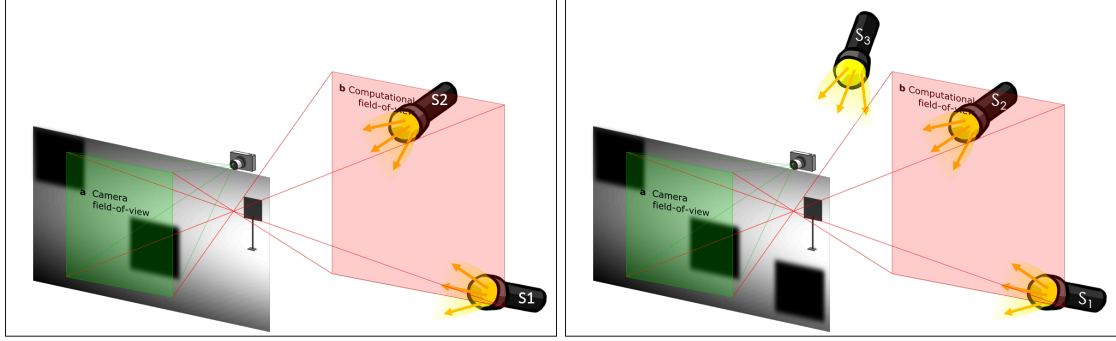
<div align="center">(a)   (b)</div>

Figure 2: **Computational field of view.** A visualization of the shadows cast by light sources that are, within the CFOV (sources at points S1 and S2 in (a) and (b)), or outside the CFOV (source at point S3 in (b)). In (a) the camera measurements contain the distinct shadows cast by the occluder due to light sources at points S1 and S2, while in (b) the shadow cast by the occluder due to a light source at point S3 falls outside of the camera FOV. Adapted from our earlier paper.[24]

Since the computational FOV is determined by visibility or occlusion from any part of the camera FOV, its boundaries can be found by projecting the boundaries of the camera FOV across the edges of the occluder as in Figure 2. In the following section, we use the notion of Fisher information to demonstrate that radiosities within the CFOV can be estimated much more reliably than those outside the CFOV.

3.1 Cramér–Rao bound

Under the discrete model for $f(\mathbf{x})$, and assuming that the camera measurements are corrupted by zero mean additive white Gaussian noise with variance σ^2, if the background contribution is negligible and $\boldsymbol{\theta}_{\mathrm{occ}}$ is assumed to be known, we obtain

$$\mathcal{I}(\mathbf{f}; \boldsymbol{\theta}_{\mathrm{occ}}) = \frac{1}{\sigma^2} \mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})^{\mathsf{T}} \mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}}), \tag{8}$$

as the Fisher information matrix for estimating the unknown vector parameter $\mathbf{f}$, from the camera measurements.[37] Here $\mathcal{I} \in \mathbb{R}^{N \times N}$. From (8) it follows that the CRB for forming an N-pixel image of the scene is given by:

$$\begin{aligned}
\mathrm{CRB}(f_n) &= [\mathcal{I}(\mathbf{f}; \boldsymbol{\theta}_{\mathrm{occ}})]_{n,n} \\
&= \sigma^2 \left[\left(\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})^{\mathsf{T}} \mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}}) \right)^{-1} \right]_{n,n},
\end{aligned} \tag{9}$$

where according to (7) the elements of the matrix $\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})$ are given by $[\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})]_{m,n} = r_n\left(\boldsymbol{p}_{\mathrm{w},m}\right)$, while $r_n\left(\boldsymbol{p}_{\mathrm{w},m}\right)$ is defined in (6).

Note that without occlusion, the visibility function $v(\mathbf{x}, \boldsymbol{p}_{\mathrm{w}}, \boldsymbol{\theta}_{\mathrm{occ}}) = 1$ for any $\mathbf{x}$ and $\boldsymbol{p}_{\mathrm{w}}$. We compute and present the CRB for each pixel at the imaging plane in Figures 3 and 4, with and without the occluder in the hidden region.

