Fast Computational Periscopy in Challenging Ambient Light Conditions through Optimized Preconditioning

Charles Saunders, *Student Member, IEEE,* and Vivek K Goyal, *Fellow, IEEE*

Abstract—Non-line-of-sight (NLOS) imaging is a rapidly advancing technology that provides asymmetric vision: seeing without being seen. Though limited in accuracy, resolution, and depth recovery compared to active methods, the capabilities of passive methods are especially surprising because they typically use only a single, inexpensive digital camera. One of the largest challenges in passive NLOS imaging is ambient background light, which limits the dynamic range of the measurement while carrying no useful information about the hidden part of the scene. In this work we propose a new reconstruction approach that uses an optimized linear transformation to balance the rejection of uninformative light with the retention of informative light, resulting in fast (video-rate) reconstructions of hidden scenes from photographs of a blank wall under high ambient light conditions.

Index Terms—Linear inverse problems, Passive Non-Line-of-Sight Imaging, Occlusion, Optimized preconditioning

1 Introduction

Systems and algorithms that enable imaging of scenes that are not otherwise accessible or visible could prove to be useful in areas such as autonomous navigation, search-and-rescue operations, and even medical imaging. Since its first conception and demonstration [1], [2], non-line-of-sight (NLOS) imaging has become an active research area with many significant advances.

NLOS imaging can be broadly split into two catagories: active, and passive. In active NLOS imaging, the observer has some control over the illumination of the scene, and in passive imaging, the observer simply measures light that is already present within the environment. Most active imaging modalities use pulsed lasers to probe the hidden scene, by bouncing light off of a visible relay surface, into the hidden area, back to the relay surface, and finally to ultrafast detectors such as single-photon avalanche diodes with time-correlated single photon counting modules [3], [4], or streak cameras [2]. The time-of-flight of the laser photons returning then provides information about the hidden scene that can be decoded. Different system architectures and information encoding principles provide different capabilities, such as 3D object imaging [3], occluder-aided imaging [5] including 2.5D full-room reconstructions [6], and recovering hidden motion [7], [8]. There are a variety of reconstruction approaches used, including back-projection [9], fast convolutional methods [3], [10], *f-k* migration [11], Fermat paths [12], speckle correlations [13], and more. Reference [14] provides a thorough review of many existing NLOS techniques. The controlled illumination and time-resolved sensing allows for impressive 3D recovery capabilities that can not be achieved by passive sensing. However, many active NLOS imaging methods require expensive, ultra-fast optics, impractical acquisition times, or high-powered lasers that are not eye safe.

In passive imaging, many techniques use the shadows cast by occluding objects to make inferences about the hidden scene [15], [16], [17], [18]. Others use spatial coherence [19], or time-resolved sensing without controlling the illumination [20]. Some techniques recover both the scene and the occluder structure using motion [21] or deep matrix factorization [22]. Passive imaging techniques avoid some of the issues with active imaging, as they require less expensive equipment, are stealthier, and often require shorter acquisition time. However, passive methods are not competitive with active imaging in terms of reconstruction accuracy or recovery of 3D information [23]. Furthermore, these methods face difficulties when there is significant ambient light intensity or very low signal light intensity. This is particularly problematic in the common situation where the light source(s) illuminating the hidden area also directly illuminate the surface in the detector's field-of-view (FOV), resulting in measurements with a low signal-to-background ratio (SBR). Some passive methods deal with this by recovering motion only, as this essentially allows the temporally constant background to be cancelled [17]. Others subtract a pre-calibrated background measurement, estimate the background contribution by fitting a linear model [24], aim to cancel the background contribution in the measurements [15], or only function well in high SBR regimes that are unlikely to be seen outside of a laboratory setting. Background estimation or cancellation has also been shown to be important for estimating other scene parameters outside of imaging [23].

In this paper, we investigate the possibility of improving robustness to ambient light in passive NLOS imaging through algorithmic development. Instead of relying on time-resolved measurements, movement in the scene, or low background light levels, we formulate a new method

- *Department of Electrical and Computer Engineering, Boston University, Boston, MA, 02215.*
  E-mail: goyal@bu.edu

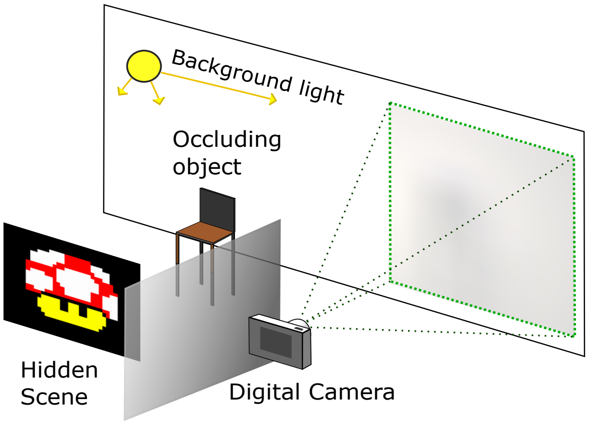

Fig. 1. **The imaging scenario**. A camera takes a photograph of a Lambertian relay surface, measuring light arriving from both the hidden scene, some of which is occluded, and also additional ambient background light. Based on Fig. 1 from [27].

inspired by generalized pseudoinverses [25], [26] that tolerates high ambient light contributions to the measurements and performs fast reconstruction of out-of-sight two-dimensional scenes using only a single photograph of a plain wall, as depicted in Fig. 1.

2 Related Work

In [15], a planar hidden scene is outside of the direct line of sight of the observer and is to be reconstructed. The observer is able to take a photograph of a Lambertian surface with constant albedo that can receive light emitted or reflected by the hidden scene. Between the hidden scene and the visible surface is an occluding object that casts subtle shadows and penumbrae that punctuate the measurement on the visible wall. These shadows turn an extremely ill-conditioned inverse problem into one that is tractable. All positions in the hidden area from which a light source would cast a unique shadow into the camera FOV comprise the 'computational field-of-view' in which recoveries can be made. However, despite some attempts to mitigate the issue, high background levels cause the reconstruction algorithm proposed in [15] to fail.

The photograph measurement of the visible wall can be modelled using

$$\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{b}, \tag{1}$$

where $\mathbf{A}$ is the light transport model, $\mathbf{f}$ is the discretized hidden scene, $\mathbf{y}$ is the camera measurement (i.e., photograph of the visible wall), and $\mathbf{b}$ is unmodelled background light that is not originating directly from within the computational FOV. The light transport matrix $\mathbf{A}$ can be computed column-by-column by rendering the partially occluded measurement that would be seen on the relay surface if only the single scene pixel is radiating light.

