Stereo Non-Line-of-Sight Imaging

Pablo Luesia-Lahoz[1*], Sergio Cartiel[1] and Adolfo Muñoz[1]

Universidad de Zaragoza - I3A.

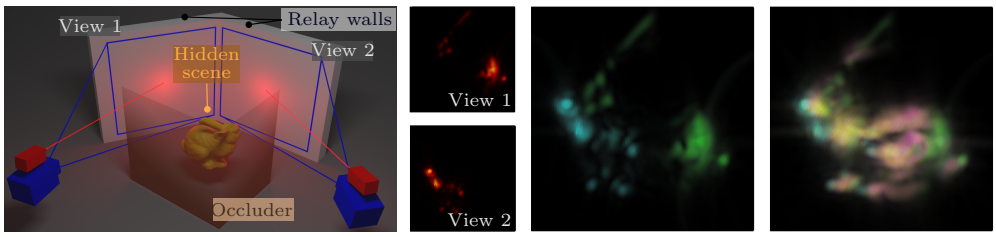
*Corresponding author(s). E-mail(s): pluesia@unizar.es;
Contributing authors: scartiel@unizar.es; adolfo@unizar.es;

Abstract

Transient non-line-of-sight imaging techniques reconstruct hidden scenes by analyzing the time of flight of light scattered off a visible secondary surface, or relay wall. Despite many promising approaches, all face the inherent problem of the missing cone, which restricts surface visibility based on their position and orientations relative to the relay wall. Drawing inspiration from stereo technologies from computer vision, we devise a setup consisting of two distinct relay walls. We leverage Phasor Fields that computationally model both relay walls as generalized virtual camera apertures. This approach allows us to combine the contributions from each relay wall, including the signal obtained by illuminating one wall and capturing the other, information that would be lost otherwise. Our results demonstrate that our proposal diminishes the effect of the missing cone by making the problem better posed. Additionally, by analyzing the visibility conditions of the missing cone, we extract orientation cues from each relay wall contribution. We use this information to enhance visualizations.

Keywords: Non-line-of-sight imaging, computational photography, stereo imaging, looking around corners

Acknowledgements. We thank the people from the Graphics and Imaging Lab for their support, and especially Jorge García, Santiago Jiménez and Julio Marco for their assistance. This work was funded by HORIZON EUROPE Research and Innovation Actions (101070310), and Ministerio de Ciencia, Innovación y Universidades/Agencia Estatal de Investigación/10.13039/501100011033 (Project PID2019-105004GB-I00).



a) Our setup. b) Independent reconstructions. c) Combination d) Ours

Fig. 1 We present a novel stereo configuration for non-line-of-sight imaging using the time of flight of the light. Two relay walls illuminate and capture a hidden scene (a). The missing cone limits the visibility in the independent reconstructions, regardless of the imaging method employed. Thus, we observe limited visibility when applying the Phasor Fields framework in independent reconstructions (b). Our proposal poses a combination of Phasor Fields virtual cameras of two views as a generalized single camera (c), and we can employ the limited visibility to extract orientation cues of the hidden surface by color coding the contribution from each wall. Moreover, our setup allows us to gain additional illumination and capture signals between relay walls, increasing the reconstruction's overall visibility (d).

1 Introduction.

With current technology, scenes can be captured at a frame rate of trillions per second, breaking the classical assumption of infinite speed-of-light and opening the field of transient imaging [1, 2]. One of the most promising applications of this technology is non-line-of-sight (NLOS) imaging.

NLOS imaging techniques use the temporal profile of light reflected at a secondary surface or relay wall (RW) to reconstruct partially or fully occluded scenes around corners. A device illuminates the RW with a short pulse of nanoseconds, which scatters light into the hidden scene. Part of this scattered light is then scattered back into the surface which is captured by a time-resolved sensor.

Numerous imaging methods exist to invert the propagation of light and obtain a reconstruction of the hidden scene from the captured signal: back-projection [3, 4], light transport inversion [5–9] or wave-based models [10–12] are some of the current solutions for this problem.

The Phasor Fields framework [10] is particularly interesting as it provides an analogy between the Fourier transform of the temporal profile of light transport and well-established wave-optics concepts. Specifically, it poses the RW as a *virtual camera* on a computational realm, which therefore transforms the real-domain NLOS problem into a computational-domain line-of-sight (LOS) one, making it more tractable and enabling accurate reconstructions.

However, besides its promising results, Phasor Fields and other NLOS imaging methods have limitations regarding their ability to reconstruct surfaces at specific locations and orientations. The problem is ill-posed and presents uncertainties and occlusions making state-of-the-art approaches unable to reconstruct specific geometries

in the scene, even if some signal is captured. This limitation is often called the missing-cone problem and it is ubiquitous in NLOS measurements (we refer the interested reader to previous works [13]).

Following the analogy between the RW and a virtual camera, in this work, we explore the possibility of using classical LOS techniques from computer vision to reconstruct the geometry of the hidden scene. In particular, we analyze stereo vision. For that purpose, we devise scenes with not one, but two RWs, in which each of those walls acts as a virtual camera in the computational domain, and apply a stereo-vision-based reconstruction technique (see Figure 1). We demonstrate that multiple captures from different RWs can be combined under Phasor Fields into a single and generalized optical system. This approach leads to a better-posed NLOS problem, where we capture the light scattered from one and reaching the other, besides their independent capture, which would be lost otherwise. As we show, our approach not only retrieves geometry but is also able to enhance reconstruction positions and orientations when compared to any of the individual RWs, even in real-world captures.

Classical stereo-vision algorithms often require the identification of common feature points on both views. However, the missing cone limits the visibility of the NLOS reconstructions, as it occurs for specific positions and orientations of surfaces with respect to the RW [13]. Two different RWs will more likely have non-overlapping missing-cone regions because the same geometrical features of the hidden object have different orientations with respect to each of the individual RWs. We therefore cannot leverage information about common features from both perspectives (as given by the two RWs) because the missing cone leads to finding very few of them in both virtual cameras.

In contrast, thanks to the knowledge about the missing cone, the fact that a particular surface is visible from one of the RWs is actually giving some cues about the orientation of such surface (otherwise it would not be visible). We leverage these missing-cone orientation cues as a feature, instead of a limitation. We label the reconstructed surfaces according to their orientations, following such orientation cues, giving a more detailed image of the reconstruction with additional orientation information.

Our approach combines different views into a single reconstruction, enhancing the visibility of the reconstructed hidden scene and gaining orientation information of the surfaces, as our results show. Hence, this work offers a step towards a generalized model of extended non-planar apertures in the NLOS imaging scenario.

2 Related Work.

2.1 NLOS imaging.

NLOS imaging aims to reconstruct target hidden objects, that are partially or fully occluded from the direct line-of-sight, and it relies on the indirect illumination scattered one or multiple times into visible surfaces. A possible categorization is passive and active methods. While passive methods rely on the existing illumination of the scene (i.e., present or emitted by the hidden scene) [14–16], active NLOS methods use controlled light sources to illuminate the hidden scene.