Notice that with no occluder present the CRBs are roughly uniform in the hidden-scene location and large (Figure 3a, as well as the 'No occluder' plots of Figure 4). Apparent in Figure 3b (and the 'Occluder present' plots of Figure 4) is a bimodal distribution of CRBs. The presence of the occluder yields significant CRB improvements (around 8 orders of magnitude) for a subset of the hidden scene pixels. These pixels with significantly lowered CRB are precisely those that fall within the CFOV (denoted by the red dashed lines) – obtained by projecting the camera FOV through the occluder edges, as shown in Figure 2. All other points that lie outside the CFOV have CRBs that are approximately unchanged, when compared to their 'no occluder' counterpart. Figures 4a and 4b show the CRBs for different reconstruction resolutions N. While the CFOV is still apparent, as we perhaps expect, there is a trend of increased CRBs for pixels in the CFOV as N increases.
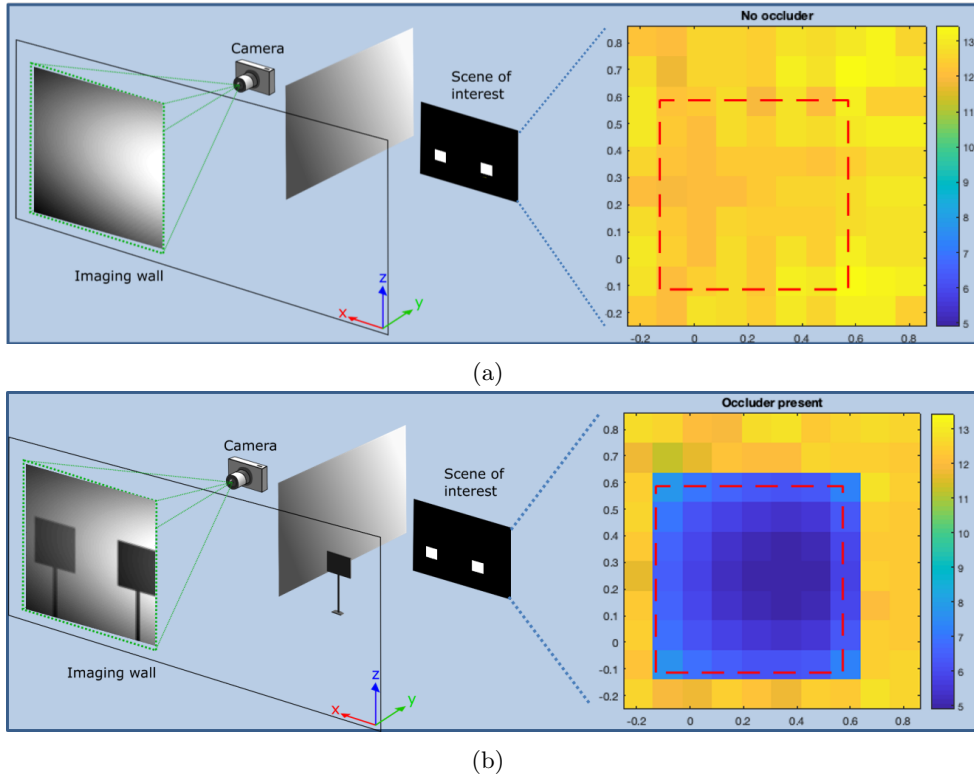
Figure 3: **Cramér–Rao bound for scene estimation.** The CRBs for each pixel are shown for $N = 10 \times 10$ imaging resolution, in the unoccluded case (**a**) and occluded setting (**b**). The dashed red lines indicate the boundary of the CFOV obtained by projecting the camera FOV through the occluder's edges and onto the hidden scene plane.
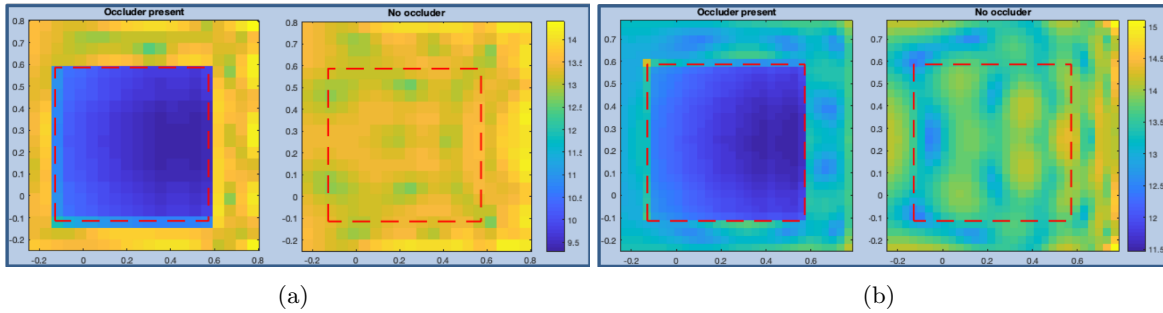

