Universidad Zaragoza
1542

Trabajo Fin de Máster

Computational aperture masks for virtual-wave non-line-of-sight imaging

Autor

Carlos Carazo de la Fuente

Director

Julio Marco Murria

Máster Universitario en Robótica, Gráficos y Visión por Computador

ESCUELA DE INGENIERÍA Y ARQUITECTURA
2024

Computational aperture masks for virtual-wave non-line-of-sight imaging

ABSTRACT

The field of non-line-of-sight imaging aims to obtain images and three-dimensional reconstructions of scenes with which the observer has no direct line of sight. To this end, it makes use of the acquisition of the illumination reflected by these scenes over a secondary surface at speeds comparable to light's, and the computational processing of the captured information. This domain offers numerous potential applications, from autonomous driving to rescue operations.

This work addresses one of its main limitations: the efficiency in the computation of the mentioned reconstructions. Current algorithms require the processing of large amounts of data, both temporal and spatial, to achieve the temporal focusing that makes it possible to obtain the pictures of the hidden scenes. In order to reduce the high computational cost that this entails, we make use of known wave propagation models, and restrain to methods that exploit the information carried by a single temporal frequency of the captured signals. This allows to reduce the number of computations greatly and obtain steady-state images, like the ones obtainable with a conventional camera. To mitigate the out-of-focus illumination caused by the shallow depth of field of these, we propose as our main contribution the incorporation of virtual pinholes to these methods: pinholes, widely known in areas like classic photography and confocal microscopy, make it possible to achieve a spatial focusing that mitigates the out-of-focus artifacts and allows to reduce the number of planes needed to reconstruct a scene.

We present the mathematical formulation that defines current wave-propagation models in non-line-of-sight imaging, as well as our contributions to it in order to improve the reconstructions that are obtained with a single temporal frequency while maintaining its efficiency. To conclude, we conduct an experimental study of the results that can be gotten with the proposed methods and the advantages that these present with respect to other existing wave-based methods.

Computational aperture masks for virtual-wave non-line-of-sight imaging

RESUMEN

El campo de non-line-of-sight imaging estudia la obtención de imágenes y reconstrucciones tridimensionales de escenas con las que el observador no tiene línea directa de visión. Para ello, se sirve de la adquisición de la iluminación reflejada por estas escenas sobre una superficie secundaria a velocidades comparables a las de la luz, y del tratamiento computacional de la información capturada. Este dominio presenta numerosas aplicaciones potenciales, desde conducción autónoma hasta operaciones de rescate.

En este trabajo se aborda una de sus principales limitaciones: la eficiencia en el cálculo de las mencionadas reconstrucciones. Los algoritmos actuales requieren del procesamiento de un elevado número de datos, tanto temporales como espaciales, para conseguir el enfoque temporal que posibilita la obtención de imágenes de las escenas ocultas. Con el objetivo de reducir el alto coste computacional que esto conlleva, hacemos uso de modelos conocidos de propagación de ondas y nos ceñimos a métodos que explotan la información contenida en una sola frecuencia temporal de las señales capturadas. Esto permite reducir el número de operaciones en gran medida y obtener imágenes estáticas, como las obtenidas con una cámara convencional. Para mitigar la iluminación fuera de foco causada por la escasa profundidad de campo de estas, proponemos como principal aportación la incorporación de *pinholes* virtuales a estos métodos: estas herramientas, ampliamente conocidas áreas como la fotografía clásica y la microscopía confocal, hacen posible lograr un enfoque espacial que mitiga los artefactos fuera de foco y permite reducir el número de planos necesarios para reconstruir una escena.

Presentamos la formulación matemática que rige los modelos actuales de propagación de ondas para non-line-of-sight imaging, así como nuestras contribuciones a la misma para mejorar las reconstrucciones obtenidas con una frecuencia temporal manteniendo su eficiencia. Para concluir, realizamos un estudio experimental de los resultados que se pueden conseguir con los métodos propuestos y las ventajas que presentan con respecto a otros ya existentes basados en propagación de ondas.

Contents

1 Introduction

Computational imaging is the field that combines the capture and computational processing of light transport data to form images that are not possible to obtain with conventional cameras and sensors. Within this, time-of-flight imaging refers to the set of techniques that involve capturing the illumination reflected by a scene at a speed that is comparable to the speed of light and extracting information about this data. This is achieved by leveraging the time-of-flight information of the captured light, which contains information about the distance travelled by the photons from the scene to the sensor. In this document we present our work in non-line-of-sight (NLOS) imaging, the particular area of time-of-flight imaging that encompasses methods that allow to obtain images and reconstructions of scenes that are occluded from the point of view of the observer, provided that there is an available surface that can be used for measuring indirect light [MSS+19], [FVW20], [SHN+16]. This field has shown promising potential for real life applications, such as autonomous driving, rescue operations or even space-related imaging.

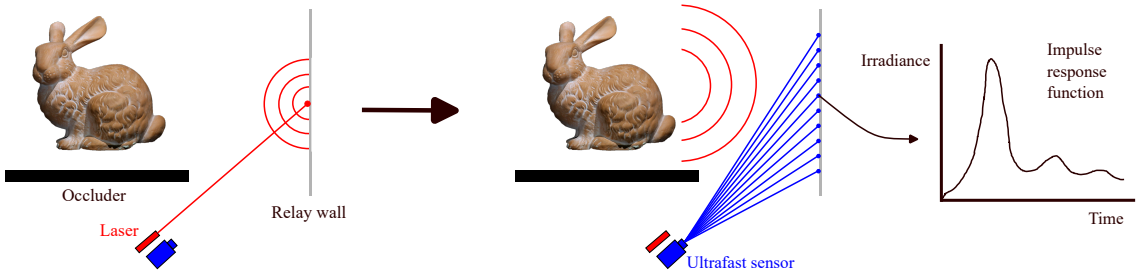

Figure 1: Setup and capture process in a typical NLOS scenario

This work is based on methods that often rely on the following active setup [BZT+15] (see Figure 1): A laser device emits an ultrafast light pulse at one point $\boldsymbol{x}_l$ on the diffuse surface that is visible (the relay wall, from now on) from the observer point. From there, light is scattered in all directions, interacts with the objects in the hidden scene and comes back to the relay wall. An ultrafast sensor (typically a single-photon avalanche diode) is then used to record the time-resolved irradiance that returns at a grid of points $\{\boldsymbol{x}_s^1, ..., \boldsymbol{x}_s^n\}$ at this surface simultaneously over a short period of time, with picosecond precision. The measured time-dependent signals approximate the impulse response function of the scene, $H(\boldsymbol{x}_l, \boldsymbol{x}_s, t)$, and the processing of all of them is what allows to achieve the desired reconstructions.

Depending on the work domain, we can make a distinction between time-based and frequency-based techniques: the first kind (backprojection and derived [VWG+12], [AGJ17], [WLL23]) leverages the fact that the speed of light is limited and known to make a time-to-distance conversion and exploit the recorded data, containing light's time-of-flight information, to determine the location of hidden surfaces that could have reflected the recorded light. This is known as temporal focusing, and with these time-based methods it is achieved by accounting, for every coordinate in the hidden scene, only for values in the impulse response function that match light's time-of-flight to that point. The majority of these reconstruction methods work under the assumption that all captured signal comes from third-bounce illumination following paths from the laser, to $\boldsymbol{x}_l$, to the hidden scene, to $\boldsymbol{x}_s$ and back to the sensor. This premise is the cause of the main challenges most methods experience: ambiguities and multi-path interference (a photon that reaches the sensor at a certain time could have come from infinite different points, or could have experienced interreflections within the scene). In order to overcome them and triangulate the hidden geometry effectively, one must combine the information contained in signals that are measured in different spatial positions. As for the frequency-based, Liu et al. [LGLM+19] presented a phasor-based formulation for NLOS imaging that computationally transforms the captured impulse response functions into a set of phasor fields representing a virtual wave captured at the relay surface after being reflected by the hidden scene. This allows us to use wave-based operators to perform virtual imaging of the hidden scene as if the relay wall was the aperture of a virtual camera that directly observes the hidden scene. With this framework, the temporal focusing

is performed by translating time-of-flight information into wave phase shifts and adding the resulting virtual waves to focus the virtual camera at any desired depth. This is employed to obtain a set of pictures of the hidden scene, each one focused at a different depth, giving as a result a focal stack of images like the ones known in conventional photography [CR99], [LSW+13], [SPT10]. As with time-based methods, temporal ambiguities are also present in phasor-based reconstructions, in this case caused by the large baselines of the virtual aperture. These cause, for every reconstruction plane, the distorted inclusion of surfaces that are actually present at a different depth in the scene (out-of-focus illumination), which is usually addressed by applying temporal filters to the captured measurements and combining the information contained in multiple temporal frequencies. Applying the phasor-based propagation models to numerous frequencies over a volumetric space turns out expensive in terms of computational cost, which makes it mandatory to employ equipment with high processing power to obtain fast reconstructions.

Motivated by this, we take the phasor fields paradigm and restrain to single-frequency models that require far fewer calculations and thus allow us to obtain NLOS reconstructions in a fraction of the time. Since with single-frequency propagations it is not possible to perform temporal focusing to mitigate the mentioned out-of-focus effects, we propose to modify the formulations to include computational photography elements in them that achieve spatial focusing; in particular, we incorporate computational pinholes to the virtual focusing of large aperture baselines and show that these computational pinholes extend the depth of field of plain single-frequency reconstructions besides filtering out-of-focus illumination. As a result, we get to improve the results of these with less reconstruction planes while being more efficient than the multi-frequency approach.

In this work, we study the potential of NLOS wave propagation models based on single-frequency phasor fields that incorporate computational pinholes: after presenting its formulation, we put it to the test with simple scenes and increase the complexity of these gradually. In order to do this, we implement a single-frequency wave simulator to design both 2D and 3D virtual scenes and compute single-frequency light transport to these and back, emulating the data that can be captured by actual NLOS setups. Finally, we display the results that can be obtained by applying the proposed methods to both simulated and actual NLOS datasets, and show the advantages it offers with respect to previous models in terms of both quality and speedup. In short, we are able to obtain NLOS reconstructions with significantly less out-of-focus artifacts than with bare single-frequency propagations in at least half of the time, both when we have some previous knowledge about the scene and when we do not.