Given this model and a potentially noisy camera measurement $\mathbf{y}$, the authors of [15] solve a total variation-regularized inverse problem (separately for each of the three RGB color channels) in order to recover the image hidden from the observer's view:

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

where matrix $\mathbf{D}$ takes differences between neighboring components. The authors note that background light which is originating from outside of the computational FOV, especially from the far field, will vary slowly spatially within the camera's FOV and can be approximated as constant, i.e., $\mathbf{b}_{i+1} \approx \mathbf{b}_i \approx b$. Therefore, this matrix aims to cancel out the background light contribution,

$$(\mathbf{D}\mathbf{y})_i = \mathbf{y}_{i+1} - \mathbf{y}_i \approx (\mathbf{a}_{i+1}^\mathsf{T}\mathbf{f} + b) - (\mathbf{a}_i^\mathsf{T}\mathbf{f} + b) \approx (\mathbf{D}\mathbf{A}\mathbf{f})_i. \tag{3}$$

This approach shows reasonable robustness to unmodelled ambient light contributions at signal-to-background ratios above 3 or so, but begins to break down at higher ambient light levels. Furthermore, this optimization problem is solved using accelerated proximal gradient descent. The speed of this is inhibited by the poor conditioning of the matrix $\mathbf{A}$. Additionally, a second iterative algorithm to solve the proximal operator of the total variation penalty is performed following every gradient step. For the imaging scenario in which the experiments were performed, $\mathrm{cond}(\mathbf{A}) \approx 7960$ and $\mathrm{cond}(\mathbf{DA}) \approx 270$.

3 Proposed Method

Taking inspiration from the finite differencing procedure used in the original computational periscopy work [15], which attempts to cancel out background contributions, we aim to develop an algorithm that both improves robustness to background light and also speeds up the reconstruction. To do so, we seek a preconditioner matrix $\mathbf{P}$ such that $\|\mathbf{P}\mathbf{b}\| \approx 0$ for any reasonable background $\mathbf{b}$, and also such that $\mathbf{PA}$ is well conditioned (ideally, $\mathbf{PA} = \mathbf{I}$). The problem of finding such a $\mathbf{P}$ can be considered a search for an approximate left inverse with additional background-cancelling properties.

To begin, we note that any background light can be modelled by integrating the response from point lights positioned over the surface emitting or reflecting the light, i.e., $\mathbf{b} = \int_{\alpha\in\beta} e(\alpha)\boldsymbol{\psi}(\alpha)\, d\alpha$, where $\boldsymbol{\psi}(\alpha)$ is the ambient light contribution to the camera measurement from a point light with position $\alpha = (x, y, z)$, $e(\alpha)$ is a scalar emittance factor, and the domain β is over the surface(s) contributing to the measurement. A simple example of a contributor of background light is a ceiling light, which could be modelled as a single point light that directly illuminates the measurement surface. Another example is unmodelled multi-bounce light that originates from the hidden scene of interest, but subsequently reflects off of other surfaces before reaching the measurement wall; in this case, β comprises all points on the final reflecting surface. The point light response is given by

$$\boldsymbol{\psi}(\alpha)_i = ((\alpha - \mathbf{c}_i)\cdot\mathbf{n})/\|\alpha - \mathbf{c}_i\|_2^3, \tag{4}$$

where $\mathbf{c}_i$ is the position the ith camera pixel sees on the wall and $\mathbf{n}$ is the wall normal ($\|\mathbf{n}\|_2 = 1$). In order to find a suitable matrix $\mathbf{P}$, one could consider solving the following problem:

$$\begin{aligned} \hat{\mathbf{P}} = \arg\min_{\mathbf{P}} \quad & \int_{\alpha\in\beta} \|\mathbf{P}\boldsymbol{\psi}(\alpha)\|_2^2\, d\alpha \\ \text{s.t.} \quad & \mathbf{PA} = \mathbf{I}, \end{aligned} \tag{5}$$

where here β is *all* possible positions from which background light can originate (i.e., everywhere in space such that a point light at that position will contribute some amount of light to the measurement surface). This aims to ensure that multiplying by $\mathbf{P}$ will cancel *any* background contribution, and the constraint ensures that the multiplication $\mathbf{PA}$ is well conditioned (here taken to the extreme of perfectly conditioned). Omitting the emittance factor is equivalent to assuming $e(\alpha) = 1$ for all α, implying that cancellation of light from all points in β is equally important.

As is, this problem is intractable. We suggest to discretize β, resulting in a matrix $\mathbf{B}$ whose M columns comprise $\boldsymbol{\psi}(\alpha)$ evaluated for a number of different values of α spanning the appropriate spatial extent (more detail on this can be found in Sec. 3.1):

$$\int_{\alpha\in\beta} \|\mathbf{P}\boldsymbol{\psi}(\alpha)\|_2^2\, d\alpha \approx \sum_i^M \|\mathbf{P}\boldsymbol{\psi}(\alpha_i)\|_2^2 = \|\mathbf{PB}\|_{\mathrm{F}}^2. \quad (6)$$

This is a reasonable approximation, as we anticipate all background light to emerge from the far field. This implies that $\boldsymbol{\psi}(\alpha) \approx \boldsymbol{\psi}(\alpha + \Delta)$ for small enough $\|\Delta\|_2$, i.e., the response varies very slowly with α and hence any reasonable $\boldsymbol{\psi}(\alpha)$ will lie in or extremely close to the subspace spanned by $\mathbf{B}$. The problem could then be solved using an iterative approach [25], but this is costly. Instead, we can also relax the constraint:

$$\widehat{\mathbf{P}} = \arg\min_{\mathbf{P}} \|\mathbf{PB}\|_{\mathrm{F}}^2 + \omega\|\mathbf{PA} - \mathbf{I}\|_{\mathrm{F}}^2, \quad (7)$$

where $\omega \in [0, \infty)$ controls the relaxation. Then, we arrive at a more familiar result:

$$\widehat{\mathbf{P}} = \arg\min_{\mathbf{P}} \|\mathbf{PB}\|_{\mathrm{F}}^2 + \|\mathbf{PA} - \mathbf{I}\|_{\mathrm{F}}^2 = \begin{bmatrix}\mathbf{I}\\ 0\end{bmatrix} [\mathbf{A}, \mathbf{B}]^\dagger, \quad (8)$$

where † represents the Moore–Penrose pseudoinverse and the factor ω has been absorbed into $\mathbf{B}$. This result can be interpreted as simply calculating the pseudoinverse of $\mathbf{A}$ augmented with additional columns containing prototype backgrounds, and then discarding the rows that pertain to an estimate of the background contribution as we are not concerned with them. Now, using $\widehat{\mathbf{P}}$ in place of the finite differencing matrix in Eq. (2), we see that forming our solution is as simple as left multiplying by $\widehat{\mathbf{P}}$ and 'denoising' the result with the proximal operator for total variation:

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\widehat{\mathbf{P}}\widehat{\mathbf{A}}\mathbf{f} - \widehat{\mathbf{P}}\mathbf{y}\|_2^2 + \lambda\|\mathbf{f}\|_{\mathrm{TV}} \quad (9)$$

$$\approx \arg\min_{\mathbf{f}} \|\mathbf{f} - \widehat{\mathbf{P}}\mathbf{y}\|_2^2 + \lambda\|\mathbf{f}\|_{\mathrm{TV}} \quad (10)$$

$$= \mathrm{prox}_{TV,\lambda}(\widehat{\mathbf{P}}\mathbf{y}), \quad (11)$$

which significantly reduces computation time compared to the algorithm in [15]. The proximal operator can be solved efficiently using a variety of methods [28], [29], [30].