Transient NLOS is an active NLOS method that leverages the capability to capture frames at trillions of frames per second by using ultra-fast capture devices [1, 17, 18] and short laser pulses. Under these circumstances, it is possible to measure the time of flight (ToF) of the photons that scatter from the hidden surfaces. It was originally proposed by Kirmany et al. [19, 20]. The problem has similarities with computed tomography problems from medical imaging and thus is related to an elliptical Radon integral. Hence, the first high-quality 3D reconstructions employed filtered ellipsoidal back projection (FBP) [3, 4], which motivated alternatives like error FBP [21] or Laplacian and Gaussian FBP [22].

Later works demonstrated that by restricting the data acquisition to a confocal configuration, i.e. collocating the sensed and illuminated points, NLOS imaging can be formulated as a deconvolution-based linear inverse problem [6, 8, 23] which is efficiently solved in the frequency domain [11].

Our work lies within the domain of transient NLOS (for simplicity just NLOS from now on). We employ Phasor Fields [10], a frequency-domain method, that does not require a confocal configuration, that makes use of well-known Fourier optics to invert the propagation of the light. Next, we describe it.

2.2 Phasor Fields Formulation.

Phasor Fields is a formulation that proposes a wave-based model to pose the NLOS as a *virtual* LOS problem [10, 24]. Classic NLOS methods develop ad-hoc strategies to reconstruct from the indirect light scattered by the hidden surfaces. In contrast, Phasor Fields can make use of well-known Fourier-space operators from traditional optics to image the hidden scene, enclosing either confocal or non-confocal capture setups.

In recent works, efficient implementations in the frequency domain exist for parallel planes [12], which has led to real-time reconstructions of hidden scenes [25]. Furthermore, research has shown that the phasor fields formulation can account for complex light interactions within the hidden scene, such as distinguishing between direct and indirect illumination [26], or allowing for beyond-the-third bounce reconstruction employing the mirror-like behavior in the virtual domain of Phasor Fields [27].

Phasor Fields has demonstrated great flexibility on the capture surfaces, showing excellent results with planar and non-planar surfaces of capture, being static or even dynamic [28, 29], or even to image through scattering media [30]. Yet, later works rely on planar projections from the non-planar RWs, limiting the results to a single aperture. In this work, we employ the Phasor Fields framework to image the hidden scene under a novel setup consisting of different apertures. We devise a combination method for multiple RWs in contrast to the single aperture of previous Phasor Fields applications.

2.3 From LOS to NLOS.

One of the most classical technologies in computer vision [31] is stereo vision, whose target consists of mimicking human depth perception using dual vision. The first application of this technology in constructing topographic elevation maps from aerial

images [32] has evolved and is still in use to estimate depth or 3D structure from a scene. Traditional techniques involve block matching and correlation-based to generate disparity maps [33], that are more sensible to occlusions, regions without texture and efficiency [34], and, nowadays approaches, exploit machine and deep learning [35].

We have already introduced the analogy of the RW from the NLOS domain to a virtual LOS camera using Phasor Fields. Our work gets influenced by stereo vision from computer vision to present a stereo setup of NLOS, saving some differences between the two domains later explained in Section 4.

Similar to stereo vision, our setup allows us to obtain more information than two independent captures on their own. Specifically, we are capable to obtain orientation information of the surfaces from the hidden scene. Some researchers have indirectly tackled this problem in the NLOS domain, generating geometry through inverse rendering [36–38] or retrieving an estimation of the normals in reconstruction time [8]. In contrast, we directly target the orientations in NLOS imaging, making use of the missing cone (see Section 3.2).

3 Background.

A prior understanding of the Phasor-Fields formulation [10] is necessary to comprehend our work. In this section, we provide insights into the formulation, together with the definition of the limitations of NLOS methods, and, specifically, the missing cone.

3.1 Phasor Fields: image formation.

Phasor Fields undergo a dual domain nature of real and computational to form the image. It is key to identify the differences between these two domains to understand the pipeline. Figure 2 shows an overview of the whole process.

Real domain. The system employs a paired system of an ultra-fast laser emitter and a time-resolved capture device with a resolution of trillions of frames per second. The illumination consists of an ultra-short pulse of light, considered a delta pulse $\delta(\mathbf{x}_l, t)$ that travels from the laser to a set L of illumination points $\mathbf{x}_l \in L$ in the RW which scatter the light into the hidden scene (Figure 2a). The hidden scene scatters part of the light back to the RW, where the capture device, focused on a grid S of capture points $\mathbf{x}_s \in S$, captures the impulse response $H(\mathbf{x}_l, \mathbf{x}_s, t)$ of the hidden scene (Figure 2b). The RW is in the direct LOS from the sensor and laser. Thus, it is a common practice to shift the impulse response $H(\mathbf{x}_l, \mathbf{x}_s, t)$ in time to discard the time of flight of the light from the laser to the RW and from the RW to the sensor. Thus, from now on, $t = 0$ corresponds to the exact moment the RW scatters the first bounce of emitted light.

Computational domain. In the real world, the light emitted from $\mathbf{x}_l$ is a delta pulse $\delta(\mathbf{x}_l, t)$. Under this assumption, the Phasor Fields formulation can virtually illuminate the hidden scene in the computational domain. Thus, the original authors define an arbitrary input signal $\mathcal{P}(\mathbf{x}_l, t)$ as a virtual illumination temporal profile to yield

$$\mathcal{P}(\mathbf{x}_s, t) = \mathcal{P}(\mathbf{x}_l, t) * H(\mathbf{x}_l, \mathbf{x}_s, t)\, d\mathbf{x}_l, \tag{1}$$