The application of the proposed methods could reduce the hardware requirements for practical real-time applications: besides lowering the processing power and memory requirements, the SPAD sensors could be substituted for amplitude-modulated continuous-frequency time-of-flight sensors that only work with a fixed frequency, reducing the cost of the setup significantly.

2 Related work

2.1 NLOS imaging

The foundations of this work lie in existing NLOS imaging methods with a special focus on the time-gated ones, that is, the approaches that leverage light's time-of-flight information to obtain reconstructions about the hidden scenes. In particular, we put our attention on active methods, where the scene is illuminated by a controlled source [CdGB+23], [KHFG14], tipically a laser that emits light pulses [BZT+15]. The availability of ultrafast sensors has given rise to a spectrum of approaches that differ in the way the temporal measurements are processed in order to extract information about the scene.

The first and best known method is backprojection [VWG+12], an algorithm that is also used in computed tomography scans and can be applied to both confocal and non-confocal measurements. It works on the temporal domain to determine, within an uncertainty ellipsoid, where the recorded photons in a temporal bin could have come from. By adding every ellipsoid in the target volumetric space, each weighted by its corresponding intensity in the measurement, a 3D heatmap is obtained where the areas with highest reprojected intensity are the more likely to contain geometry. If confocal measurements are available, NLOS reconstructions can be treated as a spatially invariant 3D deconvolution problem thanks to the light-cone transform approach [OLW18], which makes it possible to invert light transport in an exact manner by undoing this operation in the Fourier domain.

In contrast with these approaches, the phasor field framework [LGLM+19], [LBV20] offers a different point of view on NLOS imaging: it provides the means to solve the same problems by considering them waves-propagation ones. In essence, it allows to treat the relay wall as a virtual lens that can be focused in any part of the scene, that is, it opens the door to the use of line-of-sight (LOS) tools in this domain. This paradigm, which can be used with both confocal and non-confocal setups, serves as a basis on top of which more elaborate implementations can be built: from light transport matrices [MJN+21], which let us isolate direct and indirect illumination reflected by the hidden scene, to fast f-k Migration [LWO19], which brings tools previously employed in seismology to the NLOS domain, or even virtual mirrors [RSM+23], which exploit light bounces beyond the third one to image scenes that are behind more than one corner.

All these methods entail costly computational processing of a target volumetric space, which makes them unfeasible for not-high-end equipment or real-time applications. By leveraging the possibility of working in the frequency domain with the phasor field framework, this problem can be overcome with more efficient implementations, e.g., with virtual zone plates [LLGM23], which allow to reduce memory requirements up to 8 times. Along these lines, we take the phasor fields paradigm as the main basis of our work, as it allows us to compute single-frequency propagations that turn out much faster and efficient in terms of computational cost, and combine it with computational apertures to address current challenges of the single-frequency approach.

2.2 Classic computational imaging

Besides working on top of the presented NLOS background, this work aims to incorporate well established techniques and setups from other computational imaging areas that have not been applied to NLOS imaging yet.

Part of the approach that we propose is inspired by confocal microscopy, the area of computational imaging that focuses on obtaining reconstructions of small samples at different depths by combining the use of lenses, dichroic mirrors and pinholes [Pad08], [Mas04], [SA77]. The microscopes that are built using these achieve sharp images thanks to the removal out-of-focus light, in the following way: the light source is focused at a point in the object at a certain depth using a pinhole, a dichroic mirror and a lens. The same lens is then used to focus the light reflected by the point at the photodetector, and an additional pinhole blocks the light that was scattered by other points in the sample. In these systems, only one pixel is imaged at a time, therefore, either the sample or the imaging system has to be moved to effectively scan the whole object to be imaged. Throughout this work, we will leverage the idea of using a pinhole [Ren09], [ADLT22] to increase the narrow depth of field given by large NLOS aperture baselines, and incorporate it to single-frequency propagations to obtain in-focus images efficiently.

As it is elaborated in Section 4, the main objective of this work is to obtain a fast, efficient NLOS imaging system by combining low-computational-cost elements from both existing NLOS literature

(single-frequency phasor fields propagations) and well established tools from more classic computational imaging areas (use of pinholes).

3 Background

In this section, in addition to covering the basics of NLOS data capture and imaging principles, we go over the wave-based NLOS methods on which this work is based, with a special attention to their theoretical foundations and formulation. As introduced, these two differentiated parts (capture and processing) work hand-in-hand to obtain images of NLOS scenes: the capture phase, which involves a physical setup, is meant to obtain time-resolved data regarding the light reflected by the hidden surfaces, whereas the processing one makes use of this information plus physical principles to obtain reconstructions and only takes place in the computational domain.

3.1 NLOS data capture and computational basis

Data capture. Here, we present an elaboration on the way NLOS measurements are taken, and the implications of this with regard to the scene reconstructions that can be obtained from them. We will restrain to non-confocal measurements, i.e., those in which only one illumination point is used in total. As mentioned in the introduction section, these are obtained as follows (see Figure 2): a laser device emits a light pulse on a point $\boldsymbol{x}_l$ at the relay surface, the one that is visible from the observer's point of view. This diffuse surface scatters light in all directions, and then irradiance is recorded for a short period of time by an ultrafast sensor (single photon avalanche diode, or SPAD) that is focused at a grid of points $\{\boldsymbol{x}_s^1, ..., \boldsymbol{x}_s^n\}$ on this same surface. This gives as a result a set of discrete samples $H(\boldsymbol{x}_l, \boldsymbol{x}_s^1, t), ..., H(\boldsymbol{x}_l, \boldsymbol{x}_s^n, t)$ of the irradiance that reaches the relay wall over time, referred to as impulse response function. This incidental irradiance will be caused by the reflections of light on the objects present at the hidden scene, whose position can be triangulated thanks to the temporal information captured in the measurements after a time-distance conversion using the known speed of light.

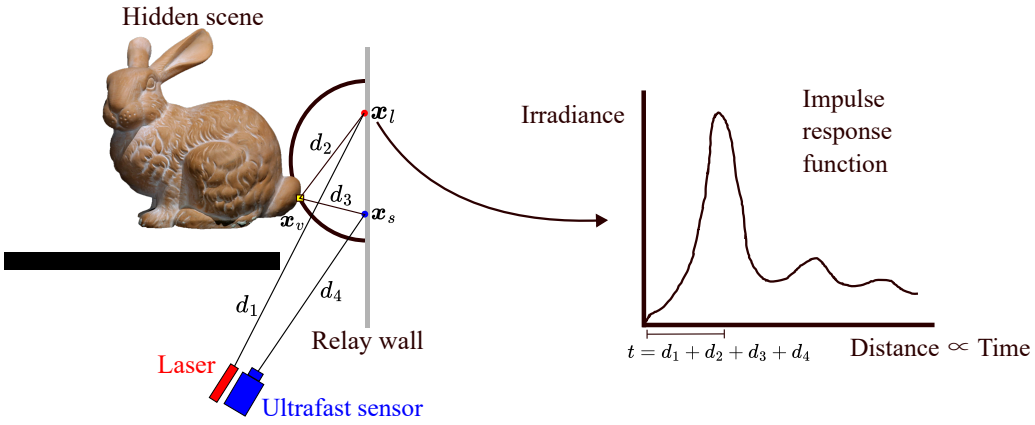

Figure 2: For a given time bin t, the surfaces that could have contributed to the intensity in $H(\boldsymbol{x}_l, \boldsymbol{x}_s, t)$ are located at some point of an ellipsoid with foci $\boldsymbol{x}_l$ and $\boldsymbol{x}_s$

The majority of NLOS imaging models work on the basis of third-bounce light transport, that is, light is assumed to travel only from the illuminated spot $\boldsymbol{x}_l$ to a point in the scene $\boldsymbol{x}_v$ (distance d_2 in the Figure) and back to the measured spot $\boldsymbol{x}_s$ (distance d_3), with no indirect interreflections between elements in the hidden scene. Under this assumption, the captured impulse response function is given by:

$$H(\boldsymbol{x}_l, \boldsymbol{x}_s, t) = \int \delta\left(t - \left(\frac{d_2 + d_3}{c}\right)\right) f(\boldsymbol{x}_v) d\boldsymbol{x}_v, \tag{1}$$

where c symbolises the speed of light and f represents the albedo of all the surfaces present in the hidden scene.

Computational basis. The main objective in NLOS reconstructions is to know the shapes and locations of the hidden geometry, for which it is necessary to invert the captured light transport back into the scene. Regardless of the computation method, this is tipically performed voxel-by-voxel: each one of these volumetric pixels $\boldsymbol{x}_v$ represents a portion of 3-dimensional space, whose size is determined by the reconstruction resolution, that will hold the light intensity that was reflected by the objects placed there. In all the formulations presented in this work that involve computing a distances to a voxel, these will be computed with respect to its center.

The third-bounce assumption mentioned previously allows us to narrow the position of all points in the scene that could have contributed to the irradiance measured in a given time bin t from one of the signals to an ellipsoid with foci $\boldsymbol{x}_l$ and $\boldsymbol{x}_s$. This is due to the fact that for a photon to arrive the sensor at time t, it must have travelled two distances d_2 and d_3 such that $t = d_2 + d_3$ (In an ellipsoid, the sum of the distances between any point and the foci is constant). It is important to note that the light pulse must also travel from the laser device to the illuminated point (distance d_1) and from the measured point to the camera sensor (distance d_4). In practice, though, these distances are just added as offsets to all propagations, since they are fixed and known from the beginning.

This information constitutes the basis of most NLOS imaging methods, although it makes it necessary to combine several measurements to compensate for the inherent ambiguity (a photon could have been reflected from infinite different points on the mentioned ellipsoid). With the backprojection method, this is achieved by adding all the uncertainty ellipsoids corresponding to every temporal bin and every measurement, each one weighted by their intensity.

The third-bounce assumption simplifies the algorithms, but does not consider multi-path interferences: third-bounce photons reflected by distant surfaces may arrive to the sensor at the same time that others that have been reflected several times by objects in the scene that are closer to the relay wall. To overcome this issue, more advanced methods like virtual mirrors [RSM+23] must be employed.

