ZONE PLATE VIRTUAL LENSES FOR MEMORY-CONSTRAINED NLOS IMAGING

Pablo Luesia-Lahoz, Diego Gutierrez, Adolfo Muñoz

Graphics and Imaging Lab, Universidad de Zaragoza - I3A

ABSTRACT

The recently introduced Phasor Fields framework for non-line-of-sight imaging allows to image hidden scenes by treating a relay surface as a virtual camera. It formulates the problem as a diffractive wave propagation, solved by the Rayleigh–Sommerfeld diffraction (RSD) integral. Efficient Phasor Fields implementations employ RSD-based kernels to propagate waves from parallel planes by means of 2D convolutions. However, the kernel storage requisites are prohibitive, hampering the integration of these techniques in memory-constrained devices and applications like car safety.

Instead of relying on expensive RSD kernels, we propose the use of alternative virtual lenses to focus on the incoming phasor field and image hidden scenes. In particular, we propose using zone plates (ZP), which require significant less memory. As our results show, our ZP virtual lenses allow us to obtain reasonable reconstructions of the hidden scene, offering an attractive trade-off for memory-constrained devices.

***Index Terms*—** diffraction lenses, non-line-of-sight imaging, computational imaging, virtual camera

1. INTRODUCTION

Transient imaging deals with the capture and reconstruction of light transport at very high temporal resolutions [1]. One of its most active applications deals with non-line-of-sight (NLOS) imaging of hidden scenes, first introduced by Kirmani et al. [2, 3]. By illuminating a secondary relay surface, it is possible to capture the time-resolved scattered light of an occluded scene onto such surface, and reconstruct, through software signal processing, the hidden scene from the time of flight of the photons (see Figure 1). A variety of approaches have been proposed to solve this problem, including filtered back-projection [4, 5, 6], direct inversion of light transport [7, 8], fermat paths [9], exploiting occlusion [10], light field tomography [11], or wave-based [12, 13, 14] methods, to name a few. We refer the reader to a recent survey for a more complete discussion [15].

This work was funded by MCIN/ AEI /10.13039/501100011033 through Project PID2019-105004GB-I00, by the Government of Aragon's Departamento de Ciencia, Universidad y Sociedad del Conocimiento through project BLINDSIGHT (ref LMP30_21), and by the European Union's European Defense Fund Program through the ENLIGHTEN project under grant agreement No. 101103242.

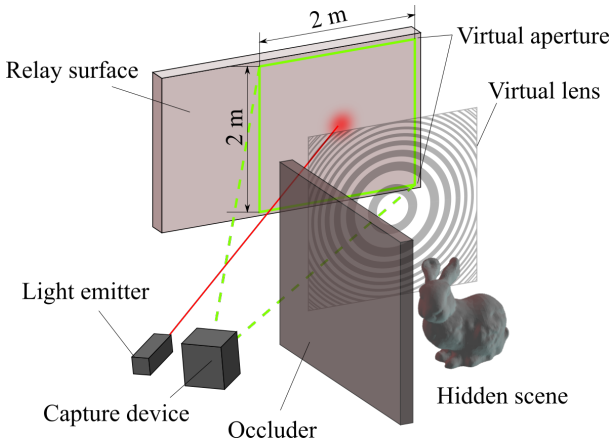

Fig. 1: The NLOS setup. A paired system of pulsed light and SPAD (single-photon avalanche diode) illuminate and capture a relay surface, to obtain the time-of-flight of the light scattered from an occluded hidden scene. In this work, we explore the idea of using virtual lenses based on zone plates to reconstruct the hidden geometry.

The wave-based Phasor Fields method generalizes NLOS imaging by treating the relay surface as a virtual lens, effectively transforming an NLOS problem into a computational representation of a traditional line-of-sight (LOS) system [12]. Under this assumption, the hidden scene can be imaged using well-known Fourier optic tools. In their original work, the authors propose using the Rayleigh-Sommerfeld Diffraction (RSD) integral to model the propagation of light waves in the scene.