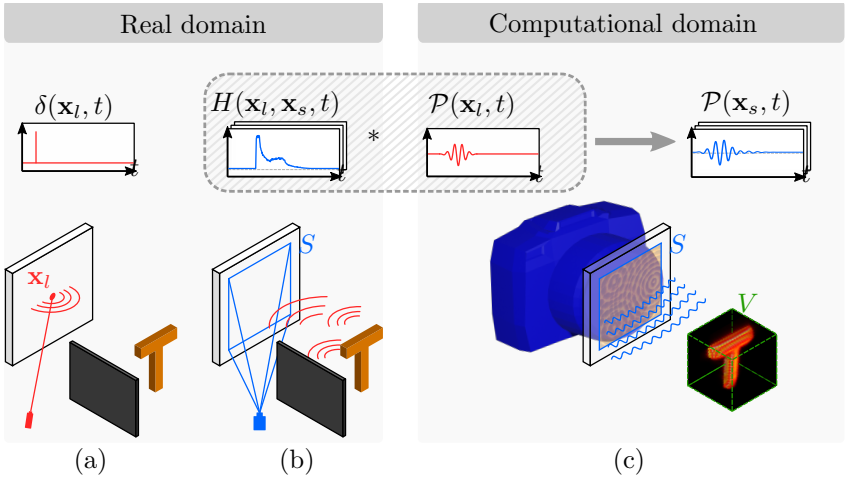


Fig. 2 Phasor Fields formulation overview. In an NLOS setup in the real world (a), a laser emits a delta pulse to a point $\mathbf{x}_l$ in the RW that illuminates the hidden scene. Then, an ultrafast capture device (b) focused at an area S in the RW captures the impulse response $H(\mathbf{x}_l, \mathbf{x}_s, t)$ from the hidden scene. In the computational domain (c), an input signal $\mathcal{P}(\mathbf{x}_l, t)$ serves as a virtual illumination of the hidden scene when convolving with the impulse response $H(\cdot)$. The outcome is a field of phasors $\mathcal{P}(\mathbf{x}_s, t)$ placed at the area S of the RW. The phasor field is equivalent to a planar wavefront, and thus it is possible to use well-known Fourier optics to reconstruct a volume V, posing the RW as a virtual camera. *Figure adapted from Royo et al. 2023 [27].*

being $*$ a convolution in time, and $\mathcal{P}(\mathbf{x}_s, t)$ the captured signal as the hidden scene was virtually illuminated with $\mathcal{P}(\mathbf{x}_l, t)$. Following the original work, the input signal for our work is a pulse wave with a Gaussian envelope with a central wavelength of $\lambda = 6\Delta S$, where ΔS is the minimum distance between the capture points of S, and a standard deviation $\sigma = 6\lambda/\sqrt{2}$.

We define $\hat{O}_\Omega$ as a single component Ω in the frequency-domain, i.e. after a Fourier transform, from a time-domain signal O. Note then that $\hat{\mathcal{P}}_\Omega(\mathbf{x}_s)$ is a phasor. Therefore, considering the whole grid of $\mathbf{x}_s \in S$ in the RW is a phasor field (i.e., a wavefront with frequency Ω), which gives the formulation its name. Then, we represent Equation 1 as

$$\hat{\mathcal{P}}_\Omega(\mathbf{x}_s) = \hat{\mathcal{P}}_\Omega(\mathbf{x}_l)\hat{H}_\Omega(\mathbf{x}_l, \mathbf{x}_s)\, d\mathbf{x}_l. \tag{2}$$

In this computational domain, the original authors interpret the phasor field formed by $\hat{\mathcal{P}}_\Omega(\mathbf{x}_s)$ as out-of-focus illumination captured from the hidden scene. Hence, they can see the original light distribution by bringing that out-of-focus illumination into focus at each specific point and use, under this premise, well-known operators from classic Fourier optics to model the propagation of the virtual wavefronts. We define our imaging operator Φ to compute the in-focus value of a point $\mathbf{x}_v \in V$ of a reconstruction volume V as

$$\hat{I}_\Omega(\mathbf{x}_v) = \Phi\left(\mathbf{x}_v, \hat{\mathcal{P}}_\Omega(\mathbf{x}_s)\right), \tag{3}$$

assuming $\hat{I}_\Omega$ as the image formation function, where Φ is a computational focusing operator equivalent to the lens of a traditional camera. This permitted the original

authors to make the analogy of the RW as a virtual camera, thus posing the NLOS problem as a virtual LOS one.

In their work, the original authors model Φ with a function that resembles the Rayleigh-Sommerfeld Diffraction (RSD) integral:

$$\hat{I}_\Omega(\mathbf{x}_v) = \int_L \int_S \hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}_s)\hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}_l)\hat{\mathcal{P}}_\Omega(\mathbf{x}_s)d\mathbf{x}_s d\mathbf{x}_l, \\ \hat{R}_\Omega(\mathbf{x}_a, \mathbf{x}_b) = \frac{e^{i2\pi\Omega|\mathbf{x}_a - \mathbf{x}_b|/c}}{|\mathbf{x}_a - \mathbf{x}_b|}, \tag{4}$$

where c is the speed of light and $|\cdot|$ is the euclidean distance. The authors named the image formation described by Equation 4 as *confocal camera* [10] (do not mistake with confocal configuration for capture [6]), and it is characterized for bringing into focus both illumination and capture to a point $\mathbf{x}_v$.

3.2 Visibility limitation of NLOS imaging.

The limited size of the RW leans to posing the NLOS imaging as an ill-posed problem. Only a part of the scattered light from the hidden scene is captured, which causes certain features to be impossible to image in the reconstructions, regardless of the imaging method employed [13]. This problem is present in other imaging fields such as medical computed tomography which originally served as a baseline for looking around corners with transient data [3], where it receives the name of the missing cone. In NLOS imaging, the lack of modulation of the signal in the Fourier domain makes it impossible to reconstruct those surfaces in the missing cone, even if some radiance reaches the RW. Under the Phasor Fields scope, the missing cone has also been explored by analyzing near-specular paths of light on the hidden scene [27].

In practice, due to the missing-cone problem, surfaces of the hidden scene can only be reconstructed for a subset of positions and orientations with respect to the RW, which is a general limiting factor for NLOS reconstructions. However, in this work, we factor in that subset of positions and orientations and use them as additional orientation cues. We use such orientation cues to label the reconstructions of the hidden scene, providing additional orientation information that resembles a normal map (see Section 5.1).

4 Our proposal.

Stereo in computer vision algorithms can recreate depth, transforming two different 2D views of the same scene into a 3D representation using line-of-sight (LOS) captures. Inspired by the same idea, we aim to use two different relay walls (RWs). Under Phasor Fields, these two walls can be seen as two cameras that provide two views in the NLOS domain of the same scene. Figure 3 shows the outline of our setup. However, there are two key differences between computer vision and NLOS imaging:

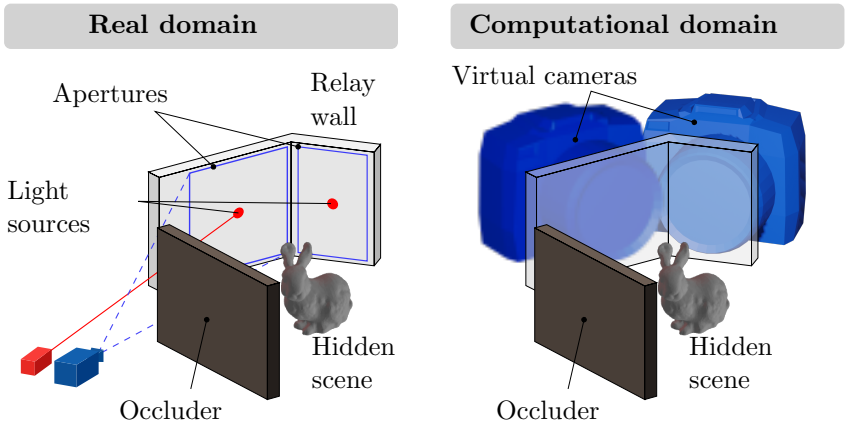


Fig. 3 Our Stereo Aperture setup. Our setup consists of two RWs for illuminating and capturing in the real domain. In the computational domain, our setup is equivalent to placing a camera on each RW. Our approach allows us to combine both cameras as a single generalized one.