Figure 4: **Cramér–Rao bound for scene estimation.** The CRBs for each pixel are shown for different imaging resolutions, with and without an occluding object. In (**a**) scene recovery resolution $N = 20 \times 20$, while in (**b**) scene recovery resolution $N = 30 \times 30$. The dashed red lines indicate the boundary of the CFOV obtained by projecting the camera FOV through the occluder's edges and onto the hidden scene plane.

Given the improved conditioning of the problem (7), we present a computational method for for inverting (7) to estimate the unknown occluder $\boldsymbol{\theta}_{\mathrm{occ}}$ and hidden scene $\mathbf{f}$ in the subsequent section.

4. COMPUTATIONAL INVERSION

Recovering $\boldsymbol{\theta}_{\mathrm{occ}}$ and $\mathbf{f}$ from the single photograph $\mathbf{y}$ is a nonlinear problem, for which we propose a two-step recovery algorithm. Firstly, an estimate $\widehat{\boldsymbol{\theta}}_{\mathrm{occ}}$ of the unknown occluder parameterization. Secondly, we compute $\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\mathrm{occ}})$ using (7) and solve the resulting linear inverse problem to recover an image of the hidden scene. We discuss each step in more detail below.

4.1 Occluder estimation

Because the number of measurements (rows of $\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})$), M, is significantly larger than the attempted recovery resolution of the hidden scene, N, the measurements $\mathbf{y}$ reside close to a low-dimensional affine subspace that is dependent on the occluder position $\boldsymbol{\theta}_{\mathrm{occ}}$ and background $\mathbf{b}$. The occluder is estimated from $\mathbf{y}$ through

$$\widehat{\boldsymbol{\theta}}_{\mathrm{occ}} = \arg\max_{\boldsymbol{\theta}_{\mathrm{occ}}} \|\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})(\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})^{\mathsf{T}}\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}}))^{-1}\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})^{\mathsf{T}}\mathbf{y}\|_2^2, \tag{10}$$

where $\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})$ is the computed light-transport matrix for an unknown occluder position $\boldsymbol{\theta}_{\mathrm{occ}}$. Omitting the unknown background term $\mathbf{b}$ does not adversely degrade the estimate (see Supplementary Information of our earlier paper[24]). Hence, the desired occluder estimate, $\widehat{\boldsymbol{\theta}}_{\mathrm{occ}}$, is the one that minimizes the Euclidean distance between the measured photograph $\mathbf{y}$ and the range space of $\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\mathrm{occ}})$ or, equivalently, maximizes the Euclidean norm of the orthogonal projection of $\mathbf{y}$ onto the range space of $\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\mathrm{occ}})$.[38] In practice, we solve (10) by doing a grid search, and orthogonally projecting $\mathbf{y}$ onto the smaller subspace spanned by the most dominant left singular vectors of $\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})$ for each $\boldsymbol{\theta}_{\mathrm{occ}}$.

4.2 Imaging the hidden light-emitting scene

Given the estimated occluder parameterization $\widehat{\boldsymbol{\theta}}_{\mathrm{occ}}$, an estimate $\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\mathrm{occ}})$ of the true light transport matrix $\mathbf{A}(\boldsymbol{\theta}_{\mathrm{occ}})$ is computed. If the estimated occluder were perfect, and model mismatch and background contributions were inconsequential, pre-multiplying the vectorized camera measurements $\mathbf{y}$ (for each color channel) by the pseudoinverse $\widehat{\mathbf{A}}^{\dagger} = (\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\mathrm{occ}})^{\mathsf{T}}\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\mathrm{occ}}))^{-1}\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\mathrm{occ}})^{\mathsf{T}}$ would yield the least-squares estimate of the hidden scene's RGB content.