This estimation is followed by a final post-repair stage, in which pixels likely to be incorrect are identified and replaced with the output of a median filter (per color channel). To identify potentially incorrect pixels, we calculate the sample mean of the magnitude of eight neighboring pixels, and we consider the current pixel to be incorrect if its magnitude is more than some threshold away from the neighborhood sample mean. This helps to improve the final

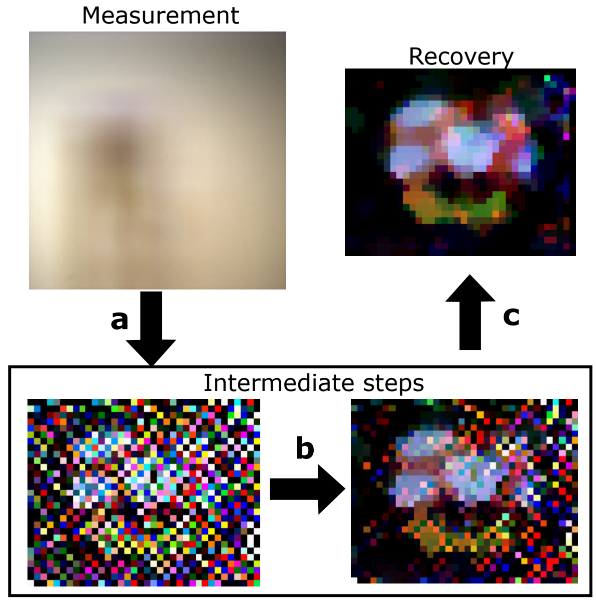

Fig. 2. **Reconstruction procedure**. **a.** Measurement is left-multiplied by $\widehat{\mathbf{P}}$. **b.** Total variation denoising using the proximal operator. **c.** Median filtering.

estimate by reducing the severity of artefacts, especially in regions of the hidden space which are worst conditioned for recovery (which can be observed using the Cramér-Rao bound [27]). The full recovery procedure is summarised in Algorithms 1 and 2 and in Fig. 2. We note that more sophisticated methods could be used, such as adaptive vector median filters [31], but we did not find that the trade-off between improvement to the reconstruction and increase in computation time was worthwhile. Similarly, one could consider total-variation regularization that is not separable among color channels, however, we found the same trade-off behaviour. We anticipate that these more sophisticated techniques may prove more useful at higher reconstruction resolutions.

The search for a generalized pseudoinverse opens up many interesting options which could be explored. For instance, while the use of the Frobenius norm here leads to a computationally efficient and familiar result, different norms will result in different, and perhaps interesting, outcomes. Other terms can be included in Eq. (7), too. For example, a sparsifying term $\|\mathbf{P}\|_1$ could result in a sparse pseudoinverse that may speed up subsequent multiplications by $\widehat{\mathbf{P}}$, but we do not explore this option further in this paper.

3.1 Constructing $\mathbf{B}$

When the sampling used to generate $\mathbf{B}$ is of sufficient breadth and density, the same $\mathbf{B}$ matrix can be used in reconstructions for different scene configurations, as long as the properties of the surface the camera is pointing at stay the same (i.e., constant albedo and Lambertian); different orientations and positions of the relay surface are then simply changes of the coordinate system.

For the experiments throughout this paper, we construct $\mathbf{B}$ by calculating the measurement due to point lights at various positions using Eq. (4). We define the center of

Algorithm 1 Proposed Algorithm

Input: $\mathbf{A}$, $\mathbf{y}$, β, λ.
Output: $\hat{\mathbf{f}}$

```
for α_i ∈ β do
    B_i = ψ(α_i)
end for
P̂ = [I; 0] [A, B]^†
for each color channel do
    f̄ = prox_{TV,λ}(P̂y)
end for
return f̂ = post-repair(f̄)
```

$$\widehat{\mathbf{P}} = \begin{bmatrix}\mathbf{I}\\0\end{bmatrix}[\mathbf{A},\mathbf{B}]^{\dagger}, \quad \bar{\mathbf{f}} = \mathrm{prox}_{TV,\lambda}(\widehat{\mathbf{P}}\mathbf{y})$$

Algorithm 2 Post-repair Algorithm

Input: $\bar{\mathbf{f}} \in \mathbb{R}^{3\times N}$, threshold κ
Output: $\hat{\mathbf{f}}$

```
f̂ = f̄
m = median-filter(f̄) (for separate color channels)
for i = [0, 1, ...N − 1] do
    Calculate neighborhood magnitude mean μ
    if |‖f̄_i‖ − μ| > κ then
        f̂_i = m_i
    end if
end for
return f̂
```

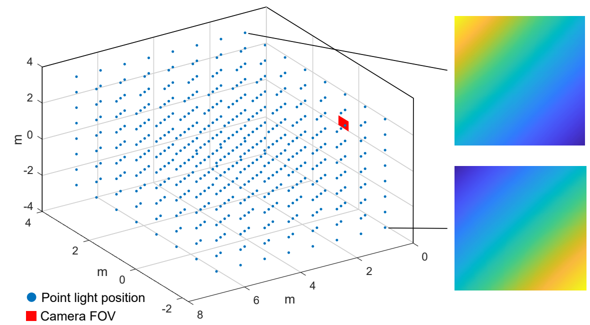

Fig. 3. **Sampling possible background contributions.** Each dot represents a point light that contributes a column to $\mathbf{B}$. Two examples are shown.

the camera's FOV on the relay surface as $(0,0,0)$, and then evaluate $\boldsymbol{\psi}(\alpha)$ for α on a 3-dimensional grid from $(-3,1,-3)$ m to $(3,7,3)$ m with 75 cm steps, as depicted in Fig. 3. This results in $\mathbf{B}$ with 512 columns, each sampling a different light position over a $6\times6\times6$ m cube. It is also possible to sum together the response from numerous point lights over a small area to more accurately sample the space – but we found that the extra expenditure did not noticeably effect the results. Similarly, finer discretization did not significantly improve reconstruction quality but increased computational effort in forming $\widehat{\mathbf{P}}$. Many other more sophisticated discretization strategies are possible and are worth exploring – for instance importance sampling the space to discretize more finely near expected background light sources (e.g., discretizing the space above the measurement plane more finely due to the expectation of ceiling lights or sunlight from above). Similarly, discretizing more finely closer to the measurement plane may improve the results as the contribution from point lights farther from the measurement plane vary more slowly with changes in position.