1. Our views consist of Phasor-Fields reconstructions, which already contain 3D information about the scene. In contrast with classical LOS imaging, they may contain some errors such as out-of-focus artifacts and limited visibility.
2. The missing-cone problem limits our feature visibility from the RWs in the NLOS domain. That prevents us from finding common feature points, which are crucial for most stereo-vision techniques.

As classical stereo algorithms are not applicable in this NLOS virtual domain, we propose an alternative approach to combine the contribution from different RWs in the same reconstruction space. This combination diminishes the lack-of-visibility effect from the missing cone on individual views if compared to only using independent RWs on their own. However, we also enable the acquisition of the hidden scene's impulse response by illuminating and capturing in different RWs, thereby increasing the recovered information.

The combined aperture resolution contrasts with the LOS multi-view approaches, from which we draw inspiration, where detached aperture information is crucial for gaining geometrical reconstructions. In this section, we prove that different reconstructions can be combined into a generalized single one, where each RWs exhibits different visibility constraints. That leads us to the following statement: the strict constraints of visibility from the missing cone allow us to label surface orientations of the reconstructed objects depending on the contribution from each RW. The missing cone transforms from a constraint to a feature.

4.1 Combination of different Phasor-Fields views.

Under the scenario of K different RWs ($K = 2$ in our experiments), we can define a general impulse response $H(\mathbf{x}'_l, \mathbf{x}'_s, t)$, where $\mathbf{x}'_l \in L'$ and $\mathbf{x}'_s \in S'$. L' and S' represent all illumination and capture points of all RWs. This way we represent the entire impulse

response illuminated and captured by our setup, even from the illumination point of an RW m to a capture point in another one n.

By applying the virtual illumination of Phasor Fields (Equation 1), and transforming it into the frequency domain, we can define a general phasor $\hat{\mathcal{P}}_\Omega(\mathbf{x}_l)$. As a result, we can directly apply the image formation for a single frequency, as shown in Equation 4. However, since the acquired data is not continuous and both the illumination and capture points are discrete due to the current capture technologies, we represent Equation 4 integral as a summation as

$$\hat{I}_\Omega(\mathbf{x}_v) = \sum_{\mathbf{x}'_l \in L'} \sum_{\mathbf{x}'_s \in S'} \hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}'_s) \hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}'_l) \hat{\mathcal{P}}_\Omega(\mathbf{x}'_s), \tag{5}$$

to define the combined image formation $\hat{I}_\Omega(\mathbf{x}_v)$ into the point $\mathbf{x}_v$ of all the RWs.

Recalling Equation 2, we express Equation 5 as

$$\hat{I}_\Omega(\mathbf{x}_v) = \sum_{\mathbf{x}'_l \in L'} \sum_{\mathbf{x}'_s \in S'} \hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}'_s) \hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}'_l) \hat{\mathcal{P}}_\Omega(\mathbf{x}'_l) \hat{H}_\Omega(\mathbf{x}'_l, \mathbf{x}'_s). \tag{6}$$

In this work, the capture points $\mathbf{x}'_s \in S'$ are grouped in grids $\mathbf{x}_s^{(k)} \in S_k$ for each RW k, thus $S_k \subseteq S'$. In contrast, there is only a single illumination point $\mathbf{x}_l^{(k)} \in L'$ at the center of each grid S_k. We define

$$H^{(m,n)}(\mathbf{x}_l^{(i)}, \mathbf{x}_s^{(k)}, t) = H(\mathbf{x}_l^{(j)}, \mathbf{x}_s^{(k)}, t)\delta(m-j)\delta(n-k) \tag{7}$$

as the contribution of our impulse response into pairs of grids S_n and illumination points $\mathbf{x}_l^{(m)}$ of two RWs n and m respectively. Equation 7 allows us to identify two different cases: i) the illumination and capture are both in the same RW ($n = m$), that we will refer to as *independent illumination*, and ii) the illumination and capture are in different RWs ($n \neq m$), which we name *dependent illumination*.

The summation of all pairs of illumination-grid $H^{(i,j)}(\cdot)$ will result in the original generalized capture. Therefore, factoring in Equation 7 in to Equation 6 yields

$$\begin{aligned} \hat{I}_\Omega(\mathbf{x}_v) &= \sum_{m,n \in K} \hat{I}_\Omega^{(m,n)}(\mathbf{x}_v), \\ \hat{I}_\Omega^{(m,n)}(\mathbf{x}_v) &= \sum_{\mathbf{x}'_l \in L'} \sum_{\mathbf{x}'_s \in S'} \hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}'_s) \hat{R}_\Omega(\mathbf{x}_v, \mathbf{x}'_l) \hat{\mathcal{P}}_\Omega(\mathbf{x}'_l) \hat{H}_\Omega^{(m,n)}(\mathbf{x}'_l, \mathbf{x}'_s). \end{aligned} \tag{8}$$

Equation 8 poses how to combine different and independent RWs into a generalized view under Phasor Fields. We leverage the separated contribution by pairs of illumination-capture points in our stereo context to define the image formation from two views as a summation of their reconstructions. Note that we also gain the information of the dependent illumination captures, previously lost.

The direct calculation of Equation 8 may be expensive in time. Therefore, for efficiency's sake, we employ a Phasor Fields implementation optimized for reconstruction

volumes formed by parallel planes to the RW [12]. By leveraging the regular grid of the captures S_k for each independent capture, we define a reconstruction volume for each $\hat{I}_{\Omega}^{(m,n)}(\mathbf{x}_v)$. Later, we project all the reconstruction planes into a common area of matching points $\mathbf{x}_v$ of the reconstruction volume V. As mismatches may occur in the projection, we use a 3D cubic interpolation, whose values are later used for the final summation of Equation 8.

The missing cone establishes a strict visibility condition, which implies that high values in $\hat{I}_{\Omega}^{(m,n)}(\mathbf{x}_v)$ of the RW k correspond to the view of a feature from a surface of the hidden-scene oriented towards that specific RW. Factoring the known information of orientation of the RWs, and the missing cone, the labeling allows us to extract orientation cues of the hidden scene as additional information. We color-code the contribution for each RW as

$$\hat{I}_{\Omega}(\mathbf{x}_v) = \sum_{m,n \in K} c^{(m,n)} \hat{I}_{\Omega}^{(m,n)}(\mathbf{x}_v), \tag{9}$$

where $c^{(m,n)}$ corresponds to the chosen color for the illumination point $\mathbf{x}_l^{(m)}$ and capture grid S_n. As Section 5.1 shows, this coding provides clear hints about hidden surfaces' orientation.