3.2 Wave-based NLOS imaging

As mentioned in 2.1, the phasor field framework introduced by Liu et al. [LGLM+19] for NLOS imaging reconstructions makes it possible to treat these as wave propagation problems. In this subsection, we deal with the concepts and propagation operators that constitute the basis of this paradigm.

Let us consider the impulse response function $H(\boldsymbol{x}_l, \boldsymbol{x}_s, t)$ recorded by the capture device. Since the relay wall measurements are taken after the emission of an approximately delta illumination pulse $\delta(\boldsymbol{x}_l, t)$, which carries information about a wide set of temporal frequencies, it is possible to compute the response of the scene to any other time-resolved illumination function $\mathcal{P}(\boldsymbol{x}_l, t)$ by means of a temporal convolution. This way, different illumination functions will translate into different ways of filtering the recorded data. By taking into account all illumination sources, the response function of the scene to the chosen illumination function at measured points $\boldsymbol{x}_s$ will be given by:

$$\mathcal{P}(\boldsymbol{x}_s, t) = \int_{\mathcal{L}} \mathcal{P}(\boldsymbol{x}_l, t) * H(\boldsymbol{x}_l, \boldsymbol{x}_s, t) d\boldsymbol{x}_l \tag{2}$$

In multi-frequency phasor-based approaches, a common illumination function consists in a pulse wave with a Gaussian envelope centered around a certain wavelength λ and with standard deviation σ, $\mathcal{P}(\mathbf{x}_l, t) = e^{i2\pi \frac{t}{\lambda} - \frac{1}{2}\left(\frac{t}{\sigma}\right)^2}$, which amounts to a filter that emphasises the information carried by a set of frequencies.

Once the response of the hidden scene to an illumination function has been computed, we need to translate this information to the frequency domain by means of the Fourier transform, $\widehat{\mathcal{P}}(\boldsymbol{x}_s, \omega) = \mathcal{F}(\mathcal{P}(\boldsymbol{x}_s, t))$, in order to be able to work with wave propagation operators. Every Fourier coefficient receives the name of phasor, and is a complex number containing the amplitude and phase of the sinusoidal function corresponding to a frequency ω in the Fourier representation. Now, a mathematical tool that serves us to invert the light transport to obtain images of it is needed, accounting for both the phase shift and the spatial decay that the computational light waves experience. Here, we introduce the main propagation operator that supports the majority of wave-based NLOS methods: given an incoherent electric field at an isotropic source plane S, the propagation of any monochromatic component $\widehat{\mathcal{P}}(\boldsymbol{x}_s, \omega)$ to a point $\boldsymbol{x}_d$ in a destination plane D is given by an Rayleigh-Sommerfeld

diffraction (RSD) propagation integral:

$$\widehat{\mathcal{P}}(\boldsymbol{x}_d, \omega) = \gamma \int_{\mathcal{S}} \widehat{\mathcal{P}}(\boldsymbol{x}_s, \omega) \frac{e^{ik|\boldsymbol{x}_d - \boldsymbol{x}_s|}}{|\boldsymbol{x}_d - \boldsymbol{x}_s|} d\boldsymbol{x}_s, \tag{3}$$

where γ symbolises an attenuation factor, often with the value of the inverse of the mean distance between planes S and D, and $k = 2\pi/\lambda$ is the wave number. By means of this operator we can obtain the back-propagated phasor field $\widehat{\mathcal{P}}(\boldsymbol{x}_v, \omega)$ at every voxel in the hidden scene, to which an image formation function $\Phi(\cdot)$ can be applied after to generate the image of the hidden scene as seen from the virtual camera. A common imaging function in NLOS methods is the composition of the inverse Fourier transform with the complex modulus, or an operator derived from it (can vary according to the visualization purpose): $\Phi(\widehat{\mathcal{P}}(\boldsymbol{x}_v, \omega)) = |\mathcal{F}^{-1}(\widehat{\mathcal{P}}(\boldsymbol{x}_v, \omega))|$.

Depending on the kind of reconstruction we want to get from the measurements, different variations of the propagation integral 3 can be used: the most common camera models are the transient and the confocal one. We will focus on the latter, since it is the one we will be applying in our method. In order to achieve temporal focusing, the confocal camera implements two computational lenses: one is meant to focus the illumination aperture, while the other focuses on the voxel to obtain its image. To accomplish this, the phasor field corresponding to frequency ω at a given voxel $\boldsymbol{x}_v$ is computed with the following RSD-like operator:

$$\widehat{\mathcal{P}}(\boldsymbol{x}_v, \omega) = \int_{\mathcal{S}} \frac{e^{ik|\boldsymbol{x}_s - \boldsymbol{x}_v|}}{|\boldsymbol{x}_s - \boldsymbol{x}_v|} \int_{\mathcal{L}} \frac{e^{ik|\boldsymbol{x}_v - \boldsymbol{x}_l|}}{|\boldsymbol{x}_v - \boldsymbol{x}_l|} \widehat{\mathcal{P}}(\boldsymbol{x}_l, \omega) \cdot \widehat{H}(\boldsymbol{x}_l, \boldsymbol{x}_s, w) d\boldsymbol{x}_l d\boldsymbol{x}_s, \tag{4}$$

where $\widehat{\mathcal{P}}(\boldsymbol{x}_l, \omega)$ and $\widehat{H}(\boldsymbol{x}_l, \boldsymbol{x}_s, w)$ represent the temporal Fourier transform of the illumination function and the impulse response function, respectively. This kind of propagation performs phase shifts that account for the light's time-of-flight from all the illuminated points $\boldsymbol{x}_l$ to the voxel in question $\boldsymbol{x}_v$, and from this to all the capture points, in order to focus the virtual camera on it. Since these computations are performed for every temporal frequency ω, the result after going back to the time domain consists in a "temporal image" for every reconstructed voxel, $I(\boldsymbol{x}_v, t) = \Phi(\widehat{\mathcal{P}}(\boldsymbol{x}_v, \omega))$, that shows the light intensity that was reflected by the object in that position at every instant t starting from the first moment that light reached it. That is, $I(\boldsymbol{x}_v, t = 0)$ will display the whole scene in focus, as most time-gated NLOS imaging methods do, because of the fact that all illumination-voxel-capture distances were already accounted for.

In the presented equations, only light paths from the illuminated point to the scene and back to the points where the measurements are taken are considered (distances $d_2 = |\boldsymbol{x}_l - \boldsymbol{x}_v|$ and $d_3 = |\boldsymbol{x}_v - \boldsymbol{x}_s|$ in Figure 2). In actual datasets where the recordings start the moment of the laser emission, two more distances have to be taken into account: from this device to the illuminated point (d_1) and from every capture point to the actual camera sensor (d_4). In this situation, these are accounted by adding them as offsets to the phase shift parts in all propagations. This can also be done beforehand, since these distances are fixed and known from the start. Their omission in theoretical elaborations simplifies the formulation of the algorithms and has no effect on reconstructions [RMBV19], [MJN+21].

3.3 Scene reconstructions in NLOS imaging

Non-line-of-sight reconstructions, regardless of the method employed, provide a 3D intensity map $I(x, y, z)$ [1] where the areas with the highest values are the most likely to contain geometry. These are usually obtained plane by plane (see Figure 3) using temporal focusing to image the objects present in each depth while mitigating out-of-focus artifacts, i.e., illumination that was reflected from objects present in other parts of the scene. In the case of wave-based methods, this temporal focusing is achieved by combining the propagations corresponding to a set of multiple frequencies.

[1] In all formulations presented in this work that involve voxel positions, their (x, y, z) coordinates are implicit in the employed notation $\boldsymbol{x}_v$.

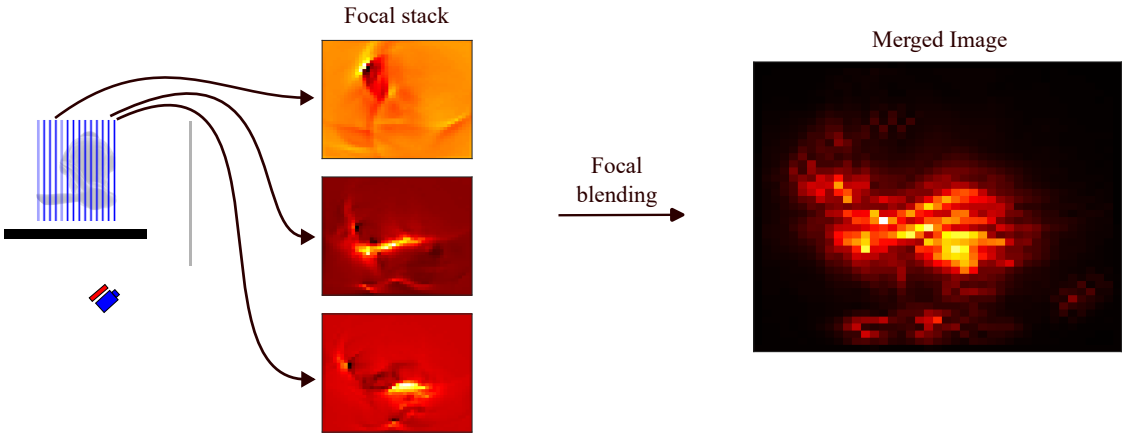

Figure 3: NLOS reconstructions involve computing a focal stack over a set of planes at different depths, which is then merged into a single image of the scene

In other words, the computed volume represents a focal stack of images where each one displays, in focus, the surfaces that are present at a different depth in the scene. In order to obtain a single picture of the scene after this set of images, a common approach that resembles classic focal blending techniques consists in projecting the maximum intensity along the depth direction (max-projection, from now on), in line with the premise mentioned in the first sentence: $I(x, z) = \max\{I(x, y, z) : y_{min} \leq y \leq y_{max}\}$.