3.2 Extension to Video Reconstruction

Often, video reconstruction under high ambient lighting conditions is simplified by taking differences through time to essentially cancel out any background contributions which do not vary in time, for example in [17]. However, reconstructions of this type are only capable of recovering *movement* within the scene, rather than the *whole scene through time*. Given the potential for fast, video-frame-rate reconstructions using the proposed algorithm, we extend here the algorithm to make more use of the temporal structure in video sequences. For the ith frame we denote the measurement $\mathbf{y}_i$. For the first frame, we solve the problem in Eq. (11). Then, after each subsequent measurement is acquired, we could solve the following problem:

$$\hat{\mathbf{f}}_i = \arg\min_{\mathbf{f}} \|\widehat{\mathbf{P}}\widehat{\mathbf{A}}\mathbf{f} - \widehat{\mathbf{P}}\mathbf{y}_i\|_2^2 + \lambda\|\mathbf{f}\|_{\mathrm{TV}} + \lambda_2\|\mathbf{f}-\mathbf{f}_{i-1}\|_1, \quad (12)$$

where $\mathbf{f}_{i-1}$ is the previous frame's estimate, and hence continuity in time is promoted. To simplify this, we can instead simply stack $\mathbf{f}$ and $\mathbf{f}_{i-1}$ into one vector and denoise this with the 3-dimensional total variation proximal operator, for which there are fast implementations [32]. Following this, we can apply the same post-repair algorithm as before. However, we also include the previous frame in time within the window (now a $3\times3\times2$ window); for each pixel in $\mathbf{f}$, we now take the median of the neighboring 8 pixels at the current frame as well as 9 from the previous frame.

3.3 Analysis of Relaxed Solution

In order for the equality in Eq. (10) to hold, we require that $\mathbf{PA} \approx \mathbf{I}$, which is not explicitly guaranteed when solving the relaxed problem in Eq. (7) for general $\mathbf{A}$ and $\mathbf{B}$. However, the relaxed problem has the attractive quality that the solution has a simple and familiar closed form. In essence, we wish to find a matrix $\mathbf{P}$ such that $\mathbf{PA} = \mathbf{I}$ and $\mathbf{B} \in \mathrm{Null}(\mathbf{P})$. This implies that the closer to orthogonal the subspaces spanned by $\mathbf{A}$ and $\mathbf{B}$ are, the better the solution to the relaxed problem will be. Empirical evidence arises from the reasonable success of the use of the finite difference matrix, $\mathbf{D}$. This matrix is justified by the authors of [15] as approximately cancelling any slowly varying background contribution from outside of the computational FOV. This can also be interpreted as any background contribution being close to lying in the null space of $\mathbf{D}$. Similarly, the discontinuities in the differences due to the penumbrae punctuating the light from the hidden scene then ensures that $\mathbf{A}$ does *not* lie close to $\mathrm{Null}(\mathbf{D})$.

This can be explored quantitatively by calculating the angle between the subspaces spanned by $\mathbf{A}$ and $\mathbf{B}$ [33]. This will differ depending on the exact imaging situation, but for the setup depicted in Fig. 1, we calculate this angle to be 89.989 degrees. This is very close to orthogonal, hence we anticipate the solution to Eq. (7) to be acceptable. Indeed, we find that $\widehat{\mathbf{P}}$ is an acceptable pseudoinverse ($\widehat{\mathbf{P}}\mathbf{A} \approx \mathbf{I}$), and that the average percentage of residual background $\frac{100}{M}\sum_{i=1}^{M}\|\widehat{\mathbf{P}}\mathbf{B}_i\|_2/\|\mathbf{B}_i\|_2$ is less than 5×10^{-13}, implying that $\widehat{\mathbf{P}}$ also successfully cancels out the background examples in $\mathbf{B}$.

From a statistical point of view, $\widehat{\mathbf{P}}\mathbf{A} \approx \mathbf{I}$ implies that the reconstruction is approximately unbiased (in the absence of ambient light). The conventional Moore–Penrose pseudoinverse minimizes the variance of the reconstruction under a white Gaussian noise model. Comparing the Cramér-Rao bounds (CRBs) for estimating $\mathbf{f}$ using only $\mathbf{A}$ (conventional pseudoinverse) or $[\mathbf{A}, \mathbf{B}]$ (as proposed here) quantifies any increased sensitivity to additive noise caused by the suppression of background. The CRBs for the ith hidden scene pixel in each case are the ith diagonal elements of $(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}$ and $([\mathbf{A}, \mathbf{B}]^\mathsf{T}[\mathbf{A}, \mathbf{B}])^{-1}$. Indeed, we find that for the scenario in Fig. 1, the CRBs are nearly identical; the ratio of the CRBs totaled over all the pixels, $(\sum_{i=1}^{N}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1})_{i,i}) \,/\, (\sum_{i=1}^{N}([\mathbf{A}, \mathbf{B}]^\mathsf{T}[\mathbf{A}, \mathbf{B}])^{-1})_{i,i})$, is 1.0001.

4 Experimental Results

4.1 Tolerating Background

To see the relationship between reconstruction quality and SBR using the proposed method and the method from [15], we perform an experiment. Firstly, eight photographs of a white, reasonably Lambertian poster board were collected under various lighting conditions (outside in sunlight, inside in sunlight, and inside under room lighting at various orientations), to sample a variety of realistic ambient lighting conditions. Additionally, this data set was augmented by flipping each image about each axis, resulting in 32 background examples. One test measurement per background example was generated by adding a scalar multiple of each of the collected ambient light measurements to the experimental data[1] from [15] *without* ambient light, to achieve a desired SBR. The augmented data is normalized to simulate a fixed total acquisition time, and white Gaussian noise was added to the measurement to maintain a realistic signal-to-noise ratio of 60 dB, similar to that of the original experimental data [15]. This has the effect of pushing the signal components closer to the noise floor as the SBR decreases. Fig. 4 shows the mean-squared errors (MSEs) of the reconstructions using both algorithms. We also show results of a second experiment of this type where the background measurements were collected in particularly adversarial environments (outside with snow, and inside with stained glass windows) in Fig. 5.