Current efficient implementations of this technique rely on RSD-based kernels to simulate this propagation with 2D convolutions [14]. These kernels depend on the parameters of the imaging setup, but not on the transient data (the hidden scene) itself. In other words, it makes sense to pre-compute the kernels for a given imaging setup, even if the objects to be imaged around the corner change. Unfortunately, the required memory to store the kernels scales with the number of Fourier components used for propagation, the resolution of the data, and the reconstruction size. For instance, a 256^2 sampling grid at the relay wall using a single illumination point, with a 256^3 voxel grid for reconstruction and 80 Fourier components in the phasor field requires more than 80 GB of memory. And

even in the case of less complex reconstructions with a single Fourier component, the kernel still needs around 1 GB of storage. This impractical memory consumption is due to the complex values that the RSD kernel requires, which increase storage compared to floating precision or integers values.

To overcome this problem, we investigate the use of alternative lenses in the virtual camera of Phasor Fields approach, to perform the propagation of the light wave. We specifically propose the use of zone plates (ZP) as virtual lenses, which use diffraction instead of refraction to bring a wavefront into focus [16]. The simplicity of manufacturing these lenses has made them very useful in coherent light scenarios, such as microscopic analysis with x-rays [17]. In our virtual light transport domain, this translates into simpler implementations with binary values, which require orders of magnitude less memory to store.

Our binary zone plate requires up to 16 times less memory than the RSD-based kernels while being also faster to compute. Our results show that the reconstructions using our ZP kernels provide an attractive trade-off between efficiency and accuracy, being a valid approximation for the exact wave diffraction of the RSD at a much smaller cost. Our approach represents a step forward toward the integration of NLOS imaging systems in embedded and small devices, e.g. in car driving systems or medical imaging.

2. BACKGROUND

2.1. Zone plates

A zone plate is a planar lens for spherical wavefronts. It is made up of concentric rings that focus light onto a focal point situated in a perpendicular direction from the center of the lens at a distance f, depending on the wavelength. Instead of refraction, the lens structure forces the incoming wavefront to diffract and generate the Arago spot on the focal point. Figure 1 shows an illustrative example, where the ZP acts as a virtual lens on the relay surface.

Given a specific focus distance f and wavelength λ, we determine a value $\mathcal{M}(\cdot)$ for a point $\mathbf{x_z} \in Z$, where Z represents the virtual lens, as

$$\mathcal{M}(\mathbf{x_z}, \lambda, f) = \frac{2}{\lambda}\left(\sqrt{r_{\mathbf{x_z}}^2 + f^2} - f\right), \tag{1}$$

where $r_{\mathbf{x_z}}$ is the distance of $\mathbf{x_z}$ to the center of the lens. The traditional Fresnel zone plate (FZP) is made up of alternating opaque and transparent rings, as

$$FZP(\mathbf{x_z}, \lambda, f) = \mathcal{M}(\mathbf{x_z}, \lambda, f) \mod 2. \tag{2}$$

The opaque rings of the FZP absorb part of the contribution from the incoming wavefront. A simple solution for increasing the energy that arrives at the focal point is to transform the opaque rings from the FZP into phase-shifting rings. The so-called phase Fresnel zone plate, or binary zone plate (bZP) shares the ring structure with the FZP. For the transparent rings, the value is 1 in both lenses, while the opaque rings with a value of 0 are set to -1 in the phase-shifting rings. So, a bZP is given by

$$bZP(\mathbf{x_z}, \lambda, f) = 2\left(\mathcal{M}(\mathbf{x_z}, \lambda, f) \mod 2\right) - 1. \tag{3}$$

2.2. Phasor Fields

Data from the hidden scene is collected by illuminating different points on the relay wall $\mathbf{x_l} \in L$, and the information is captured by a grid of ultra-fast sensors $\mathbf{x_s} \in S$, which can obtain a temporal domain impulse response $H(\mathbf{x_s}, \mathbf{x_l}, t)$ from the hidden scene, where t represents time [12].

$\mathcal{P}(\mathbf{x_l}, t)$ and $\mathcal{P}(\mathbf{x_s}, t)$ are two phasor field wavefronts. We define $\mathcal{P}(\mathbf{x_l}, t)$ as an input signal, which virtually illuminates the scene to produce the response wavefront $\mathcal{P}(\mathbf{x_s}, t)$ computed as

$$\mathcal{P}(\mathbf{x_s}, t) = \int_L \left[\mathcal{P}(\mathbf{x_l}, t) *_t H(\mathbf{x_s}, \mathbf{x_l}, t)\right] d\mathbf{x_l}, \tag{4}$$

where $*_t$ represents a convolution in time.