However, when the unknown background $\mathbf{b}$ is significant we *(i)* may attempt to estimate it, or *(ii)* cancel out its contributions to the measurements $\mathbf{y}$. We opt for the later for its simplicity and effectiveness. Since any light originating from outside the computational FOV has slow spatial variation, the background components in neighboring camera pixels – say b_m and b_{m+1} – are approximately equal. Thus computing the difference between neighbouring camera measurements gives $y_{m+1} - y_m \approx (\mathbf{a}_{i+1}^{\mathsf{T}}\mathbf{f} + b) - (\mathbf{a}_i^{\mathsf{T}}\mathbf{f} + b) \approx (\mathbf{a}_{i+1} - \mathbf{a}_i)^{\mathsf{T}}\mathbf{f}$, with cancelled background contribution. Repeating this differencing for all neighboring camera measurement pixels reduces the affine model (7) to $\mathbf{D}\mathbf{y} \approx \mathbf{D}\mathbf{A}\mathbf{f}$, with $\mathbf{D}$ denoting the differencing matrix.

To improve robustness of recovering $\mathbf{f}$, using this new linear system, even in the presence of noise and model mismatch, we propose the following optimization problem:

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\mathbf{D}\widehat{\mathbf{A}}\mathbf{f} - \mathbf{D}\mathbf{y}\|_2^2 + \lambda\|\mathbf{f}\|_{\mathrm{TV}}, \tag{11}$$

where the operator $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation (TV) semi-norm and λ is the TV-regularization parameter. We introduce the TV norm (specifically, the isotropic TV) in our inversion to promote spatial correlations in our reconstruction[39, 40] and solved (11) using the fast iterative shrinkage-thresholding algorithm (FISTA).[40]
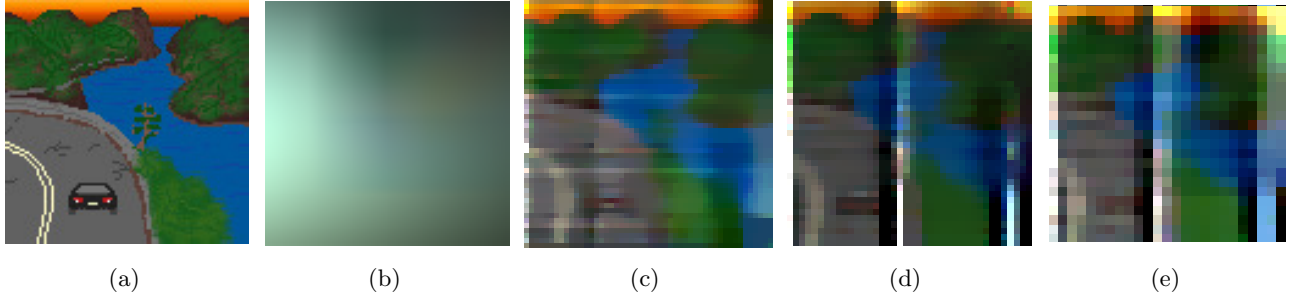
(a)	(b)	(c)	(d)	(e)

Figure 5: (a) Ground truth scene (100×100-pixels). (b) Simulated digital photograph of visible wall (1008×1008-pixels per color channel). (c) Recovered image of scene (at 50×50-pixel resolution). (d) Recovered image of scene (at 40×40-pixel resolution). (e) Recovered image of scene (at 25×25-pixel resolution).