Furthermore, we compare the efficacy of the proposed method to the algorithm from [15] by reconstructing a hidden image from experimental data in the presence of ambient light. In the Supplementary Material of [15], an experiment was included where reconstructions were formed of a hidden image for decreasing SBR, achieved by slowly increasing the ambient light level in the room. In this experiment, depicted in Fig. 1, a hidden image was displayed on a 0.305 m × 0.404 m computer monitor that was approximately 1 m from the visible relay wall. Between the hidden monitor and the relay was a chair model which occludes some light paths. In Fig. 6, we show the resulting

1. Available at: https://github.com/Computational-Periscopy/Ordinary-Camera

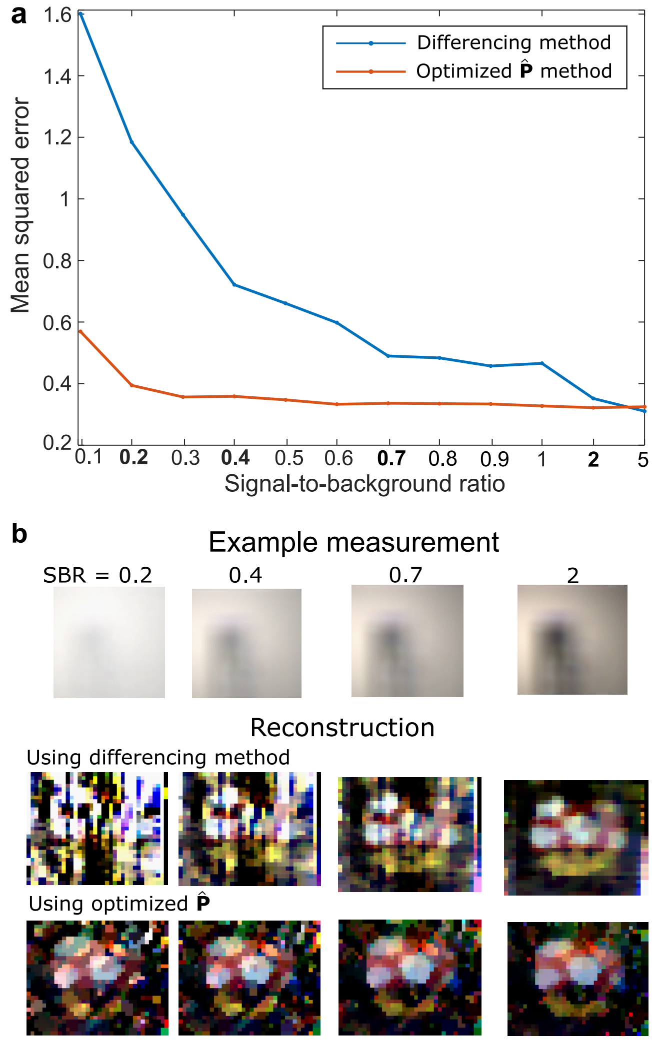

Fig. 4. **a.** A plot of reconstruction MSE against SBR. Experimental background measurements (32 in total, 1 used per trial) were added to a background-free experimental measurement, to achieve a fixed SBR. **b.** An example of the generated measurements followed by reconstructions using the method in [15] and the proposed method using the optimized preconditioner $\hat{\mathbf{P}}$, respectively.

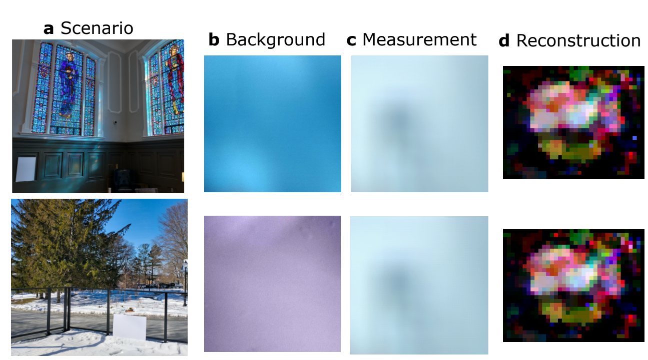

Fig. 5. **a** Scenario in which background was measured (see white posterboard). Top: Stained glass windows. Bottom: Direct sunlight and snow reflections. **b** Background measurement (scaled from 0-1 for display only). **c** Composite measurement (SBR = 0.25). **d** Reconstruction.

reconstructions using the proposed algorithm compared with the original results from [15], and show the computation time for each method (excluding the time taken to generate the light transport matrix $\mathbf{A}$, which is common to each method). This matrix is computed column-by-column, calculating the light contribution (including shadows cast by the chair-shaped object) from each scene pixel one-by-one using available code[1].

4.2 Video Recovery

Inspired by the speed of the proposed algorithm, we now aim to recover a sequence of images, or video, in the hidden area using the procedure outlined in Sec. 3.2. To do so, we simulate the measurement due to each frame of the video by left multiplying the vectorized ground truth frames by our transport model $\mathbf{A}$ (for the scene configuration depicted in Fig. 1). We then augment this measurement with one of the experimentally measured backgrounds to achieve an SBR of 0.5, and finally we add white Gaussian noise to achieve a signal-to-noise ratio of 60 dB, which is fairly typical of the experimental measurements. The first frame is recovered with the algorithm proposed for still images, and the subsequent frames are recovered using the proposed video approach. The results of this are shown in Fig. 7.

5 Discussion

5.1 Tolerating Background

We see from the experimental reconstruction results in Fig. 6 that the algorithm from [15] begins to struggle to form acceptable reconstructions at SBR lower than 4. However, our proposed method suffers minimal degradation even down to SBR of 0.95, and performs similarly at higher SBR. Furthermore, the proposed algorithm does not require the costly projected gradient descent algorithm with a poorly conditioned $\mathbf{A}$ as in [15]. Hence, the computation time is improved by a factor of $\sim 80\times$ when averaged over the four reconstructions. If the post-processing time taken to form $\widehat{\mathbf{P}}$ is discounted (as it is only needs to be calculated once for a particular scenario), the computation time for one reconstruction is reduced by a factor of $3125\times$ (0.024 s vs. 75 s), which approaches video frame rates (around 40 fps). In Fig. 4(a), we see that the reconstruction from the proposed algorithm is not significantly affected by the background level, performing similarly at SBRs down to as low as 0.2. As can be seen in the example measurements (Fig. 4(b)), this is a realistic SBR that one may expect to see in a variety of scenarios – with only faint penumbrae present in the measurement.

To better analyse the effect of left-multiplying by $\widehat{\mathbf{P}}$, we show in Fig. 8 a comparison of a measurement with SBR of 0.25, the same measurement but background-free, and $\mathbf{A}\widehat{\mathbf{P}}\mathbf{y}$. We see that $\mathbf{A}\widehat{\mathbf{P}}\mathbf{y}$ appears markedly similar to the background-free measurement, which suggests the $\widehat{\mathbf{P}}$ operator is successfully cancelling out the background contribution whilst retaining the information important to forming a reconstruction, i.e., the penumbrae. In Fig. 9 we compare two reconstructions to highlight the importance of the background model. The first uses the procedure as outlined in Fig. 1, and the second follows the same procedure

TABLE 1
Time taken in seconds to form reconstructions of different resolutions using the method in Saunders et al. [15] and the proposed method. Forming $\widehat{\mathbf{P}}$ is only performed once for a particular scene configuration.