The reconstruction of a wavefront from a reflected wave is a problem already solved in LOS imaging. We define the image formation operation $\Phi(\cdot)$ as

$$I(\mathbf{x_v}, t) = \Phi(\mathbf{x_v}, \mathcal{P}(\mathbf{x_s}, t)), \tag{5}$$

resulting $I(\mathbf{x_v}, t)$ the image formation at time t via the propagation of the wavefront from the plane S back to the point $\mathbf{x_v}$ of the scene. In the Fourier domain, the diffractive wave propagation is the RSD integral. Hence, for a phasor in the Fourier domain $\mathcal{P}_\mathcal{F}(\mathbf{x_s}, \omega)$ with frequency ω, the image formation operator is

$$\Phi_\mathcal{F}(\mathbf{x_v}, \mathcal{P}_\mathcal{F}(\mathbf{x_s}, \omega)) = \int_S \mathcal{P}_\mathcal{F}(\mathbf{x_s}, \omega) \frac{e^{-i2\pi\omega|\mathbf{x_v}-\mathbf{x_s}|/c}}{|\mathbf{x_v}-\mathbf{x_s}|} d\mathbf{x_s}, \tag{6}$$

where c is the speed of light and $\Phi_\mathcal{F}(\mathbf{x_v}, \mathcal{P}_\mathcal{F}(\mathbf{x_s}, \omega))$ is the propagation of a monochromatic wavefront $\mathcal{P}_\mathcal{F}(\mathbf{x_s}, \omega)$ with frequency ω from S to the point $\mathbf{x_v}$ in the Fourier domain.

Upon discretization of the captured signal at S, the computation of Equation 6 integral is a summation for a specific point $\mathbf{x_v}$. By defining a volume V, which is a voxelization of points $\mathbf{x_v} \in V$, the summation can be applied for each point to yield the complete 3D reconstruction of the scene.

3. OUR METHOD

The Phasor Fields implementation is highly optimized for parallel reconstruction planes $\mathbf{p_v} \in V$ and equally spaced sensor grids $\mathbf{x_s} \in S$ [14]. Equation 6 propagation only differs in the z axis for each of the reconstruction planes $\mathbf{p_v}$. By