5. EXPERIMENTAL VALIDATION

5.1 Synthetic data

In this section, we present estimation results based on synthetically generated data with Matlab. The simulated scene uses a 5 cm-by-5 cm planar square card as the occluder, placed at position the (0.52, 0.57, 0.30) m, such that it is fronto-parallel to both the hidden scene and visible wall planes. In the results below, we assume that the occluder shape is known, thus $\boldsymbol{\theta}_{\mathrm{occ}} = (0.52, 0.57, 0.30)$ m, and only recover it's position. A ground truth hidden scene (100 by 100 pixels) displayed in Figure 5a, is used to generate 3 megapixel RGB camera photograph (1 MP per color channel), shown in Figure 5b. We model camera imperfections by corrupting the simulated camera photograph with additive white Gaussian noise (AWGN) to give an SNR of 45 dB.

From the noisy photograph, we estimate the occluder parameterization $\widehat{\boldsymbol{\theta}}_{\mathrm{occ}}$ by first solving (10), and then estimate the hidden scene by solving (11). While the measurements are taken for a 100×100-pixel scene, we solve the problem for different scene resolutions, namely $N = 50 \times 50$, $N = 40 \times 40$ and $N = 25 \times 25$, and show the estimated images in Figures 5c, 5d and 5e, respectively. Corresponding occluder position estimates for each of the three target resolutions are shown in Table 1; these estimates are close to the true occluder position.

5.2 Real data

We now present reconstructions formed from real measured data. Specifically we consider two hidden scene setups. In the first setup, the hidden scenes were displayed on a Dell LCD monitor (model 2001FP, with 4:3 aspect ratio and 1280-by-1024 resolution); in the second setup, the floodlit hidden scenes comprised of colored card shapes on a black background real geometric scenes. The occluder is an opaque chair constructed from cardboard, while the visible surface is a white diffuse Elmer's foam board. A digital camera (FLIR Grasshopper3 model GS3-U3-41S4C-C) with 2016-by-2016 resolution (4.1 megapixels) – equipped with a Tamron M118FM16 lens with 16 mm focal length and f/1.4 aperture – was used to obtain a single photograph of the visible surface. Control of scene and camera was through a Lenovo ThinkPad P51s laptop computer. The setup is shown in see Figure 1. More details on the experimental materials and methods can be found in our earlier article.[24]

In Figure 6, we show reconstructions of three hidden scenes (scene 'A' is light emitting, while scenes 'B' and 'C' are light reflecting scenes). From the camera measurements shown in the second column, we reconstruct images (third column) of the corresponding ground truth hidden scene (first column).

Table 1: **Occluder localization results.** Estimated occluder positions using proposed localization algorithm for each example hidden scene.

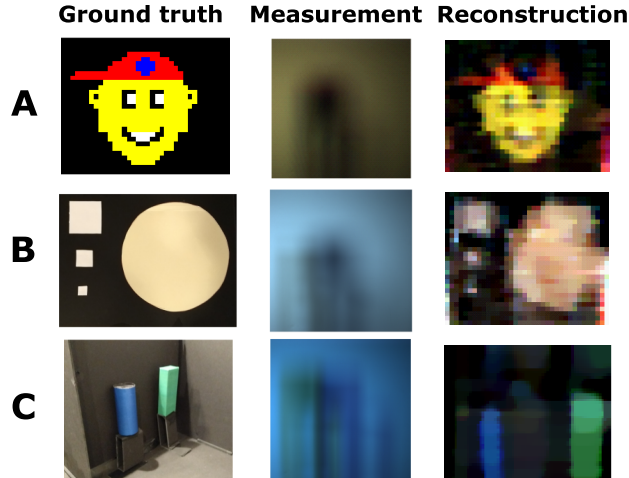| Occluder location [m]	$(\boldsymbol{\theta}_{\mathrm{occ}})_x$	$(\boldsymbol{\theta}_{\mathrm{occ}})_y$	$(\boldsymbol{\theta}_{\mathrm{occ}})_z$
Ground truth	0.520	0.570	0.300
50 by 50	0.515	0.563	0.301
40 by 40	0.528	0.582	0.300
25 by 25	0.511	0.585	0.308