Resolution	Saunders et al. [15]	Proposed method	
		Forming $\widehat{\mathbf{P}}$	Reconstruction
36×29	68.755	4.621	0.019
54×43	145.099	19.357	0.042
72×58	198.789	59.484	0.082
90×72	262.255	157.473	0.131

but in place of $\hat{\mathbf{P}}$ uses an inverse of $\mathbf{A}$ formed by retaining only the top 40% singular vectors, to help regularize the solution. The data is augmented with a background measured in the presence of stained glass windows, as in Fig. 5, with SBR of 0.5. Without using the proposed inverse, the reconstruction fails.

In Fig. 10 we show the $\mathbf{P}$ calculated given the scenario used in all of the presented results. Interestingly, we see structure that is reminiscent of the finite differencing matrix employed in previous work [15].

5.2 Computational Complexity

In Table 1, the time taken to form reconstructions of various resolutions using either the method from [15] or the proposed method is shown. The proposed method performs significantly faster, especially when one considers forming multiple reconstructions in the same scene configuration (e.g. a video), as the matrix $\widehat{\mathbf{P}}$ need only be constructed once for a particular scenario. At higher reconstruction resolution, the time to form $\widehat{\mathbf{P}}$ increases quite significantly. At much higher resolutions, it may be computationally infeasible to calculate the pseudoinverse directly, and instead iterative methods may be required instead. The actual image reconstruction, however, is still extremely fast compared to the method from [15]. At high reconstruction resolutions, simply storing $\widehat{\mathbf{P}}$ in memory may also be challenging. This could be a situation where aiming to learn a *sparse* pseudoinverse could be of interest, in order to reduce the storage requirements. Literature on generalized pseudoinverses details how this can be achieved [26].

5.3 Video Reconstruction

In Fig. 7 we demonstrate the use of the video recovery procedure outlined in Sec. 3.2 on simulated measurements. Despite high background levels (SBR = 0.5) and higher resolution (55×40 pixels) we are able to form the reconstructions at a rate of 16 frames per second. We see that later frames are recovered with superior MSE than the first two frames. This is due to the extra regularization we are able to use in the video recovery: the first frame is recovered with the still-image algorithm, and the second and subsequent frames improve with temporal regularization, with the influence of the (lower quality) first frame diminishing. This can be thought of as a short 'burn in' period in which the reconstructions improve before reaching a steady state.

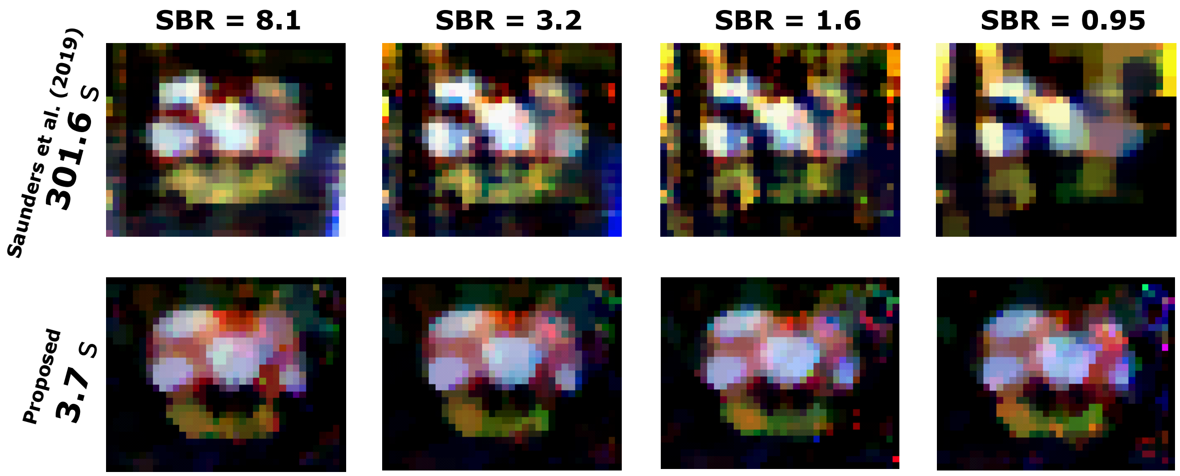

Fig. 6. **Reconstruction results with decreasing SBR.** In the top show, we show state-of-the-art experimental results from the Supplementary Material of [15]. In the second row are reconstructions formed using the proposed method. Also included is the run time to form all four reconstructions using each method. The 3.7 s in the second row includes the pre-processing step of forming $\widehat{\mathbf{P}}$ – following this, the mean time per reconstruction is 0.024 s.

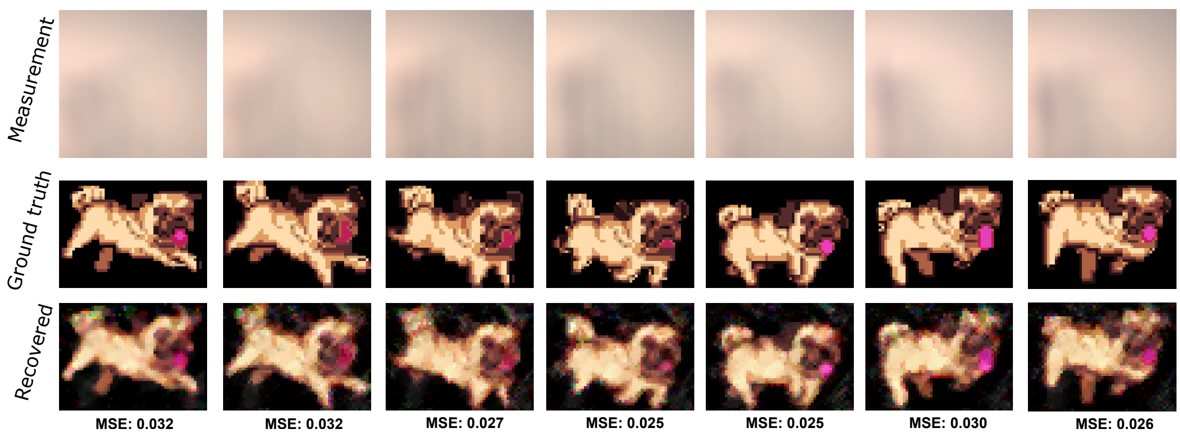

Fig. 7. **Simulated video reconstruction results at SBR = 0.5.** The first row shows the measurement. The second row shows the ground truth video frames. The bottom row shows the recovered 55×40 pixel video frames, at a rate of 16 frames per second. The mean square error (MSE) of each reconstruction is listed beneath.