4 Contributions

As explained in Section 3.3, the process to obtain an image of a non-line-of-sight scene requires the computation of a focal stack made up of several reconstruction planes, each one focused at a different depth. Then, the information contained in this can be merged into a single picture (usually by applying max-projection, as introduced in Section 3.3) where all the scene elements appear in focus, regardless of their position. The whole process, especially the first part, has a high computational cost due to the amount of voxels and frequencies that need to be processed to minimize the out-of-focus effects caused by the large size of the aperture. The temporal focusing with multiple frequencies explained in Section 3.2 applied to a reconstruction plane only yields a picture of that particular plane, which makes it mandatory to compute multiple planes to obtain a full representation of the hidden scene. In our work, we propose to employ single-frequency reconstructions in contrast to multi-frequency or time-based reconstructions, which has several considerations: working with a single frequency illumination function implies that the temporal information of the original measurements is lost, as the new obtained signals represent the response of the scene to an illumination source that is constant in time, and therefore it is not possible to implement temporal focusing. As a consequence, we will obtain steady-state images of the scene, like the ones that can be captured with a conventional camera, where all the objects appear in the picture (not necessarily in focus). The main limitation of this approach lies in the narrowness of the depth of field, caused by the considerable size of the aperture and the fact that it is not possible to leverage temporal information, which gives the surfaces that are not in the reconstruction plane a blurry (out of focus) appearance in the images. In order to overcome it, we incorporate pinholes, known in conventional photography for their good focusing properties, into single-frequency reconstruction formulations. This will allow us to reduce the required number of images to see the whole hidden scene, as each one will display surfaces in focus in a wider depth range.

To sum up: in contrast to mitigating out-of-focus effects by performing temporal focusing with multi-frequency propagation, we propose a new method that uses single-frequency propagations through computational pinholes to achieve spatial focusing. As a result, we will be able to obtain images of a non-line-of-sight scene efficiently by processing single-frequency signals over a significantly lower number of volumetric planes. Along the lines of most of existing methods, throughout this work we will assume a delta emission from a single point $\boldsymbol{x}_l$ and only third-bounce light paths.

4.1 Single-frequency scene reconstructions

Here, we go over the mathematical tools that we will be applying to single-frequency data in order to obtain the NLOS reconstructions, and the advantages that this approach offers with respect to the multi-frequency one:

As seen in Section 3.2, the captured phasor field at the relay wall can be projected back to the scene to obtain an image of the scene at any voxel $\boldsymbol{x}_v$, by means of the confocal camera operator (Equation 4). In this work we are only interested in the information carried by one frequency ω, and therefore we need an illumination function that translates into a delta function in the frequency domain: a sinusoidal one. In order to work with it using the phasor representation, we choose its general complex form: $\mathcal{P}(\boldsymbol{x}_l, t) = e^{i\omega t}$. Since we will only be using the information relative to this sole frequency, the continuous domain on the ω variable becomes the evaluation of a sole fixed value ($\widehat{\mathcal{P}}(\boldsymbol{x}_v, \omega)$ and $\widehat{\mathcal{P}}(\boldsymbol{x}_l, \omega)$ turn into $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v)$ and $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l)$, respectively):

$$\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v) = \int_{\mathcal{S}} \frac{e^{ik|\boldsymbol{x}_s - \boldsymbol{x}_v|}}{|\boldsymbol{x}_s - \boldsymbol{x}_v|} \int_{\mathcal{L}} \frac{e^{ik|\boldsymbol{x}_v - \boldsymbol{x}_l|}}{|\boldsymbol{x}_v - \boldsymbol{x}_l|} \widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l) \cdot \widehat{H}(\boldsymbol{x}_l, \boldsymbol{x}_s, \omega) d\boldsymbol{x}_l d\boldsymbol{x}_s, \tag{5}$$

Moreover, the fact that we are constraining our reconstructions to one illumination point (besides single-frequency data) allows us to omit the illumination focusing part, as it can be extracted as a common coefficient that will not affect the final images due to the following facts:

- The attenuation term $\dfrac{1}{|\boldsymbol{x}_v - \boldsymbol{x}_l|}$ would only act as a scale factor.
- We only visualise the modulus (or an operator derived from it) of the computed phasors $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v)$, which would by unnafected by a phase change $e^{ik|\boldsymbol{x}_v - \boldsymbol{x}_l|}$.

Consequently, in our reconstructions we will be using a simpler and more efficient operator, which computes the single-frequency propagation and accumulation from the phasor field at $\mathcal{S}$ to the voxel $\boldsymbol{x}_v$ at the reconstructed plane:

$$\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v) = \int_{\mathcal{S}} \frac{e^{ik|\boldsymbol{x}_s - \boldsymbol{x}_v|}}{|\boldsymbol{x}_s - \boldsymbol{x}_v|} \widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l) \cdot \widehat{H}(\boldsymbol{x}_l, \boldsymbol{x}_s, \omega) d\boldsymbol{x}_s \tag{6}$$

All these computation savings contrast with the computationally expensive multi-frequency approach, where the whole operator from Equation 4 (the illumination focusing part cannot be omitted to achieve temporal focusing) must be applied to multiple frequencies. After the phasor propagations to every target voxel in the scene, the reconstruction is finally obtained by visualising the modulus (or its square) of the computed voxel phasors: $I(\boldsymbol{x}_v) = \Phi(\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v)) = |\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v)|$. In this case, it is not necessary to apply an inverse Fourier transform to them, since they do not hold any temporal information.

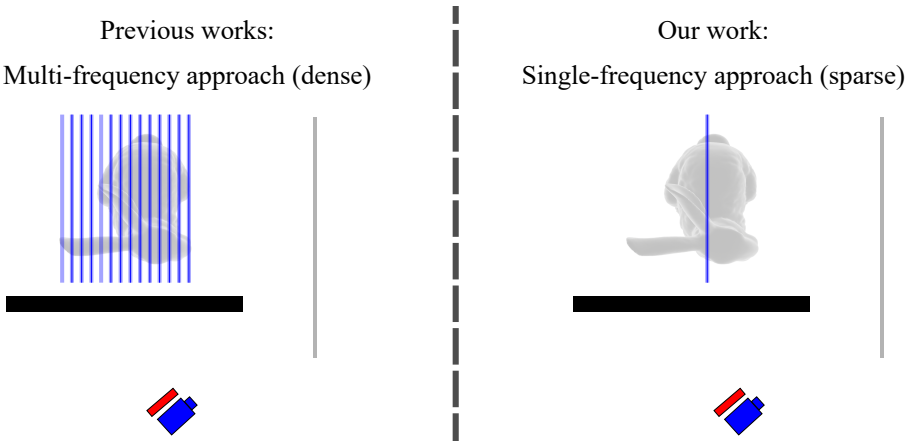

Figure 4: Multi-frequency phasor fields computations require dense reconstructions due to temporal focusing, while with single-frequency all the scene is visible in one reconstruction plane due to the loss of temporal information

Lastly, as explained in the introduction of this section, each reconstruction plane will yield a steady-state image that will show all the objects in the hidden scene. Thanks to this, we will be able to visualise the full scene with one or a few reconstruction planes (see Figure 4), whereas multiple-frequency propagations require to reconstruct a whole set of volumetric planes in order to see the surfaces present at different depths in the hidden scene.

4.2 Computational aperture masks

Performing single-frequency reconstructions like the ones we propose with the information captured by a wide aperture (usual capture grids in NLOS imaging measure about 1x1 metres) cause our virtual camera to have a narrow depth-of-field, due to the ambiguities in the incoming light paths. As a result, only the surfaces that are placed at one particular plane are displayed in focus in a computed image, while the rest of objects present at different depths appear out-of-focus. To address this issue, we turn to a common imaging resource: pinholes. In conventional imaging setups, these are physical apertures placed before the camera sensor that allow to obtain images with a large depth of field, i.e., where elements placed in a wide range of depths in a scene appear in focus. This is due to its small diameter, which only allows a narrow light beam to pass through, so that the light that is emitted or reflected by each point of the imaged scene is captured by a small region in the sensor.

The main approach proposed in this work consists in making use of virtual versions of these pinholes to enlarge the depth of field in NLOS images, motivated by the reduction of the number of reconstruction planes needed to image, in focus, all the elements present in a scene.

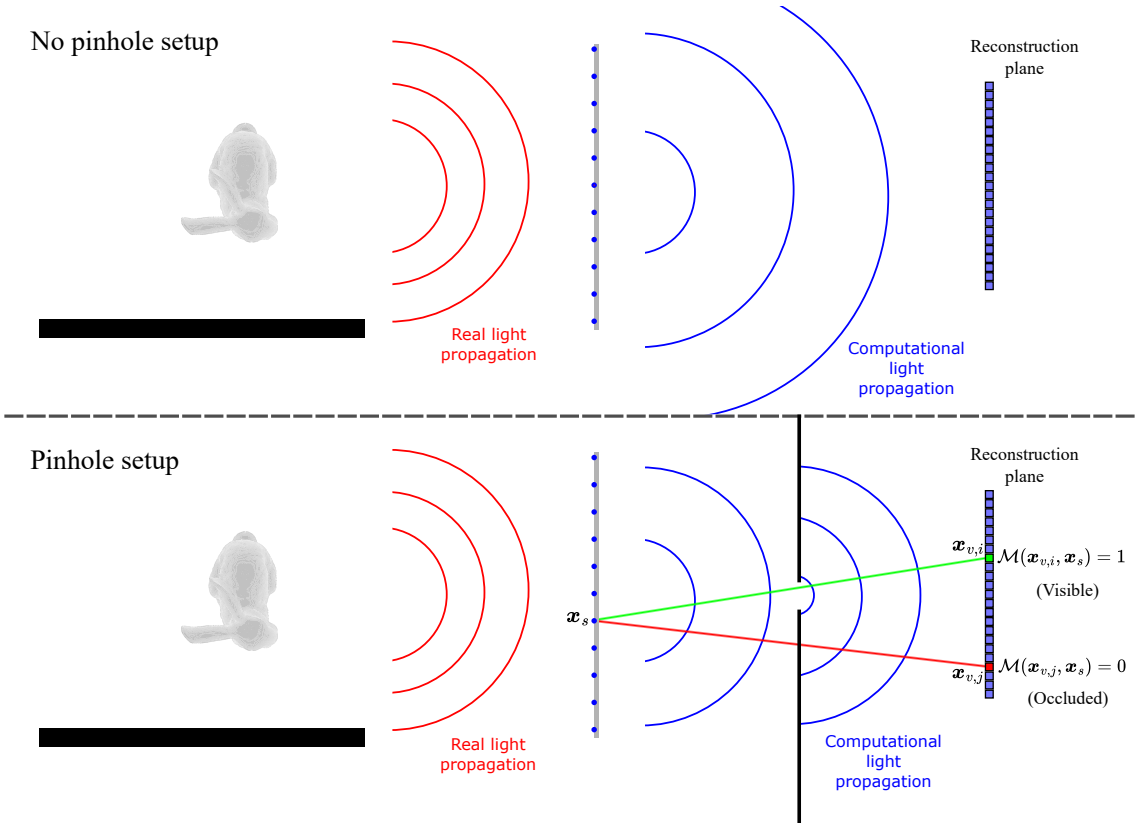

Figure 5: Top: visualization of the wave-based light transport that is usually performed in NLOS imaging. Bottom: we propose to place virtual occluders with pinholes during reconstruction time between the capture grid and the reconstruction plane to filter out-of-focus illumination effectively