5 Experiments

Our experimental validation demands precise control of the position and rotation of the hidden objects, sensors, illumination, and RWs. Therefore, we opt to synthesize the transient data required to validate our proposal, as it provides sufficient flexibility in managing and controlling the scene, while maintaining the quality and realism to validate our proposal. Previous work [39] has already shown that there is a good agreement between real captures and time-resolved light transport simulations. Specifically, we use a transient renderer optimized for the NLOS configuration [40] to generate the time-resolved responses. The only exception is the experiment represented in Figure 6, which has been generated from measured transient data.

The use of synthetic data enables us to evaluate performance using quantitative metrics derived from the ground truth. Specifically, we report the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) for our synthesized results. PSNR is computed on the reconstructed volume after removing values close to zero in both the ground truth and the reconstruction. SSIM is calculated on the projected 2D images presented in this paper.

To simulate our setup, we place two RWs, one adjacent to the other, at a particular angle (90° in our experiments unless specified otherwise). Each RW has a size of 2×2 meters, with 256×256 capture points distributed in a regular grid, and a single illumination point at the center.

5.1 Stereo combination experiments.

Section 4.1 demonstrates that, under Phasor Fields, we can sum independent reconstructions into a common volume, and this combination behaves as a single optical

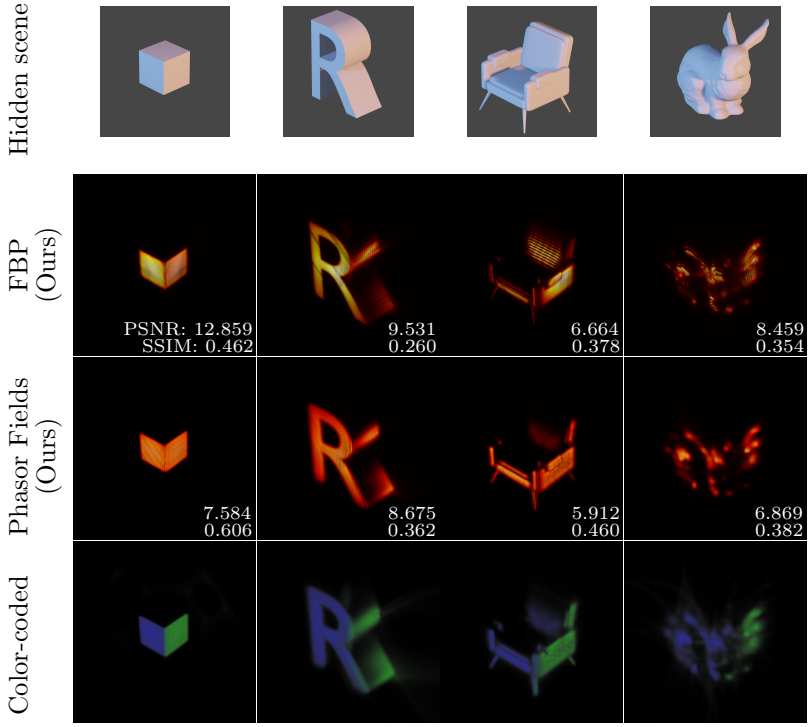


Fig. 4 Reconstructions of two views with independent illumination. From top to bottom: the hidden geometry, the reconstruction with filtered backprojection (FBP), the reconstruction from two views according to Equation 8, and the reconstruction color-coded from each RW (Equation 9) in blue and green. Our combination improves the visual recognition of the hidden geometries when using two independent illuminations, i.e., illuminating and capturing on each RW independently. Furthermore, by simply color-coding by contribution, we can identify the surfaces by their orientation.

system. This mathematical analysis stands for backprojection [3], which is a linear aggregation model. In their original work, Velten et al. filtered the reconstruction by applying a Laplacian filter to increase the visibility of the hidden-scene features, resulting in filtered backprojection (FBP). In this section, we present the FBP approach for experimental validation. We apply the Laplacian filter independently, following the original work, and combine the views as discussed in Section 4.1. We do not provide a mathematical model that proves that this approach can generalize to multiple views, nor devise a specific filter that would work after combining multiple views, both of which would be interesting avenues for future work.

We reconstruct four hidden scenes with our combination method: a cube, a chair, an R letter with volume, and the Stanford bunny. We show the results of using only the *dependent illumination* in Figure 4, and they support our claims for Phasor Fields and FBP. Part of the surfaces that are not visible due to the missing cone from one aperture are visible from the other. The outcome is that, at the combination (second and third rows of Figure 4), the reconstruction increases the overall visibility if compared with any of the individual apertures. Consequently, it is more intuitive to understand the geometry of the hidden object. The bunny is a particularly challenging case due to

the number of surfaces in the missing cone of both apertures. Yet, it is possible to identify the general shape of the bunny, with the face and body, even with missing information (e.g., the bunny ears).

The fourth row of Figure 4 shows the result colored depending on the contribution from each relay wall. We can observe that, by simply color-coding the image, we obtain similarities with a normal map. Furthermore, we observe that this matches the orientations of the visible surfaces from the meshes. This is consistent with our key observation of using the missing cone as a feature, which allows us to label the orientation of a surface knowing the contribution and orientation of the RW that reconstructs it.

The aggregation of the independent reconstructions into the common projection volume is a linear function that incorporates a temporal overhead of $\mathcal{O}(n^3)$, where n is the lateral size of the reconstruction volume. Thus, the complete temporal analysis for reconstructing yields $\mathcal{O}(k(h+n^3)$, being k the number of relay walls, and h the reconstruction time for each independent reconstruction. Existing efficient implementations of Phasor Fields for parallel planes report $h = n^3 \log n$, whereas for FBP this is $h = n^5$. For Figure 4, the Phasor Fields approach takes on average 104 s per reconstruction, while FBP takes 156788 s.

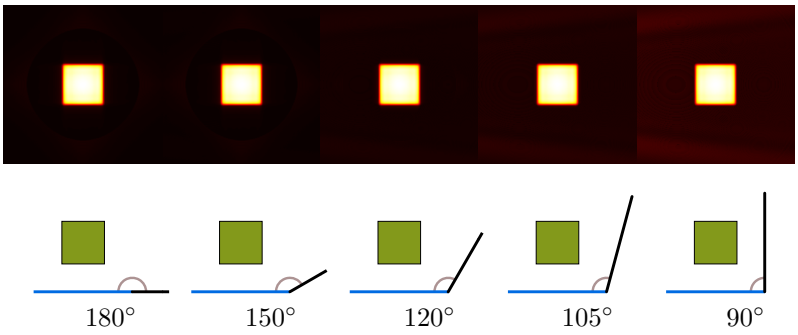


Fig. 5 Visibility under multi-path interference: From left to right: reducing the rotation of a secondary wall to the RW. Top: reconstruction front view. Bottom: schematic top view. The secondary wall adds some background radiance due to multi-path interference when reducing its rotation. Even so, the reconstruction is sharp and clear in all cases, and the noise is negligible compared to the signal.