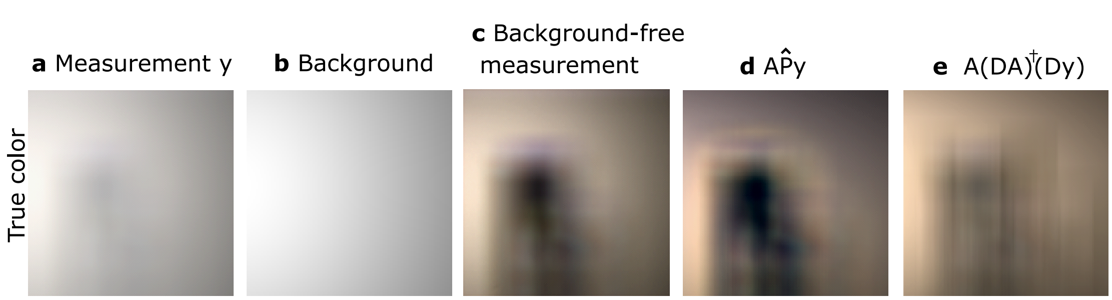

Fig. 8. **Measurement compared to background-free measurement and $\mathbf{A}\widehat{\mathbf{P}}\mathbf{y}$. a.** Measurement $\mathbf{y}$. **b.** The background contribution to the measurement. **c.** Measurement without background contribution. **d.** The result of $\mathbf{A}\widehat{\mathbf{P}}\mathbf{y}$. **e.** The result of pseudoinverse with differencing, $\mathbf{A}(\mathbf{DA})^{\dagger}\mathbf{Dy}$.

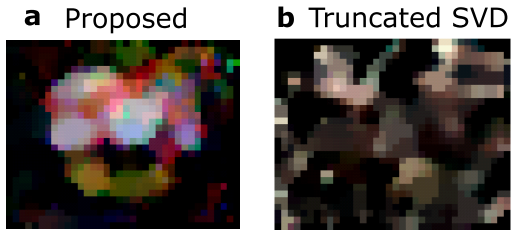

Fig. 9. **Comparing inverses**. **a.** Proposed, background cancelling inverse. **b.** Truncated SVD inverse. Both reconstructions using experimental data augmented with stained glass background in Fig. 5.

6 Conclusion

We have proposed a new algorithm that enables passive non-line-of-sight imaging using a photograph from an ordinary digital camera in high ambient light scenarios. The proposed algorithm significantly improves upon previous results in terms of both tolerance to background light levels and also reconstruction speed. Further investigation into improved post-repair algorithms could improve the reconstructions further.

Exploring the use of generalized pseudoinverses in NLOS imaging is an intriguing avenue with many possibilities, especially in the context of improving robustness to background contributions and in reducing reconstruction time. Doing so overcomes a major hurdle in making passive NLOS practical, without making sacrifices such as imaging only objects in motion, requiring pre-calibrated background measurements, or requiring lengthy offline recovery procedures. We hope this may inspire similar improvements to other NLOS imaging modalities and further exploration of the application of variations of generalized pseudoinverses.

Acknowledgments

This work was supported in part by the US Defense Advanced Research Projects Agency (DARPA) REVEAL Program under contract number HR0011-16-C-0030 and the US National Science Foundation under grant number 1955219.

References

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

[2] 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 Commun.*, vol. 3, p. 745, 2012.

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

[4] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, "Non-line-of-sight imaging using a time-gated single photon avalanche diode," *Opt. Express*, vol. 23, no. 16, pp. 20 997–21 011, 2015.

[5] C. Thrampoulidis, G. Shulkind, F. Xu, W. T. Freeman, J. H. Shapiro, A. Torralba, F. N. C. Wong, and G. W. Wornell, "Exploiting occlusion in non-line-of-sight active imaging," *IEEE Trans. Computational Imaging*, vol. 4, no. 3, pp. 419–431, Sep. 2018.

[6] J. Rapp, C. Saunders, J. Tachella, J. Murray-Bruce, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, R. M. A. Dawson, F. N. C. Wong, and V. K. Goyal, "Seeing around corners with edge-resolved transient imaging," *Nature Commun.*, vol. 11, no. 1, p. 5929, Nov. 2020.

[7] S. Chan, R. E. Warburton, G. Gariepy, J. Leach, and D. Faccio, "Non-line-of-sight tracking of people at long range," *Opt. Express*, vol. 25, no. 9, pp. 10 109–10 117, 2017.

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

[9] V. Arellano, D. Gutierrez, and A. Jarabo, "Fast back-projection for non-line of sight reconstruction," *Opt. Express*, vol. 25, no. 10, pp. 783–792, 2017.

[10] B. Ahn, A. Dave, A. Veeraraghavan, I. Gkioulekas, and A. Sankaranarayanan, "Convolutional approximations to the general non-line-of-sight imaging operator," in *Proc. IEEE Int. Conf. Computer Vision*, Oct. 2019, pp. 7888–7898.

[11] D. B. Lindell, G. Wetzstein, and M. O'Toole, "Wave-based non-line-of-sight imaging using fast f-k migration," *ACM Trans. Graph*, vol. 38, no. 4, p. 13, 2019.

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

[13] C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Baraniuk, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," *Optica*, vol. 7, no. 1, pp. 63–71, Jan. 2020.

[14] D. Faccio, A. Velten, and G. Wetzstein, "Non-line-of-sight imaging," *Nature Reviews Physics*, vol. 2, no. 6, pp. 318–327, Jun. 2020.

[15] C. Saunders, J. Murray-Bruce, and V. K. Goyal, "Computational periscopy with an ordinary digital camera," *Nature*, vol. 565, no. 7740, pp. 472–475, 2019.

[16] S. W. Seidel, Y. Ma, J. Murray-Bruce, C. Saunders, W. T. Freeman, C. C. Yu, and V. K. Goyal, "Corner occluder computational periscopy: Estimating a hidden scene from a single photograph," in *Proc. IEEE Int. Conf. Computational Photography*, May 2019, pp. 25–33.

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

[18] S. W. Seidel, J. Murray-Bruce, Y. Ma, C. Yu, W. T. Freeman, and V. K. Goyal, "Two-dimensional non-line-of-sight scene estimation from a single edge occluder," *IEEE Trans. Computational Imaging*, vol. 7, pp. 58–72, 2021.

[19] M. Batarseh, S. Sukhov, Z. Shen, H. Gemar, R. Rezvani, and A. Dogariu, "Passive sensing around the corner using spatial coherence," *Nature Commun.*, vol. 9, no. 1, p. 3629, Sep. 2018.

[20] J. Boger-Lombard and O. Katz, "Passive optical time-of-flight for non line-of-sight localization," *Nature Commun.*, vol. 10, no. 1, p. 3343, Jul. 2019.