This new NLOS imaging system is built as follows (see Figure 5): firstly we place, computationally, an infinite opaque plane parallel to the relay wall at a chosen distance from it. In this virtual occluder we can create a circular hole at any position and with any radius, which will let virtual light waves projected from the capture grid pass through to any reconstruction plane, giving place to diffraction effects similar to the ones caused by actual pinholes. This way, for every reconstruction voxel, only a small region of the capture grid (the points that are visible from the voxel through the pinhole) will contribute to its reconstruction, effectively removing out-of-focus illumination that could be introduced by other capture phasors. These occlusiones can be modeled with the following binary mask for every voxel-capture point pair:

$$\mathcal{M}(\boldsymbol{x}_v, \boldsymbol{x}_s) = \begin{cases} 1, & \text{if } \boldsymbol{x}_s \text{ is visible from } \boldsymbol{x}_v \text{ through the pinhole} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

and incorporated into Equation 6 to get the final propagation operator that we will be using in our reconstructions:

$$\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v) = \int_{\mathcal{S}} \frac{e^{ik|\boldsymbol{x}_s - \boldsymbol{x}_v|}}{|\boldsymbol{x}_s - \boldsymbol{x}_v|} \widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l) \cdot \widehat{H}(\boldsymbol{x}_l, \boldsymbol{x}_s, \omega) \cdot \mathcal{M}(\boldsymbol{x}_v, \boldsymbol{x}_s) d\boldsymbol{x}_s \tag{8}$$

In practice, this binary mask with the visibility information is computed the following way (3-dimensional case): let $\boldsymbol{x}_s$ and $\boldsymbol{x}_v$ be the positions of a point in the captured phasor field and a point located at the image plane, respectively, and let l be the line that contains these points. Assuming a coordinate system with its origin at the center of the relay wall where its Y coordinate represents depth in the scene, the x, z components of a point in l at depth d can be computed as:

$$\begin{cases} x = x_s + \dfrac{v_x}{v_y} d \\ z = z_s + \dfrac{v_z}{v_y} d \end{cases} \tag{9}$$

where $v \equiv (v_x, v_y, v_z) = \overrightarrow{\boldsymbol{x}_s \boldsymbol{x}_v}$. If the pinhole has radius r and the X, Z coordinates of its center are (p_x, p_z), the capture point $\boldsymbol{x}_s$ will be visible from $\boldsymbol{x}_v$ if $||(x, z) - (p_x, p_z)|| < r$ ($||\cdot||_\infty$ for square-shaped pinholes, $||\cdot||_2$ for circular ones).

In the following subsections, we present two different ways of incorporating virtual pinholes at the moment of computing the reconstructions: the first one consists in using one pinhole at a fixed position per image plane, while in the other its position varies with every voxel.

4.2.1 Fixed pinhole position per reconstruction plane

The first approach we propose to incorporate virtual pinholes to obtain NLOS reconstructions consists in using only one of these per reconstruction plane, in the following way (see Figure 6):

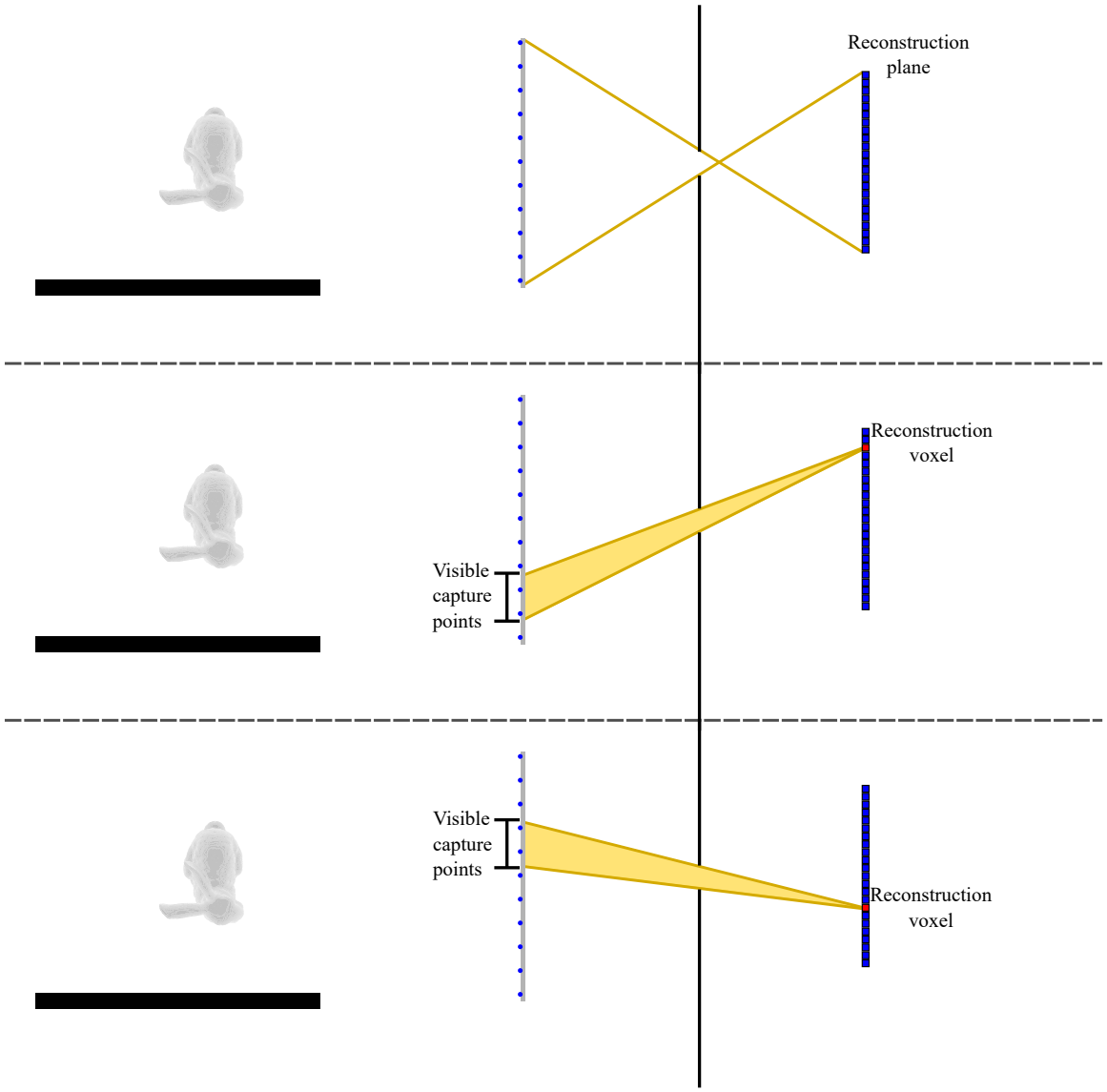

Figure 6: Top views of the resulting setup using one pinhole position per reconstruction frame. With this approach, every reconstruction voxel receives light contributions only from capture points that are not occluded by the virtual wall

For every plane in the scene we want to image, we can place an infinite virtual occluder wall with a hole that is centered with respect to the capture grid, between this and the reconstruction plane. In this virtual system, the actual reconstruction plane must be placed at a given distance beyond the pinhole (focal length), in order to leverage the focusing properties of this new element: as explained

previously, for every voxel in the reconstruction plane, only the computational waves corresponding to capture points that are visible through the pinhole will contribute to its image. As it is displayed in the figure, the boundaries of the reconstruction frame are not arbitrary, but determined by the field of view of the pinhole setup: we take the maximum lateral width and height that allow to receive contributions from any capture point from the relay wall. Voxels outside this frame will not display anything, since all the capture points would be occluded by the virtual wall.

The resemblance of this method to a pinhole camera allows us to leverage its focusing properties, in particular an increased depth of field, which makes it possible to obtain reconstructions of hidden scenes with a significantly lower number of images, since these will display elements in focus present in a wider depth range.

4.2.2 Multiple pinhole positions per reconstruction plane

In contrast with the method presented above, the position of the pinhole on the virtual occluder can be modified during the scene computations: instead of using a single pinhole position for every reconstruction plane, we propose another approach consisting in placing this element dynamically, right in front (same X, Z coordinates) of the voxel that is being imaged (see Figure 7).