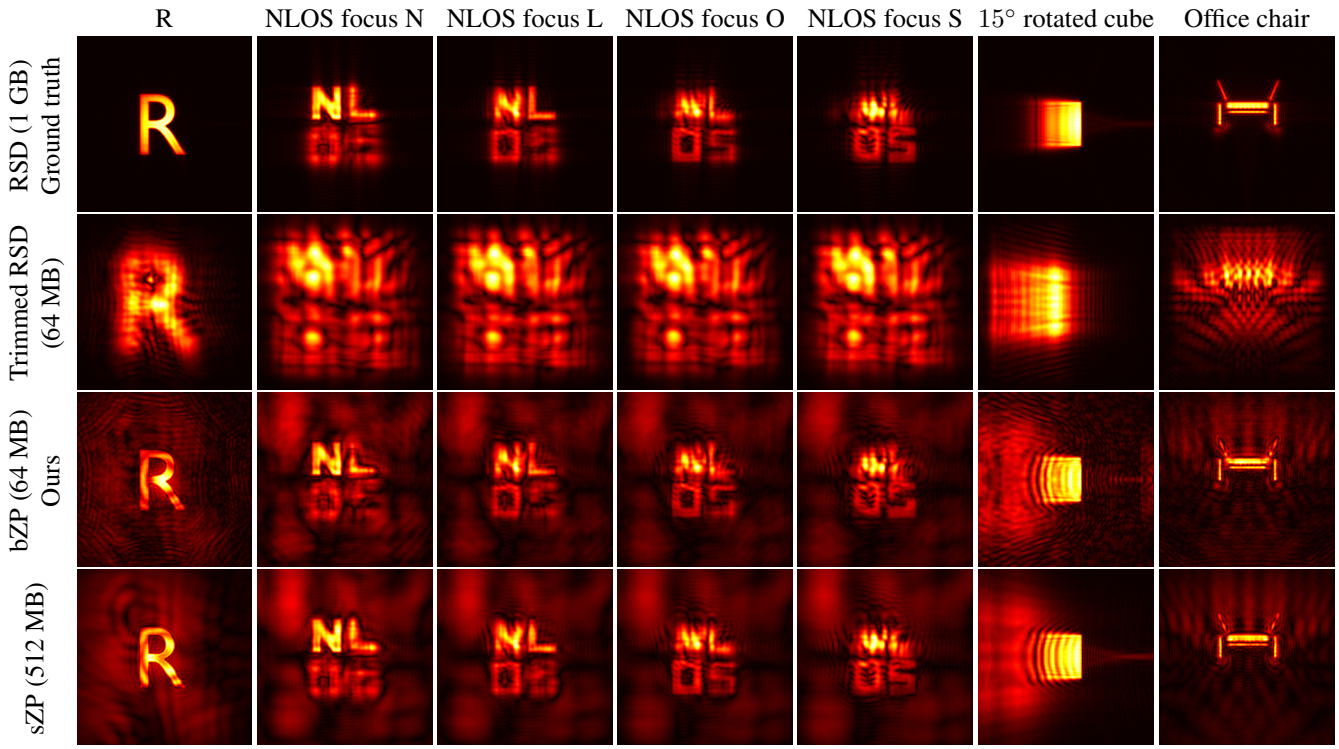

Fig. 2: Reconstructions of the synthesized hidden scenes using different kernels. From top to bottom: the exact RSD kernel (1GB), a trimmed RSD kernel (64MB), our bZP kernel (64 MB), and a sZP kernel (512 MB). The trimmed RSD kernel fails to properly reconstruct the hidden geometry, and the sZP kernels do not improve the reconstructions of the bZP, at the cost of expensive memory consumption. In contrast, our bZP kernel allows to image the hidden scene, with clearly identifiable geometry and sharp edges, despite the residual noise in empty areas, presenting an attractive memory trade-off compared with the RSD.

assuming Cartesian coordinates of $\mathbf{x_v} = (x_v, y_v, z_v)$ and $\mathbf{x_s} = (x_s, y_s, 0)$, where $\mathbf{x_v} \in \mathbf{p_v}$, we define a RSD kernel as

$$\mathcal{R}(x_s, y_s, z_v, \omega) = \frac{e^{-i2\pi\omega\sqrt{(x_s^2+y_s^2+z_v^2)}/c}}{\sqrt{(x_s^2+y_s^2+z_v^2)}}, \quad (7)$$

and thus, the propagation of a wave to a plane $\mathbf{p_v}$ at distance z_v from S is

$$\Phi_{\mathcal{F}}(\mathbf{p_v}, \omega) = \mathcal{P}(x_s, y_s, 0, \omega) * \mathcal{R}(x_s, y_s, z_v, \omega), \quad (8)$$

where $*$ represents a 2D spatial convolution, which can be efficiently computed with a 2D Fourier transform.

The image formation for the reconstruction as given in Equation 5 poses the propagation from the plane S to the volume V as a LOS problem, transforming the relay surface into a virtual camera that focuses using the exact diffraction propagation of Equation 6. Simulating the effect of changing the lens of a physical camera device, our proposal involves replacing the RSD kernel from Equation 8 with a bZP-based kernel in order to estimate the propagation of the wavefront into the scene through software. Assuming Cartesian coordinates $\mathbf{x_z} = (x_z, y_z, 0)$ and $\mathbf{x_z} = \mathbf{x_s}$, we can use the Equation 3 definition to generate our bZP kernel as

$$\Phi_{\mathcal{F}}(\mathbf{p_v}, \omega) = \mathcal{P}(x_s, y_s, 0, \omega) * bZP(x_s, y_s, 0, \omega, z_v). \quad (9)$$

While the RSD kernel values are complex numbers, our bZP kernel values cover only two integers (-1 and 1), which can be represented as a single signed byte without any precision loss. At the complete kernel size scale, the proportion of memory requirements is kept for the same number of values between the kernels, representing 16 times less memory for our complete bZP kernel storage.