Figure 6: **Ground truths, measurements and reconstructions for three different scenes: (A)** An emitting scene displayed on an LCD monitor. **(B)** A reflecting scene comprising colored card shapes on a black background. **(C)** A reflecting scene comprising 3D shapes made from colored cards. Results taken from the Supplementary Information of our earlier paper.[24]
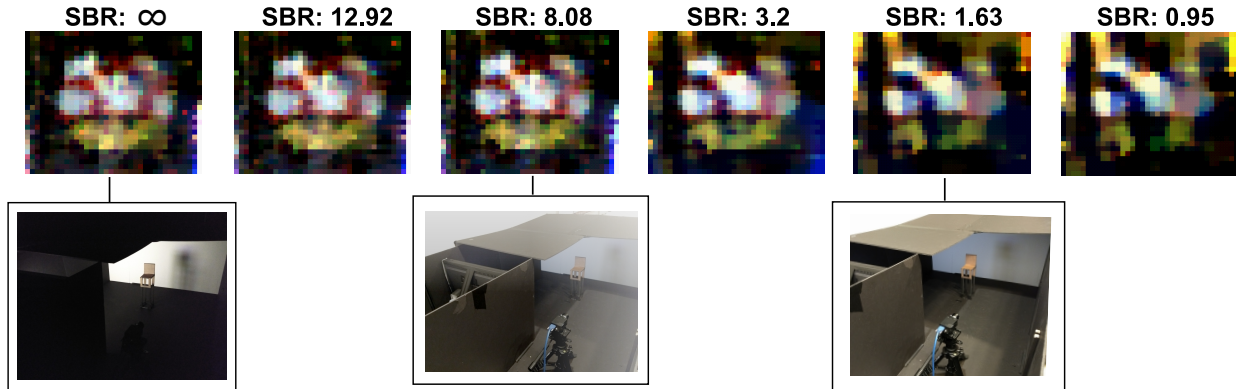

Figure 7: **Reconstructions with increasing ambient light level.** Robustness to light is demonstrated by performing reconstructions using measurements taken with decreasing signal-to-background-ratio (SBR), defined here as the ratio of the mean of signal sans ambient light to the mean of ambient light measurement. Results taken from the Supplementary Information of our earlier paper.[24]

The estimated images of the hidden scene closely match the ground truth scenes, even for passively reflecting scenes and for 3D scenes.

5.2.1 With ambient background illumination

In Section 3, it was shown that all light originating outside of the CFOV of the system will have a smooth and slowly varying contribution to the camera measurements. Therefore, all such contributions are approximately cancelled by differencing and solving (7) to obtain an image of the hidden scene (see Section 4.2).

Figure 7 shows reconstructions obtained from measurements made in the presence of ambient light. The quality of the reconstructions degrade gradually with increasing ambient light levels, and at roughly equal signal and ambient light strength, the obtained reconstruction is still visually interpretable. This gradual degradation is likely due to the ambient light level reducing the *effective dynamic range* of the camera. We also note that there are significant light contributions from outside the CFOV present in the non-emitting scene measurements caused by reflections of the floodlight from the side walls (Fig. 6 (B) and (C)), yet the reconstructions do not show any significant degradation in reconstruction quality due to this.

6. CONCLUSION

Using our proposed method, we demonstrated the recovery of a full-color 2D view of a variety of hidden scenes, whether self-emitting or reflecting, and 2D or 3D, with surprising accuracy (see Figures 5 and 6) from a single color photograph of a visible wall. In the reconstructions, even the smaller details are resolvable, such as the teeth and the smallest square in Fig. 6**A** and **B**, respectively. Simulated results summarized in Table 1 and Figure 5 showed reliable recovery of the occluder location and imaging of a fairly complicated scene, respectively. Our method for computational periscopy was enabled by penumbrae cast on the visible wall by an opaque occluder in the hidden scene. Through a Cramér–Rao bound analysis, we showed that the occluder yields a computational field of view – a portion of the hidden region that is well conditioned for recovery. Using real experimental data, we also demonstrated the robustness of the method to ambient illumination, or light emanating from regions outside the computational field of view.

ACKNOWLEDGMENTS

REFERENCES

[1] Kirmani, A., Hutchison, T., Davis, J., and Raskar, R., "Looking around the corner using transient imaging," in [*Proc. IEEE 12th Int. Conf. Computer Vision*], 159–166 (2009).