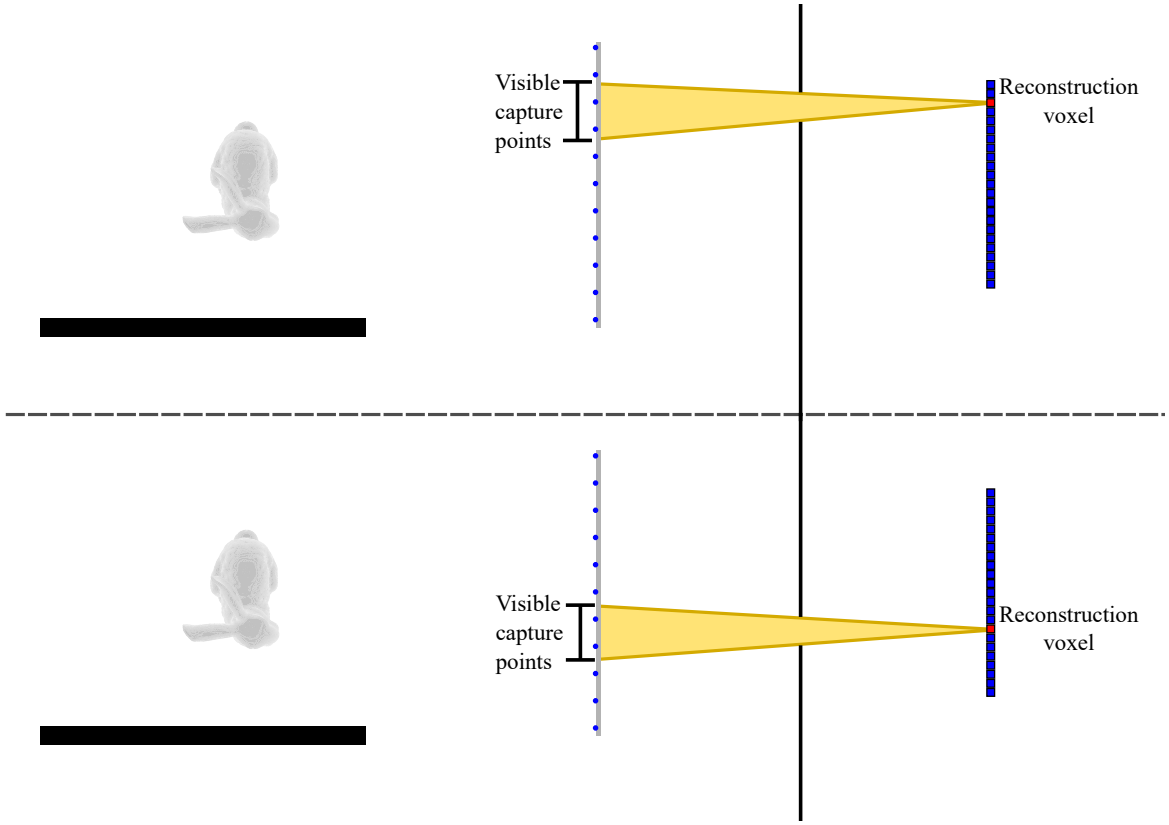

Figure 7: The second approach involving virtual pinholes that we propose consists in varying their position according to the voxel that is being processed, in order to gather more relevant capture contributions (tops view of the setup)

This way, the phasors that will be propagated and accumulated in relation to the voxel will correspond to its closest capture points at the relay wall, whose contribution is more relevant to its image formation than the rest. As we show experimentally, this variation that resembles confocal microscopy setups contributes to remove lateral artifacts that may arise in the images by using the single-position approach, since now the voxels that are far from the center receive more illumination and from capture points that contain more important information about their reflectance.

4.3 Proposed pipeline

We provide a pipeline to reconstruct the hidden scene using our single-frequency wave propagation through computational pinholes by imaging the hidden scene at a single or multiple image planes at different distances from the relay surface. Imaging through multiple image planes allows us to improve the quality of the reconstructions by constructing virtual focal stacks of the hidden scene, mitigating out of focus effects that may still be present in individual images despite our pinhole aperture. For this we take inspiration from light-field photography, which is able to capture and process a focal stack of a hidden scene to obtain all-in-focus images while using light-efficient large apertures.

In the first place, a NLOS dataset $H(\boldsymbol{x}_l, \boldsymbol{x}_s, t)$ is needed. This also includes the positions of the laser device and the recording sensor, as well as the positions in the relay wall where these aim at and the irradiance recordings.

One of our contributions consists in the implementation of a simulator that is able to generate single-frequency datasets from any 2D or 3D virtual scene defined by us, to which any wave-based method can be applied in order to obtain images of it. The application of the reconstruction methods proposed in this work to these datasets will serve as a proof of concept to test the performance and efficiency of the algorithms and compare them with other single-frequency approaches. Besides removing the need to rely on actual datasets or transient rendering engines, this tool provides us with the possibility to explore the whole parameters space: from the size and density of the captured phasor field in the relay wall to the shapes and depths of the objects placed in the virtual scene. By defining the positions of these, along with the ones of the point emitter source and the capture grid, and making use of the tools presented in Section 3.2, we can emulate virtually the way light waves propagate between them to generate single-frequency NLOS datasets (see Figure 8):

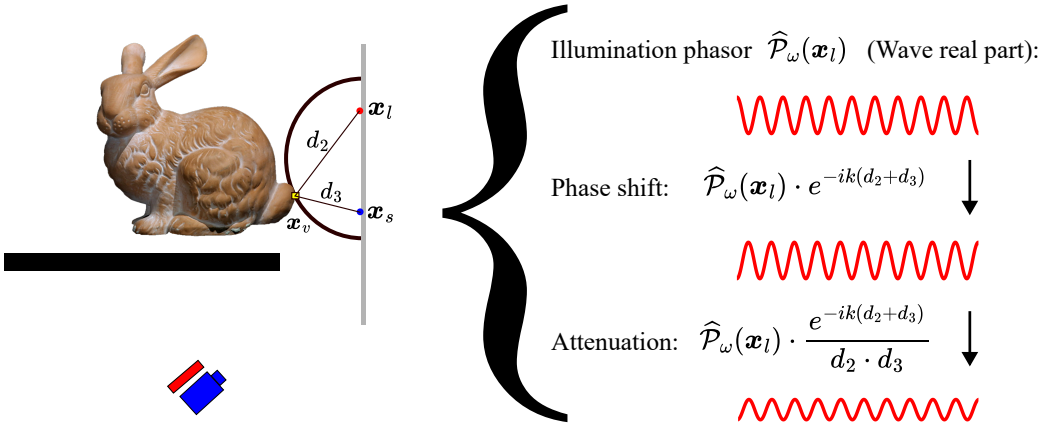

Figure 8: For every point in the virtual scene that contributes to the data accumulation, the initial illumination phasor experiences a phase shift and decay that depend on the distances between the illuminated point, the scene point and the capture point

Let $\boldsymbol{x}_l$ be the virtual point that represents the position that is illuminated with the laser device, and $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l)$ the phasor containing the emission information relative to an emission frequency w (amplitude and phase). Then, by applying Equation 3 twice, the phasor at a position $\boldsymbol{x}_s$ on the relay wall that accumulates all the irradiance contributions from the scene V can be computed as:

$$\widehat{H}_\omega(\boldsymbol{x}_l, \boldsymbol{x}_s) = \int_V \widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l) \frac{e^{-ik(|\boldsymbol{x}_l - \boldsymbol{x}_v| + |\boldsymbol{x}_v - \boldsymbol{x}_s|)}}{|\boldsymbol{x}_l - \boldsymbol{x}_v| \cdot |\boldsymbol{x}_v - \boldsymbol{x}_s|} d\boldsymbol{x}_v \tag{10}$$

This operation accounts for both the phase shift the initial phasor $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l)$ undergoes to all the points in the scene and back, and the linear decay in irradiance. In practice, we will use the equivalent formula for scenes composed of a discrete set of points $V = \{\boldsymbol{x}_{v,1}, ..., \boldsymbol{x}_{v,n}\}$:

$$\widehat{H}_\omega(\boldsymbol{x}_l, \boldsymbol{x}_s) = \sum_{i=1}^{n} \widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l) \frac{e^{-ik(|\boldsymbol{x}_l - \boldsymbol{x}_{v,i}| + |\boldsymbol{x}_{v,i} - \boldsymbol{x}_s|)}}{|\boldsymbol{x}_l - \boldsymbol{x}_{v,i}| \cdot |\boldsymbol{x}_{v,i} - \boldsymbol{x}_s|} \tag{11}$$

Note that if the data is generated this way, there is no need to apply an illumination function in the propagation operator from equation 6, since it already contains only single-frequency information.

Once a NLOS single-frequency dataset is available, obtained either by our generation method or resorting to captured or simulated datasets, we can apply one of the approaches elaborated in Section 4.2 to every reconstruction plane P_i: for every voxel $\boldsymbol{x}_v$ in it, the capture points $\boldsymbol{x}_s$ that are visible from it through the aperture are determined according to the chosen pinhole variant (fixed or multiple position), obtaining a binary mask $\mathcal{M}(\boldsymbol{x}_v, \boldsymbol{x}_s)$ with the same dimensions as the capture grid containing the information of these occlusions. The phasor $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v)$ at every reconstruction voxel is computed by propagating and accumulating the phasors corresponding to visible capture points using Equation 8, and then the image I_i of the whole plane is obtained by applying the desired imaging function $\Phi(\cdot)$ to all the phasors corresponding to the voxels in it. If multiple image planes are computed, they can be merged into a single image of the whole scene afterwards by applying max-projection or a focal blending technique $\mathcal{B}(I_1, ..., I_n)$.

A summarised version of the general method is shown in algorithm 1.

Algorithm 1 Proposed pipeline

Input: NLOS dataset $H(\boldsymbol{x}_l, \boldsymbol{x}_s, t)$, reconstruction planes $P_1, ..., P_n$
Output: Focal stack of images $I_1, ..., I_n$, merged image I
  for i = 1, ..., n **do**
    for voxel $\boldsymbol{x}_v \in F_i$ **do**
      Determine $\mathcal{M}(\boldsymbol{x}_v, \boldsymbol{x}_s)$
      Compute $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v) = \int_{\mathcal{S}} \frac{e^{ik|\boldsymbol{x}_s - \boldsymbol{x}_v|}}{|\boldsymbol{x}_s - \boldsymbol{x}_v|} \widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l) \cdot \widehat{H}(\boldsymbol{x}_l, \boldsymbol{x}_s, \omega) \cdot \mathcal{M}(\boldsymbol{x}_v, \boldsymbol{x}_s) d\boldsymbol{x}_s$
    end for
    Obtain plane image: $I_i(\boldsymbol{x}_v) = \Phi(\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v))$
  end for
  Merge the images: $I(\boldsymbol{x}_v) = \mathcal{B}(I_1(\boldsymbol{x}_v), ..., I_n(\boldsymbol{x}_v))$

5 Results

As elaborated throughout this work, our main goal is to obtain NLOS reconstructions that turn out more efficient that the ones achievable with usual phasor fields-based methods, which require the processing of multiple frequencies and planes. Since plain single-frequency reconstructions display plenty of out-of-focus effects due to their shallow depth of field, we have incorporated computational photography elements (pinholes) to them in order to address this issue, allowing us to obtain both efficient and in-focus reconstructions. In this chapter we compare the results that can be obtained by applying the tools proposed in Section 4 against the ones yielded by both single-frequency and multiple-frequency phasor-based methods that do not make use of the proposed pinholes, over one and multiple reconstruction planes. These will include not only the resulting images of the hidden scenes, but also the execution times to compare the computational efficiency of all of them under the same equipment. For the comparisons with the raw single-frequency approach, the methods have been applied to data generated with the simulator presented in Section 4.3, computed by reproducing actual 2D and 3D NLOS setups: a point emitter source $\boldsymbol{x}_l$ that represents the location illuminated by the laser device is placed at the center of the relay wall, and then, given a fixed virtual frequency ω, the phasor field that represents the response of the scene is computed in a grid of points S at this relay wall by accumulating single-frequency propagations. Every scene is represented by a set of spatial delta points, placed separately or in groups to emulate volumetric objects, that act as scatterers and over which the third-bounce light reflections (Equation 11) are computed.