4. RESULTS AND DISCUSSION

We have synthesized data to analyze the reconstructions using our bZP virtual lens, and to compare with the RSD propagation as ground truth. The synthesized data provides us flexibility while presenting a good agreement with the real captured data [18]. By using a transient renderer optimized for NLOS [19], we have generated four different scenes, with a common setup of an equally spaced sensor grid S of $2x2$ m, with 256^2 points $\mathbf{x_s} \in S$ in the relay surface, and a single illumination point $L = \mathbf{x_l}$ at the center of the grid. The

	R	NLOS letters focus N	NLOS letters focus L	NLOS letters focus O	NLOS letters focus S	15° rotated cube	Office chair
Trimmed RSD (64 MB)	1817.58	2851.79	3111.89	2907.99	3184.99	2491.67	1584.89
Our bZP (64 MB)	1248.21	**679.97**	**484.46**	**448.92**	**469.38**	1242.42	327.88
sZP (512 MB)	**1045.09**	755.59	554.07	477.37	525.62	**1174.90**	**271.48**

Table 1: RMSE of the reconstructions using different kernels relative to the RSD kernel of 1 GB reconstruction as ground truth. Our bZP kernel obtains better metrics than the trimmed RSD kernel of same memory in all the experiments. Note that, while obtaining similar values of the metric between our bZP and the sZP kernels, ours require 8 times less memory.

hidden scenes consist of four different geometries: a letter R, the four letters of NLOS at different distances, focusing on each one in sequence, a cube rotated 15º, and an office chair. We have used a single frequency propagation of wavelength $\lambda = 4\Delta\mathbf{x_s}$, where $\Delta\mathbf{x_s}$ is the minimum distance between samples in S, to reconstruct a 256^3 volume V. We have also trimmed an RSD kernel to provide fair comparisons, matching its memory storage to our bZP kernel.

Figure 2 (first three lines) shows the results for the plane of focus of the hidden geometries comparing the exact RSD solution (1GB of memory), the equal-memory trimmed RSD version, and our proposed bZP virtual lens (the last two needing only 64MB of memory). Compared to the exact diffractive wave propagation of the RSD kernels, focusing with our bZP kernel allows imaging the hidden scene at the cost of introducing noise in the empty areas. Despite this noise, the geometry of the hidden objects is clearly identifiable, with sharp edges. Note that we have not filtered the noise to better illustrate the outcome of our method; this kind of filtering is common even in well-established techniques like backprojection, and could be applied to our results as well. In contrast, the equal-memory trimmed RSD kernel reconstruction fails to properly reconstruct the hidden geometry, generating unfocused surface artifacts due to the smaller aperture used for the wave propagation. The assumption is supported by the numerical comparison of the Root Mean Square Error (RMSE) values presented in the two first rows of Table 1: the reconstructions of our bZP kernels obtain lower values in all the scenes compared with the trimmed RSD.

Additionally, since the bZP introduces abrupt discontinuities whereas a wavefront presents a smooth phase evolution, we also show the results of a continuous *sinusoidal* zone plate (sZP) for completeness purposes, defined as

$$sZP(\mathbf{x_z}, \lambda, f) = \cos\left(\pi \mathcal{M}(\mathbf{x_z}, \lambda, f)\right), \qquad (10)$$

following the notation for Equation 10. Equivalent to our Equation 9 bZP kernel, the new sZP kernel would become

$$\Phi_{\mathcal{F}}(\mathbf{p_v}, \omega) = \mathcal{P}(\mathbf{x_s}, \omega) * sZP(x_s, y_s, 0, \omega, z_v). \qquad (11)$$

As the fourth line in Figure 2 shows, the results do not improve our bZP lens. Furthermore, Table 1 present a comparison for the two lenses metrics (second and third rows) with polarized results. Moreover, the sZP continuous kernel values now span the real range $[-1..1]$, which requires float-precision values. Thus, its memory requirements discourage the use of the sZP as an attractive alternative to our proposed bZP.

The propagation of all the kernels requires similar computational time, since its implementation is the same convolution in the 2D Fourier space. In terms of kernel generation, our bZP is somewhat faster; the exact RSD kernel takes 3.5 s, while our bZP kernel only needs 1 s. For the sZP kernel, the time is increased to 1.4 s. The trimmed RSD kernel only takes 0.2 s, but at the cost of not being able to image the hidden scene. All these times have been obtained in a single thread of a i7-11700KF (3.6 GHz) processor, with 32 GB of RAM.

5. CONCLUSIONS

The Phasor Fields framework allows to theoretically transform the relay wall into a virtual lens, thus making a NLOS problem into a LOS one. This in turn allows researchers to exploit the wealth of knowledge of Fourier optics to come up with variations of the basic framework. In this work, we have tackled the problem of memory consumption when using RSD-based kernels to propagate light with convolutions.