[2] Kirmani, A., Jeelani, H., Montazerhodjat, V., and Goyal, V. K., "Diffuse imaging: Creating optical images with unfocused time-resolved illumination and sensing," *IEEE Signal Process. Lett.* **19**(1), 31–34 (2012).

[3] Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M. G., and Raskar, R., "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nature Commun.* **3**, 745 (2012).

[4] Gupta, O., Willwacher, T., Velten, A., Veeraraghavan, A., and Raskar, R., "Reconstruction of hidden 3D shapes using diffuse reflections," *Opt. Express* **20**(17), 19096–19108 (2012).

[5] Heide, F., Hullin, M. B., Gregson, J., and Heidrich, W., "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graphics* **32**(4), 1 (2013).

[6] Xu, K., Jin, W., Zhao, S., Liu, J., Guo, H., Qiu, S., and Wu, D., "Image contrast model of non-line-of-sight imaging based on laser range-gated imaging," *Opt. Eng.* **53**(6), 061610 (2013).

[7] Heide, F., Xiao, L., Heidrich, W., and Hullin, M. B., "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], 3222–3229 (2014).

[8] Laurenzis, M. and Velten, A., "Nonline-of-sight laser gated viewing of scattered photons," *Opt. Eng.* **53**(2), 023102 (2014).

[9] Buttafava, M., Zeman, J., Tosi, A., Eliceiri, K., and Velten, A., "Non-line-of-sight imaging using a time-gated single photon avalanche diode," *Opt. Express* **23**(16) (2015).

[10] Gariepy, G., Tonolini, F., Henderson, R., Leach, J., and Faccio, D., "Detection and tracking of moving objects hidden from view," *Nature Photonics* **10**, 23–27 (2016).

[11] Kadambi, A., Zhao, H., Shi, B., and Raskar, R., "Occluded imaging with time-of-flight sensors," *ACM Trans. Graphics* **35**(2), 1–12 (2016).

[12] Chan, S., Warburton, R. E., Gariepy, G., Leach, J., and Faccio, D., "Non-line-of-sight tracking of people at long range," *Opt. Express* **25**(9), 10109 (2017).

[13] Tsai, C.-y., Kutulakos, K. N., Narasimhan, S. G., and Sankaranarayanan, A. C., "The geometry of first-returning photons for non-line-of-sight imaging," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], 7216–7224 (2017).

[14] O'Toole, M., Lindell, D. B., and Wetzstein, G., "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature* **555**, 338–341 (Mar. 2018).

[15] Heide, F., O'Toole, M., Zang, K., Lindell, D. B., Diamond, S., and Wetzstein, G., "Non-line-of-sight imaging with partial occluders and surface normals," *ACM Trans. Graph.* **38**, 22:1–22:10 (May 2019).

[16] Lindell, D. B., Wetzstein, G., and O'Toole, M., "Wave-based non-line-of-sight imaging using fast f-k migration," *ACM Trans. Graph* **38**(4), 13 (2019).

[17] Liu, X., Guillén, I., Manna, M. L., Nam, J. H., Reza, S. A., Le, T. H., Gutierrez, D., Jarabo, A., and Velten, A., "Virtual wave optics for non-line-of-sight imaging." arXiv:1810.07535 [cs.CV] (July 2019).

[18] Katz, O., Heidmann, P., Fink, M., and Gigan, S., "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," *Nature Photonics* **8**(10), 784–790 (2014).

[19] Smith, B. M., O'Toole, M., and Gupta, M., "Tracking multiple objects outside the line of sight using speckle imaging," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], 6258–6266 (June 2018).

[20] Viswanath, A., Rangarajan, P., MacFarlane, D., and Christensen, M. P., "Indirect imaging using correlography," in [*Imaging and Applied Optics 2018 (3D, AO, AIO, COSI, DH, IS, LACSEA, LS&C, MATH, pcAOP)*], CM2E.3, Optical Society of America (2018).

[21] Klein, J., Peters, C., Martín, J., Laurenzis, M., and Hullin, M. B., "Tracking objects outside the line of sight using 2D intensity images," *Scientific Reports* **6**, 32491 (Oct. 2016).