As for the comparisons against multi-frequency methods, we have employed publicly-available NLOS datasets obtained with transient rendering engines [GMO+19]. As a summary, we show that:

- Our single-frequency simulator is able to reproduce the phasor fields that are captured at the relay wall.
- In simulated 2D and 3D scenes, the use of a fixed pinhole during single-frequency propagations allows us obtain a larger depth of field, yielding better reconstructions of the hidden scene than single- or multi-frequency approaches with a single plane. As a result, we reduce the computation times by halft at least, while obtaining better quality pictures. If our fixed pinhole method is not able to image the elements present within the whole depth range with a single reconstruction plane, we can apply the proposed pipeline to obtain an image of the whole scene with a low number of computed planes.
- In publicly-available datasets displaying surfaces in a wide range of depths, we are able to image all of them with a single reconstruction plane with the proposed multiple-position pinhole approach, as opposed to single and multiple-frequency methods without pinhole, which require to reconstruct several planes in the scene. Again, besides removing out-of-focus illumination, we are able to reduce the computation times significantly.
- We show that our movable pinhole method improves the visibility of the surfaces that are present in the missing cone of multi-frequency pinhole-free NLOS imaging methods.

To obtain the best possible resolution in all comparisons, we follow the well-established Rayleigh criterion, which defines the minimum angular resolution of an imaging system as directly proportional to the imaging wavelength, and inversely proportional to the imaging aperture size, with the following expression:

$$\theta = 1.22\lambda/D, \tag{12}$$

where θ represents angular resolution, λ represents the imaging wavelength, and D represents the aperture size. This criterion also applies to NLOS imaging systems [OLW18], [LGLM+19], where λ is the virtual imaging wavelength, and D is the size of the captured relay surface. To achieve the best reconstruction resolution possible, we minimize the imaging wavelength based on the sampling rate of our relay surface: theoretically, for discretely sampled apertures, the minimum imaging wavelength must be equal or longer than two times the minimum distance between sampled points $\boldsymbol{x}_s$, We follow this rationale to choose the imaging frequency $\omega = 1/\lambda$ in all our results. For multi-frequency comparisons, we follow this criterion to choose the central wavelength of the illumination function (Equation 2).

As for the implementation of all the compared algorithms (multi-frequency, single-frequency, and ours), it has been carried out in the MATLAB platform, which has allowed us to leverage optimised array operations during the virtual wave propagations and the computation of occlusions in the proposed pinhole setups. These are able to perform multiple component-wise computations at the same time that turn out much more efficient than iterative loops, in line with the motivation of this work. All the results we display have been computed with the same PC, equipped with a 11th Gen Intel(R) Core(TM) i5-11300H @ 3.10GHz processor.

5.1 2D analysis

The first experiments to test our fixed pinhole method have been performed in 2D scenes, in order to check the viability of the proposed setup in the simplest system. The goal of the initial tests, conducted by simulating scenes composed only point scatterers at different depths, is to find if these are resolvable with our fixed pinhole method and to which extent, and the advantages it can offer with respect to propagations without the use of a pinhole. These tests have been performed in simulation, as follows: in the first place, we define the positions of the capture grid, containing 500 measure points spread along a 1 metre long virtual relay wall, and the point that is illuminated virtually at its center (see Figure 9), resembling actual NLOS capture setups. Then, we place a set of 5 points that will act as the reflector objects in the hidden scene, spread along width and depth as illustrated in the figure.

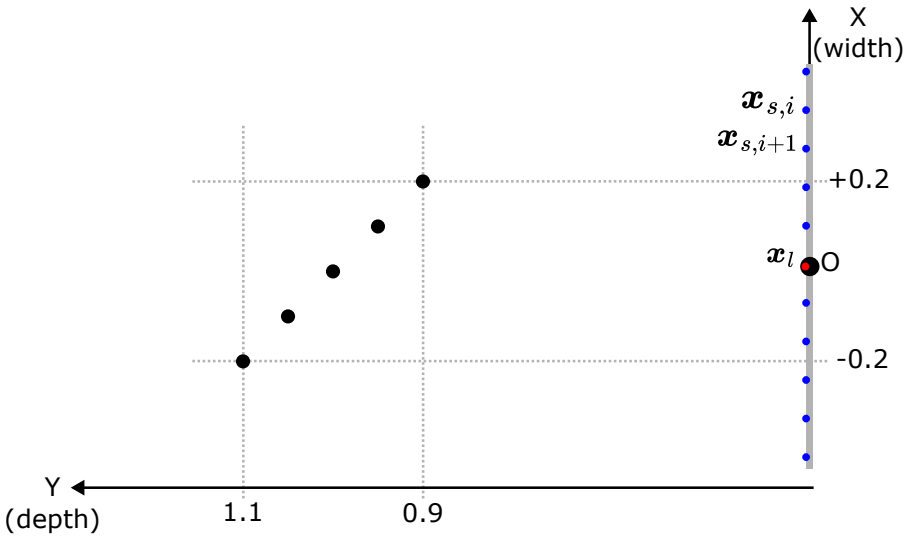

Figure 9: Diagram with the disposition of the virtual setup in our 2D scene

To get the single-frequency response of this scene to an illumination phasor $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_l)$ with a wavelength of 3 mm, we turn to the procedure explained in Section 4.3 (Equation 11): given its initial (arbitrary) value, we compute for every capture point the accumulation of the third bounce reflections of every point present in the scene, obtaining $\widehat{H}_\omega(\boldsymbol{x}_l, \boldsymbol{x}_s)$. This data will be the input to the single-frequency reconstruction algorithms we will be employing.

Single-frequency imaging. As a reference of what can be achieved with state-of-the-art single-frequency propagations with no pinhole, we compute a focal stack of 5 images that display the scene as seen from the relay wall (see Figure 10), each one obtained by applying the RSD propagator from Equation 6 to the voxels that form 5 physical planes, placed behind the relay wall at the same distances as the points in the scene (mirrored with respect to the relay wall), and taking the square of the modulus of the resultant phasors $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v)$ (we are applying the imaging function $\Phi(\cdot) = |\cdot|^2$, see Section 3.2) for visualization. As elaborated in Section 4, the computation of several reconstruction planes is necessary in order to see the objects that are placed at different depths, due to the narrow depth of field of this method. In figure 10, bottom, we display the resultant 1D images obtained in these planes, where the abscissae represent the X coordinates in the reconstruction plane, and the ordinates the value of the

imaging function Φ applied to the computed phasors $\widehat{\mathcal{P}}_\omega(\boldsymbol{x}_v)$ in each voxel position $\boldsymbol{x}_v$. As it can be appreciated, each point is clearly visible in its corresponding image plane, but note that in all cases the rest of objects present at different depths appear out-of-focus.

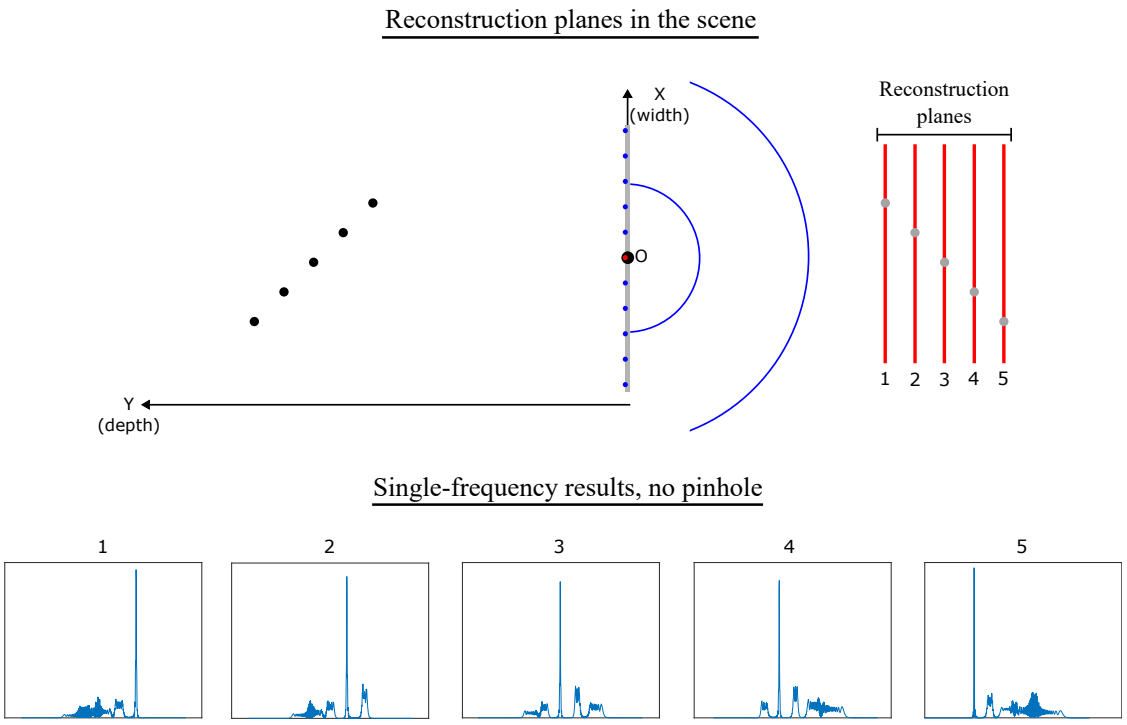

Figure 10: Top: the reconstruction planes are placed at the same depths as the elements present in the 2D scene, symmetrically with respect to the relay wall. Bottom: 1D images that are obtained applying regular single-frequency phasor propagations to the planes displayed above