The incorporation of distinct multiple walls also produces multi-path interference in the areas near the corner. Light scatters between the walls as with the hidden scene, and this radiance is all captured at the RW. Intuitively, this leads to a degradation in the visibility features of the reconstruction as new uncertainties in the signal appear. Yet, our Phasor Fields reconstructions only bring into focus the third bounce of the light (see Section 3.1). The influence from the secondary wall belongs to the RW's missing cone. Thus, the reconstruction does not show any feature or artifact corresponding to this effect, but a low-frequency noise in the background, corresponding to the energy leakage at shorter angles. We show this effect in Figure 5, where we simulated a single reconstruction scenario (with one RW) with a cube as a hidden

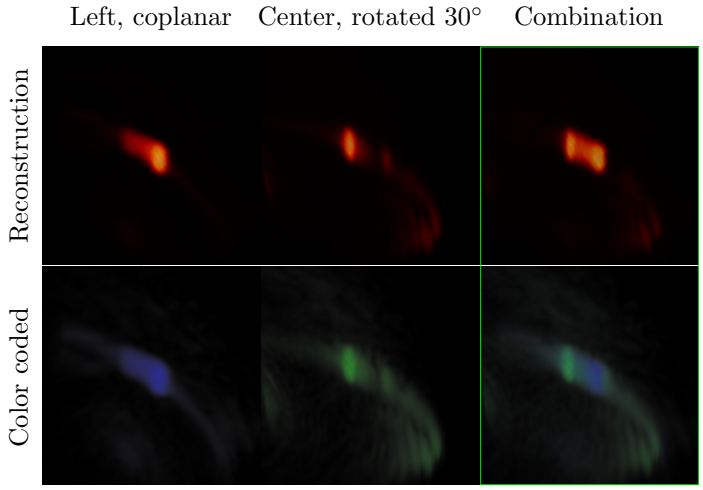


Fig. 6 Real capture combinations. We frame a rotated and translated RW with respect to the scene as an equivalence to our approach. We use different independent illumination captures to increase the visible surfaces of a real captured patch, increasing the identifiable boundaries in our combination.

scene, and a secondary wall at different rotations to the RW. In all the cases, the cube features are clear and sharp, and the background noise is negligible compared to them.

5.2 Combination of captured results.

Our generalized camera model from Equation 8 supports any RW configuration, regardless of rotation or position. So far, we only considered scenarios where both views are simultaneously present at the moment of capture. However, as stated before, the contribution to the reconstruction of multi-path interference is a minor background noise. Consequently, we can establish an equivalent setup using a single physical RW, which, under rotation and displacement between captures, acts as two views of the same hidden scene with independent illumination and without multi-path interference.

In Figure 6, we present both the independent and combined reconstruction results of a patch rotated and displaced between the two real captures. By estimating the relative transformations of the RW with respect to the hidden scene, we set this as a well-suited problem for our approach. As a result, our stereo approach gains complementary visible features from both RWs, compensating from one view the missing cone of the other. Hence, the patch boundaries become more identifiable in the combination, whereas they are not in their individual reconstructions.

5.3 Analysis on the RWs contributions.

Although our approach with independent illumination enhances the visibility of the hidden scene, certain features are not visible in the reconstruction. Yet, our approach enables us to capture the dependent illumination between RWs, otherwise lost, that we leverage as additional captures from the hidden scene.

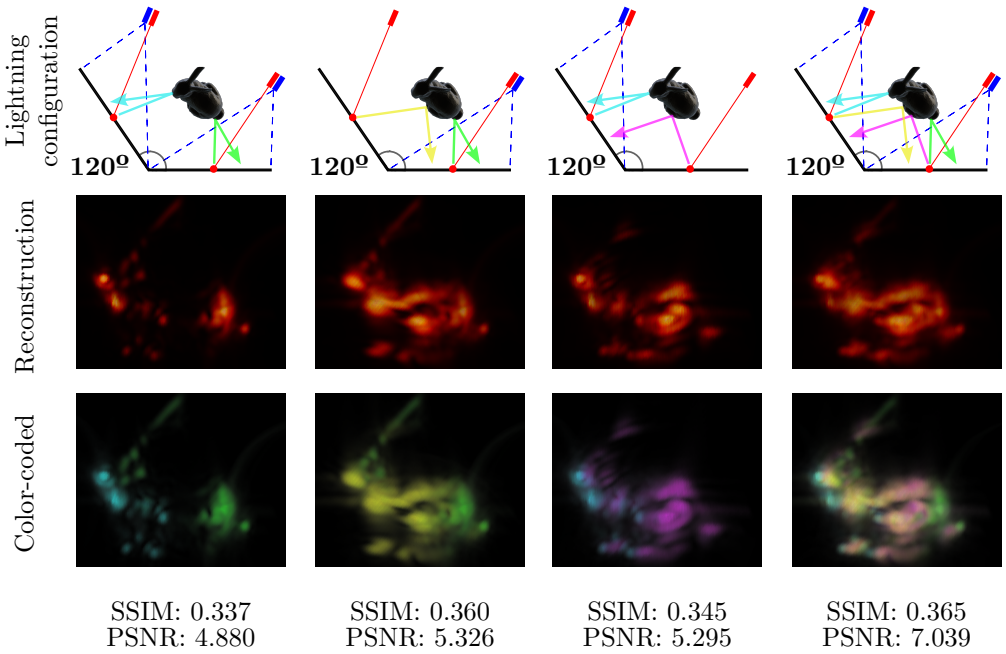


Fig. 7 Bunny scene reconstructions under dependent and independent illumination. Simulated bunny scene with two relay walls (RWs) forming 120° between them. From left to right: reconstruction using only independent illumination, reconstruction capturing at the first RW using both illuminations, reconstruction capturing at the second RW using both illuminations, and reconstruction using all captures and illuminations. In the color-coded (second row), we show the contribution to the visibility from each pair of illumination-capture. Our approach enables us to capture the dependent illumination of one RW reaching the other, otherwise lost, and to increase the overall visibility of the hidden scene.

In Figure 7, we evaluate the impact of dependent illumination in our combination method with the bunny, where features are hidden in our previous experiments, e.g., the ear, the face, and the paws. Using two RWs with a rotation of 120°, surfaces that remain hidden from both views under independent illumination (first column) become visible with the dependent illumination (second and third columns). Our approach using Equation 8 supports the combination of all views (fourth column) to significantly enhance overall visibility when compared to using only independent illumination. The SSIM and PSNR metrics support these claims, achieving maximum value at all views combination. Additionally, these new dependent illumination views can also be color-coded, which provides additional orientation hints.