Our proposal overcomes these memory consumption problems by introducing zone plates as virtual lenses. As we have shown, this offers an attractive trade-off between quality and needed resources, making it a viable option for small and portable devices with strong memory constraints, such as embedded and small devices integrated in car navigation systems, or medical exploration devices.

There are still many interesting avenues of research. Some other diffraction operators, lenses or kernels could be adapted to the virtual camera system to improve the reconstructions in terms of time, memory or quality. Future work could try these new operators, for example for non-parallel volume reconstructions.

6. REFERENCES

[1] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar, "Femto-photography: capturing and visualizing the propagation of light," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, pp. 1–8, 2013.

[2] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar, "Looking around the corner using transient imaging," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 159–166.

[3] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar, "Looking around the corner using ultrafast transient imaging," *International journal of computer vision*, vol. 95, no. 1, pp. 13–28, 2011.

[4] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nature communications*, vol. 3, no. 1, pp. 1–8, 2012.

[5] Victor Arellano, Diego Gutierrez, and Adrian Jarabo, "Fast back-projection for non-line of sight reconstruction," *Optics express*, vol. 25, no. 10, pp. 11574–11583, 2017.

[6] Byeongjoo Ahn, Akshat Dave, Ashok Veeraraghavan, Ioannis Gkioulekas, and Aswin C Sankaranarayanan, "Convolutional approximations to the general non-line-of-sight imaging operator," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7889–7899.

[7] Matthew O'Toole, David B Lindell, and Gordon Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature*, vol. 555, no. 7696, pp. 338–341, 2018.

[8] Weihao Xu, Songmao Chen, Yuyuan Tian, Dingjie Wang, and Xiuqin Su, "Fast non-line-of-sight imaging based on product-convolution expansions," *Optics Letters*, vol. 47, no. 18, pp. 4680–4683, 2022.

[9] Shumian Xin, Sotiris Nousias, Kiriakos N Kutulakos, Aswin C Sankaranarayanan, Srinivasa G Narasimhan, and Ioannis Gkioulekas, "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*, 2019, pp. 6800–6809.

[10] Joshua Rapp, Charles Saunders, Julián Tachella, John Murray-Bruce, Yoann Altmann, Jean-Yves Tourneret, Stephen McLaughlin, Robin MA Dawson, Franco NC Wong, and Vivek K Goyal, "Seeing around corners with edge-resolved transient imaging," *Nature communications*, vol. 11, no. 1, pp. 5929, 2020.

[11] Xiaohua Feng and Liang Gao, "Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging," *Nature communications*, vol. 12, no. 1, pp. 2179, 2021.

[12] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten, "Non-line-of-sight imaging using phasor-field virtual wave optics," *Nature*, vol. 572, no. 7771, pp. 620–623, 2019.

[13] David B Lindell, Gordon Wetzstein, and Matthew O'Toole, "Wave-based non-line-of-sight imaging using fast fk migration," *ACM Transactions on Graphics (ToG)*, vol. 38, no. 4, pp. 1–13, 2019.

[14] Xiaochun Liu, Sebastian Bauer, and Andreas Velten, "Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems," *Nature communications*, vol. 11, no. 1, pp. 1–13, 2020.

[15] Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez, "Recent advances in transient imaging: A computer graphics and vision perspective," *Visual Informatics*, vol. 1, no. 1, pp. 65–79, 2017.

[16] GW Webb, IV Minin, and OV Minin, "Variable reference phase in diffractive antennas: Review, applications, new results," *IEEE Antennas and Propagation Magazine*, vol. 53, no. 2, pp. 77–94, 2011.

[17] Harrison H Barrett, "Fresnel zone plate imaging in nuclear medicine," *Journal of nuclear medicine*, vol. 13, no. 6, pp. 382–385, 1972.

[18] Julio Marco, Quercus Hernandez, Adolfo Munoz, Yue Dong, Adrian Jarabo, Min H Kim, Xin Tong, and Diego Gutierrez, "Deeptof: off-the-shelf real-time correction of multipath interference in time-of-flight imaging," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 6, pp. 1–12, 2017.

[19] Diego Royo, Jorge García, Adolfo Muñoz, and Adrian Jarabo, "Non-line-of-sight transient rendering," *Computers & Graphics*, vol. 107, pp. 84–92, 2022.