[22] Bouman, K. L., Ye, V., Yedidia, A. B., Durand, F., Wornell, G. W., Torralba, A., and Freeman, W. T., "Turning corners into cameras: Principles and methods," in [*Proc. 23rd IEEE Int. Conf. Computer Vision*], 2270–2278 (2017).

[23] Baradad, M., Ye, V., Yedidia, A. B., Durand, F., Freeman, W. T., Wornell, G. W., and Torralba, A., "Inferring light fields from shadows," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], 6267–6275 (2018).

[24] Saunders, C., Murray-Bruce, J., and Goyal, V. K., "Computational periscopy with an ordinary digital camera," *Nature* **565**(7740), 472 (2019).

[25] Saunders, C., Murray-Bruce, J., and Goyal, V. K., "Computational periscopy without time-resolved sensing," in [*Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP)*], CM2A.4, Optical Society of America (2019).

[26] Yedidia, A. B., Baradad, M., Thrampoulidis, C., Freeman, W. T., and Wornell, G. W., "Using unknown occluders to recover hidden scenes," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], (June 2019).

[27] Lindell, D. B., Wetzstein, G., and Koltun, V., "Acoustic non-line-of-sight imaging," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], (June 2019).

[28] Maeda, T., Wang, Y., Raskar, R., and Kadambi, A., "Thermal non-line-of-sight imaging," in [*Proc. IEEE Int. Conf. Computational Photography*], 1–11 (May 2019).

[29] Laurenzis, M., Christnacher, F., Klein, J., Hullin, M. B., and Velten, A., "Study of single photon counting for non-line-of-sight vision," **9492**, 94920K (2015).

[30] Laurenzis, M., Christnacher, F., and Velten, A., "Study of a dual mode SWIR active imaging system for direct imaging and non-line-of-sight vision," in [*Proc. SPIE Laser Radar Technology and Applications XX and Atmospheric Propagation XII*], **9465**, 946509 (2015).

[31] Tsai, C.-Y., Sankaranarayanan, A. C., and Gkioulekas, I., "Beyond volumetric albedo – a surface optimization framework for non-line-of-sight imaging," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], (June 2019).

[32] Xin, S., Nousias, S., Kutulakos, K. N., Sankaranarayanan, A. C., Narasimhan, S. G., and Gkioulekas, I., "A theory of fermat paths for non-line-of-sight shape reconstruction," in [*Proc. IEEE Conf. Computer Vision and Pattern Recognition*], (June 2019).

[33] Thrampoulidis, C., Shulkind, G., Xu, F., Freeman, W. T., Shapiro, J. H., Torralba, A., Wong, F. N. C., and Wornell, G. W., "Exploiting occlusion in non-line-of-sight active imaging," *IEEE Trans. Computational Imaging* **4**, 419–431 (Sept 2018).

[34] Yedidia, A., Thrampoulidis, C., and Wornell, G., "Analysis and optimization of aperture design in computational imaging," in [*Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*], 4029–4033 (April 2018).

[35] Seidel, S. W., Ma, Y., Murray-Bruce, J., Saunders, C., Freeman, W. T., Yu, C. C., and Goyal, V. K., "Corner occluder computational periscopy: Estimating a hidden scene from a single photograph," in [*Proc. IEEE Int. Conf. Computational Photography*], 1–9 (May 2019).

[36] Kajiya, J. T., "The rendering equation," in [*Proc. 13th Annual Conf. Computer Graphics and Interactive Techniques (SIGGRAPH '86)*], 143–150 (1986).

[37] Kay, S. M., [*Fundamentals of Statistical Signal Processing*], Prentice Hall PTR (1993).

[38] Vetterli, M., Kovačević, J., and Goyal, V. K., [*Foundations of Signal Processing*], Cambridge Univ. Press (2014).

[39] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm," *SIAM J. Imaging Sciences* **2**(1), 183–202 (2009).

[40] Beck, A. and Teboulle, M., "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Processing* **18**(11), 2419–2434 (2009).