In Figure 8, we vary the rotation angle between the two RWs to analyze the performance of our combined reconstruction method. The rotation angle influences the contribution of each illumination-capture pair. The overall reconstruction (second row) remains qualitatively similar for all the experiments. Yet, we observe a shift in the separated contributions in the color-coded (third row). Note that the variation from yellow dominant in the experiment of 165°, to pink dominant in the 90° experiment, indicates a variation in the dependent illumination. The causes are the visibility

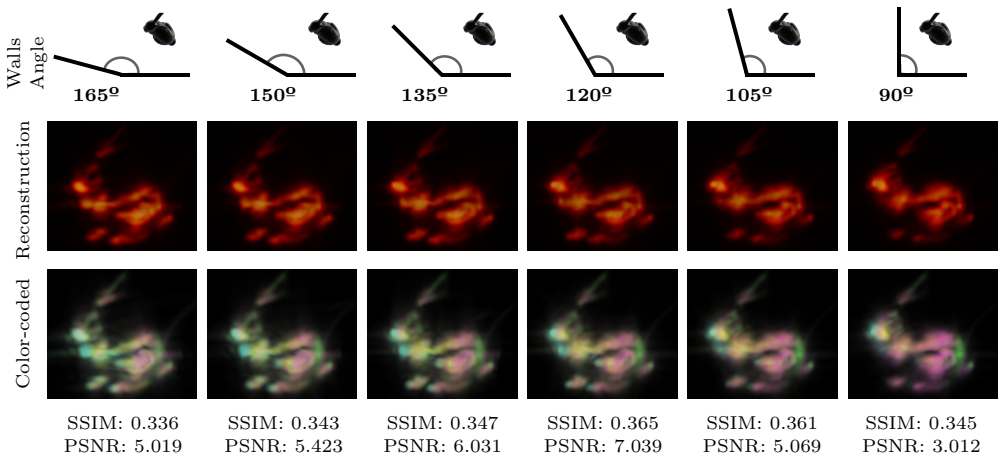


Fig. 8 Bunny scene reconstruction under varying relay wall angles. Simulated bunny scene with two relay walls (RWs) forming different angles between them. (Middle row) The reconstructed scene geometry using the combined illumination and capture data from both RWs. (Bottom row) A color-coded visualization showing the contribution to the final reconstruction from each illumination-capture pair (colors correspond to the scheme in Figure 7).

change of the features in the hidden scene, and distance attenuation from the RW to the geometry. This variation directly impacts the SSIM and PSF of the final reconstruction. The 120° experiment achieves the best metrics because the bunny is aligned with the bisector of this angle. For the analysis on the Phasor Fields behavior under different aperture sizes, we refer the interested reader to previous work [41].

6 Discussion.

Drawing inspiration from the stereo vision technology from computer vision, we introduce a novel configuration consisting of two walls acting as apertures in the NLOS computational domain. This dual-aperture setup increases the amount of captured light reflected on the hidden scene if compared with any of the individual virtual apertures.

Our proposal does come with some limitations, as capturing and processing the additional independent and dependent views requires more acquisition and computational time respectively. Nonetheless, our results demonstrate that our approach can be applied in similar real-world scenarios. Moreover, exploring the interactions between non-independent views increases the overall visibility, effectively quadrupling the available views with two RWs and making the NLOS imaging problem less ill-posed.

As the missing-cone problem generally limits the reconstruction to planar surfaces parallel to the RW, current real captures consist of hidden scenes whose elements are mainly planar and lack volume. Our method, being based on multiple RWs, enables the possibility of obtaining some volume in the hidden scene, with the additional benefit of providing color-coded surface orientations.

Our experiments support our claims. The increased number of apertures helps to retrieve more information from the hidden scene. The direct effect is the decrease in the missing-cone effect and the emergence of more visible features in the reconstruction. We manage to recover further information by exploiting the missing cone to recover surface orientations from the hidden scene. In the future, researchers may delve into sophisticated geometry reconstruction algorithms that benefit from these orientation cues and overcome some current parameterizations of the NLOS imaging methods, i.e., the volume voxelization.

The previous statements come along with several applications for our setup. In the short term, the independence between relay walls enables the use of different devices for capturing. This implies that our approach is highly parallelizable by relay walls, even in concurrent and distributed systems. Further optimizations for Phasor Fields could be applied, e.g., parallelizing by planes, frequencies, or voxels. Other researchers' endeavors, such as the utilization of Phasor Fields with non-planar surfaces [29], complement our work, and together they serve as a step towards completely generalized apertures with arbitrary geometries, which are closer with respect to real-world scenarios, e.g., medical imaging, underground exploration, and car safety. On top of that, there is potential for foster avenues to expand upon our efforts to integrate multiple different apertures as a single composited one to form an extended aperture paradigm. If we take a look at recent advances in the NLOS domain to image beyond the third bounce [27], recursive imaging to see around many corners with the extended aperture paradigm may obtain astonishing results in the NLOS imaging scope.

In conclusion, we have applied a well-known technology, such as stereo vision, in the NLOS imaging scope. The Phasor Fields framework pursues an assumption of a classic optics system, and now our work helps to reduce the gap between more traditional optics and NLOS imaging.

References

[1] Velten, A., Wu, D., Jarabo, A., Masia, B., Barsi, C., Joshi, C., Lawson, E., Bawendi, M., Gutierrez, D., Raskar, R.: Femto-photography: capturing and visualizing the propagation of light. ACM Transactions on Graphics (ToG) **32**(4), 1–8 (2013)

[2] Jarabo, A., Masia, B., Marco, J., Gutierrez, D.: Recent advances in transient imaging: A computer graphics and vision perspective. Visual Informatics **1**(1), 65–79 (2017)

[3] Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M.G., Raskar, R.: Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. Nature communications **3**(1), 1–8 (2012)

[4] Arellano, V., Gutierrez, D., Jarabo, A.: Fast back-projection for non-line of sight reconstruction. Optics express **25**(10), 11574–11583 (2017)

[5] Heide, F., Xiao, L., Heidrich, W., Hullin, M.B.: Diffuse mirrors: 3d reconstruction

from diffuse indirect illumination using inexpensive time-of-flight sensors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3222–3229 (2014)

[6] O'Toole, M., Lindell, D.B., Wetzstein, G.: Confocal non-line-of-sight imaging based on the light-cone transform. Nature **555**(7696), 338–341 (2018)

[7] Heide, F., O'Toole, M., Zang, K., Lindell, D.B., Diamond, S., Wetzstein, G.: Non-line-of-sight imaging with partial occluders and surface normals. ACM Transactions on Graphics (ToG) **38**(3), 1–10 (2019)