[21] A. B. Yedidia, M. Baradad, C. Thrampoulidis, W. T. Freeman, and G. W. Wornell, "Using unknown occluders to recover hidden scenes," in *IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 12 223–12 231.

[22] M. Aittala, P. Sharma, L. Murmann, A. Yedidia, G. Wornell, B. Freeman, and F. Durand, "Computational mirrors: Blind inverse light transport by deep matrix factorization," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, 2019, pp. 14 311–14 321.

[23] C. Saunders, R. Bose, J. Murray-Bruce, and V. K. Goyal, "Multi-depth computational periscopy with an ordinary camera," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, May 2020, pp. 9299–9309.

[24] J. Klein, C. Peters, J. Martín, M. Laurenzis, and M. B. Hullin, "Tracking objects outside the line of sight using 2D intensity images," *Scientific Reports*, vol. 6, no. 1, p. 32491, Oct. 2016.

[25] I. Dokmanić and R. Gribonval, "Beyond Moore-Penrose part I: Generalized inverses that minimize matrix norms," Jul. 2017, working paper or preprint. [Online]. Available: https://hal.inria.fr/hal-01547183

[26] I. Dokmanić, M. Kolundžija, and M. Vetterli, "Beyond Moore-Penrose: Sparse pseudoinverse," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2013, pp. 6526–6530.

[27] J. Murray-Bruce, C. Saunders, and V. K. Goyal, "Occlusion-based computational periscopy with consumer cameras," in *SPIE Wavelets and Sparsity XVIII*, D. Van De Ville, M. Papadakis, and Y. M. Lu, Eds., vol. 11138, 2019, pp. 287–298.

[28] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

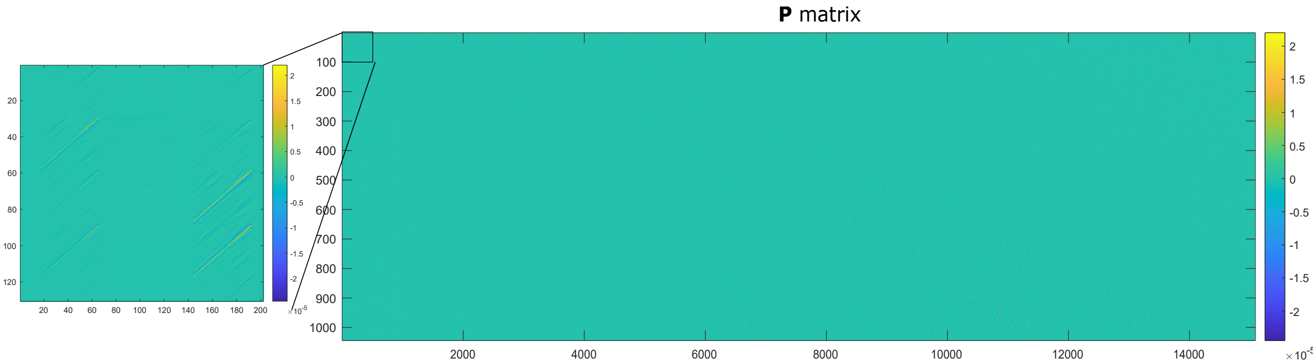

Fig. 10. The $\mathbf{P}$ matrix found for the imaging scenario considered in this paper.

[29] M. K. Ng, P. Weiss, and X. Yuan, "Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods," *SIAM J. Scientific Computing*, vol. 32, no. 5, pp. 2710–2736, 2010.

[30] L. Condat, "A generic proximal algorithm for convex optimization—application to total variation minimization," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 985–989, 2014.

[31] R. Chanu and K. M. Singh, "Vector median filters: A survey," in *Int. J. Computer Science and Network Security*, vol. 16, 2016, pp. 66–84.

[32] S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, "An augmented Lagrangian method for total variation video restoration," *IEEE Trans. Image Processing*, vol. 20, no. 11, pp. 3097–3111, 2011.

[33] A. Björck and G. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of Computation*, vol. 27, no. 123, pp. 579–594, Jul. 1973.

Charles Saunders received the M.Eng. degree in Acoustical Engineering from the University of Southampton, UK in 2016. He is currently a PhD student in Electrical Engineering at Boston University. His research interests include computational imaging (especially NLOS imaging) and optimization.

Vivek K Goyal (S'92–M'98–SM'03–F'14) received the B.S. degree in mathematics and the B.S.E. degree in electrical engineering from the University of Iowa, where he received the John Briggs Memorial Award for the top undergraduate across all colleges. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, where he received the Eliahu Jury Award for outstanding achievement in systems, communications, control, or signal processing.

He was a Member of Technical Staff in the Mathematics of Communications Research Department of Bell Laboratories, Lucent Technologies, 1998–2001; and a Senior Research Engineer for Digital Fountain, Inc., 2001–2003. He was with the Massachusetts Institute of Technology 2004–2013, where he was the Esther and Harold E. Edgerton Associate Professor of Electrical Engineering. He was an adviser to 3dim Tech, Inc. (winner of the 2013 MIT $100K Entrepreneurship Competition Launch Contest Grand Prize), and was subsequently with Nest, an Alphabet company, 2014–2017. He is now with the Department of Electrical and Computer Engineering of Boston University. His research interests include computational imaging, human perception, decision making, sampling, quantization, and source coding theory.

Dr. Goyal is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, Eta Kappa Nu, OSA, SIAM, and AAAS. He was awarded the 2002 IEEE Signal Processing Society (SPS) Magazine Award, the 2014 IEEE Int. Conf. Image Processing Best Paper Award, the 2017 IEEE SPS Best Paper Award, the 2019 IEEE SPS Best Paper Award, and an NSF CAREER Award. Work he supervised won student best paper awards at the IEEE Data Compression Conf. in 2006 and 2011, the IEEE Sensor Array and Multichannel Signal Processing Workshop in 2012, and the IEEE Int. Conf. Image Processing 2018, as well as the 2020 IEEE SPS Young Author Best Paper Award and seven thesis awards. He served on the IEEE SPS Image and Multiple Dimensional Signal Processing Technical Committee (TC), the SPS Image, Video, and Multidimensional Signal Processing TC, the Steering Committee of the IEEE Transactions on Multimedia, the Scientific Advisory Board of the Banff International Research Station for Mathematical Innovation and Discovery, and as an IEEE SPS Distinguished Lecturer. He current serves on the Editorial Boards of IEEE Transactions on Computational Imaging, IEEE Open Journal of Signal Processing, and *Foundations and Trends in Signal Processing*. He was a Technical Program Co-chair of the 11th Int. Conf. Sampling Theory and Applications and a Co-chair of the SPIE Wavelets and Sparsity conference series 2006–2016. He is a Fellow of the OSA and a co-author of *Foundations of Signal Processing* (Cambridge University Press, 2014).