[8] Young, S.I., Lindell, D.B., Girod, B., Taubman, D., Wetzstein, G.: Non-line-of-sight surface reconstruction using the directional light-cone transform. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1407–1416 (2020)

[9] Xu, W., Chen, S., Tian, Y., Wang, D., Su, X.: Fast non-line-of-sight imaging based on product-convolution expansions. Optics Letters **47**(18), 4680–4683 (2022)

[10] Liu, X., Guillén, I., La Manna, M., Nam, J.H., Reza, S.A., Huu Le, T., Jarabo, A., Gutierrez, D., Velten, A.: Non-line-of-sight imaging using phasor-field virtual wave optics. Nature **572**(7771), 620–623 (2019)

[11] Lindell, D.B., Wetzstein, G., O'Toole, M.: Wave-based non-line-of-sight imaging using fast fk migration. ACM Transactions on Graphics (ToG) **38**(4), 1–13 (2019)

[12] Liu, X., Bauer, S., Velten, A.: Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems. Nature communications **11**(1), 1–13 (2020)

[13] Liu, X., Bauer, S., Velten, A.: Analysis of feature visibility in non-line-of-sight measurements. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10140–10148 (2019)

[14] Bouman, K.L., Ye, V., Yedidia, A.B., Durand, F., Wornell, G.W., Torralba, A., Freeman, W.T.: Turning corners into cameras: Principles and methods. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2270–2278 (2017)

[15] Saunders, C., Murray-Bruce, J., Goyal, V.K.: Computational periscopy with an ordinary digital camera. Nature **565**(7740), 472–475 (2019)

[16] Krska, W., Seidel, S.W., Saunders, C., Czajkowski, R., Yu, C., Murray-Bruce, J., Goyal, V.: Double your corners, double your fun: the doorway camera. In: 2022 IEEE International Conference on Computational Photography (ICCP), pp. 1–12 (2022). IEEE

[17] Shin, D., Xu, F., Venkatraman, D., Lussana, R., Villa, F., Zappa, F., Goyal, V.K.,

Wong, F.N., Shapiro, J.H.: Photon-efficient imaging with a single-photon camera. Nature communications **7**(1), 12046 (2016)

[18] Buttafava, M., Zeman, J., Tosi, A., Eliceiri, K., Velten, A.: Non-line-of-sight imaging using a time-gated single photon avalanche diode. Optics express **23**(16), 20997–21011 (2015)

[19] Kirmani, A., Hutchison, T., Davis, J., Raskar, R.: Looking around the corner using transient imaging. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 159–166 (2009). IEEE

[20] Kirmani, A., Hutchison, T., Davis, J., Raskar, R.: Looking around the corner using ultrafast transient imaging. International journal of computer vision **95**, 13–28 (2011)

[21] La Manna, M., Kine, F., Breitbach, E., Jackson, J., Sultan, T., Velten, A.: Error backprojection algorithms for non-line-of-sight imaging. IEEE transactions on pattern analysis and machine intelligence **41**(7), 1615–1626 (2018)

[22] Laurenzis, M., Velten, A.: Feature selection and back-projection algorithms for nonline-of-sight laser–gated viewing. Journal of Electronic Imaging **23**(6), 063003–063003 (2014)

[23] Xin, S., Nousias, S., Kutulakos, K.N., Sankaranarayanan, A.C., Narasimhan, S.G., Gkioulekas, I.: A theory of fermat paths for non-line-of-sight shape reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6800–6809 (2019)

[24] Luesia-Lahoz, P., Gutierrez, D., Muñoz, A.: Zone plate virtual lenses for memory-constrained nlos imaging. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5 (2023). IEEE

[25] Nam, J.H., Brandt, E., Bauer, S., Liu, X., Renna, M., Tosi, A., Sifakis, E., Velten, A.: Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second. Nature communications **12**(1), 6526 (2021)

[26] Marco, J., Jarabo, A., Nam, J.H., Liu, X., Cosculluela, M.A., Velten, A., Gutierrez, D.: Virtual light transport matrices for non-line-of-sight imaging. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 2440–2449 (2021)

[27] Royo, D., Sultan, T., Muñoz, A., Masumnia-Bisheh, K., Brandt, E., Gutierrez, D., Velten, A., Marco, J.: Virtual mirrors: Non-line-of-sight imaging beyond the third bounce. ACM Transactions on Graphics **42**(4) (2023) https://doi.org/10.1145/3592429

[28] La Manna, M., Nam, J.-H., Reza, S.A., Velten, A.: Non-line-of-sight-imaging using

dynamic relay surfaces. Optics express **28**(4), 5331–5339 (2020)

[29] Gu, C., Sultan, T., Masumnia-Bisheh, K., Waller, L., Velten, A.: Fast non-line-of-sight imaging with non-planar relay surfaces. In: 2023 IEEE International Conference on Computational Photography (ICCP), pp. 1–12 (2023). IEEE

[30] Luesia, P., Crespo, M., Jarabo, A., Redo-Sanchez, A.: Non-line-of-sight imaging in the presence of scattering media using phasor fields. Optics Letters **47**(15), 3796–3799 (2022)

[31] Szeliski, R.: Computer Vision: Algorithms and Applications. Springer, ??? (2022)

[32] Hannah, M.J.: Computer Matching of Areas in Stereo Images. Stanford University, ??? (1974)

[33] Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision **47**, 7–42 (2002)

[34] Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Transactions on pattern analysis and machine intelligence **30**(2), 328–341 (2007)

[35] Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., Bry, A.: End-to-end learning of geometry and context for deep stereo regression. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 66–75 (2017)

[36] Plack, M., Callenberg, C., Schneider, M., Hullin, M.B.: Fast differentiable transient rendering for non-line-of-sight reconstruction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3067–3076 (2023)

[37] Iseringhausen, J., Hullin, M.B.: Non-line-of-sight reconstruction using efficient transient rendering. ACM Transactions on Graphics (ToG) **39**(1), 1–14 (2020)

[38] Yi, S., Kim, D., Choi, K., Jarabo, A., Gutierrez, D., Kim, M.H.: Differentiable transient rendering. ACM Transactions on Graphics (TOG) **40**(6), 1–11 (2021)

[39] Marco, J., Hernandez, Q., Munoz, A., Dong, Y., Jarabo, A., Kim, M.H., Tong, X., Gutierrez, D.: Deeptof: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. ACM Transactions on Graphics (ToG) **36**(6), 1–12 (2017)

[40] Royo, D., García, J., Muñoz, A., Jarabo, A.: Non-line-of-sight transient rendering. Computers & Graphics **107**, 84–92 (2022)

[41] Guillén, I., Liu, X., Velten, A., Gutierrez, D., Jarabo, A.: On the effect of

reflectance on phasor field non-line-of-sight imaging. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 9269–9273 (2020). IEEE