Fixed pinhole imaging. In Figure 11, top, we perform a similar experiment but adding a virtual occluder wall with a circular pinhole in its center, with the same radius as the employed wavelength, as proposed in Section 4.2.1, and propagate for every voxel only the capture points that are visible through the pinhole according to Equation 8, it is possible to obtain an image where all the points appear in focus, regardless of their depth in the scene. In this case, what we get is an image with a large depth of field that is focused beyond the virtual pinhole, but shows in focus the objects that are placed at the central depth of the scene, as a conventional camera does. For this, we do not place the reconstruction plane at its corresponding distance from the relay wall as in the previous case, but behind the occluder wall that must be put at the double of distance than the plane of which we want to obtain a focused image. It is worth mentioning (Figure 11, bottom) that if no pinhole is used, the propagations to the very same plane using Equation 6 only amount to out-of-focus illumination. This single-plane reconstruction took 0.47 seconds to compute, as opposed to the 3.15 seconds that were needed to obtain the previous focal stack composed of 5 pictures using propagations without the use of a pinhole, using the same equipment. Therefore, with some previous knowledge about the depths in the scene we are able to obtain a fair representation of it in approximately 15% of the time using the proposed fixed-pinhole approach.

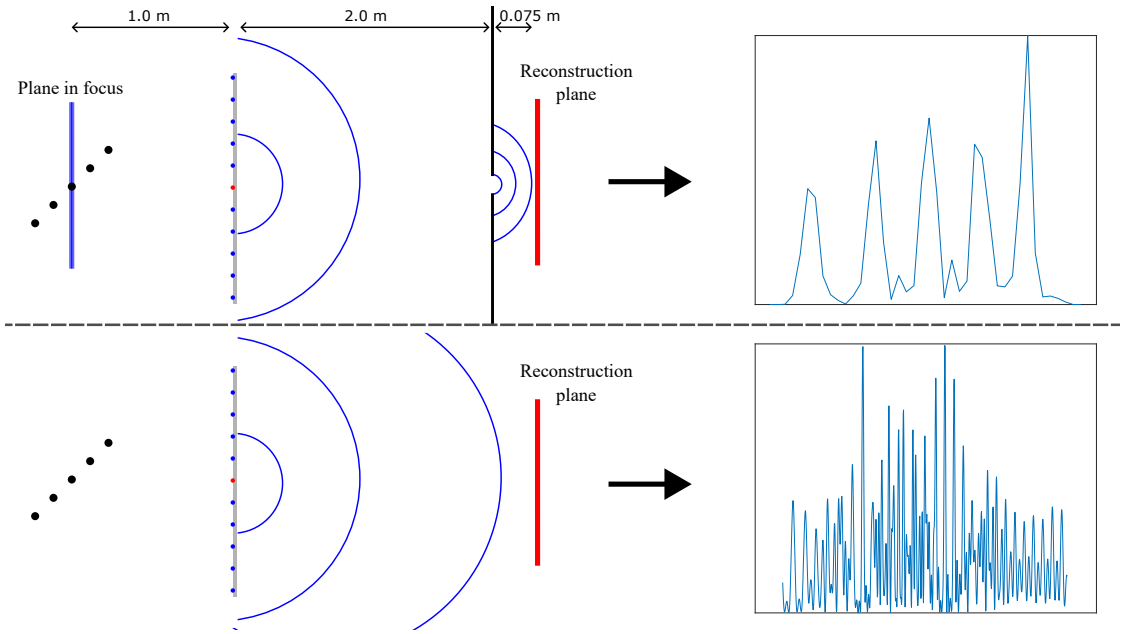

Figure 11: Top: if the phasor propagations pass through a pinhole, we are able to focus at the center of the hidden scene and get a (1D) image displaying all its elements with a single reconstruction plane. Bottom: if the illumination is not focused with our virtual pinhole, the same propagations lead to a completely out-of-focus image of the scene

5.2 3D analysis

In this section, we present the results that can be obtained using the proposed pinhole approaches with 3D scenes, and highlight the advantages of applying it in comparison with existing single and multi-frequency wave-based methods that do not use virtual pinholes. In particular, we show that we can save time and computational resources in the reconstructions, besides dealing with out-of-focus artifacts that appear when no pinholes are employed. All the pictures that are showed have been obtained using a 64x64 pixels resolution, and as in all NLOS imaging methods, these will display the scene as seen from the relay wall.

5.2.1 Validation of the generated data

As a way to validate the method proposed in 4.3 to generate single-frequency data, a planar scene from an official dataset [GMO+19] containing a letter "Z" placed 1 metre away from the relay wall has been has been replicated with our simulator using several volumetric segments. These segments are composed of individual points separated by 1 cm, the same distance as the wavelength of the virtual waves that have been propagated. To replicate the capture setup of the original dataset, we firstly build a virtual relay wall where its central point will act as the illuminated point and define the position of the capture grid, composed of 256x256 points, equally spaced along a centered square with an area of 1 m^2. By applying Equation 11 to all these points, we are able to compute the response of our virtual designed scene to a constant source of illumination in the chosen frequency (1 cm wavelength), accounting for the third bounce reflections between from the illuminated point to the area object in the scene and back to the relay wall.

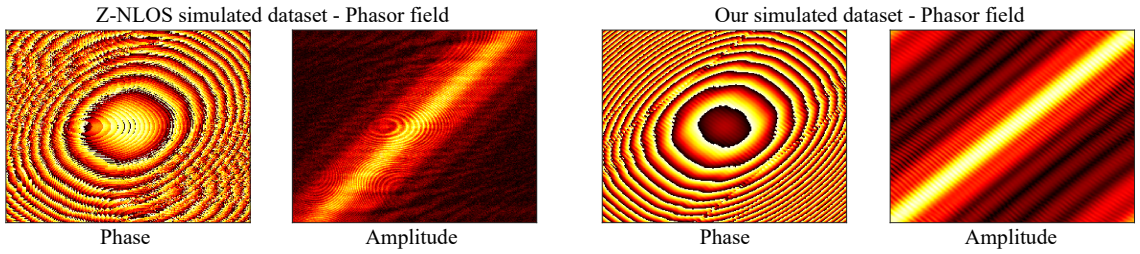

Figure 12: Phase and amplitude of the phasor field captured in the original "Z" dataset and the one generated with our method

Since the original dataset was captured with a temporal delta illumination function, it contains the phasor fields relative to all frequencies, and therefore we must filter the information regarding the frequency corresponding to a wavelength of 1 cm for a fair comparison. This can be done by convolving every captured signal with a sinusoidal complex-valued function with this frequency, as stated in 3.2, or by performing a Fourier transform and selecting the coefficients that correspond to the wanted frequency. In Figure 12, we show heatmaps with the phase and amplitude of the phasor field corresponding to a wavelength of 1 cm that was captured in the relay wall of the original setup (left), along with the one computed with our simulator (right). It can be seen that they match for the most part (except for some small differences caused by the influence of the top and bottom segments of the letter and higher-order light bounces captured in the dataset), which supports the method we have followed to obtain third-bounce single-frequency data.

5.2.2 3D reconstructions - Our simulated data

We have leveraged our single-frequency data generation algorithm to simulate the response of several 3D scenes, in order to apply and test the proposed fixed-position pinhole approach and compare the results to those that can be obtained with regular phasor fields propagations, both visually and efficiency-wise. We also include an example of execution of the full proposed pipeline from Section 4.3 that displays how we can obtain an efficient NLOS reconstruction when one single picture is not enough to image the whole scene due to the presence of objects in a wide range of depths.

In Figure 13, we show a diagram with the simulated scene we will be working with; it is composed of several segments placed at different positions and depths, which will allow us to appreciate the advantages of our methods regarding the depth of field in the images and the out-of-focus artifacts that may arise in the reconstructions.

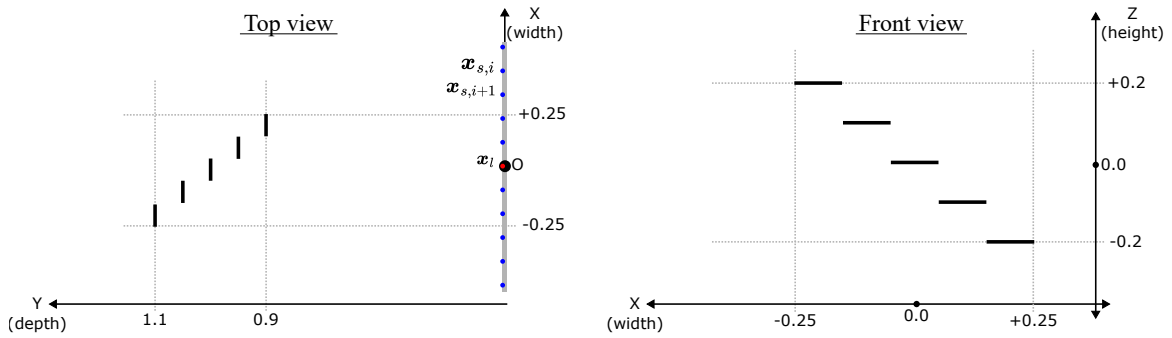

Figure 13: Disposition of the segments composing our virtual scene, seen from both above (left) and from the relay wall (right)

We start by defining the positions not only of the objects in the scene, but also of the elements on the relay wall: the illumination point and the capture grid. In line with official NLOS datasets, we set 256x256 measure points over a 1 m^2 square area on the virtual relay wall, and place the point

that is illuminated at its center. To get the response of the scene to the illumination source, we employ Equation 11 to account for the third-bounce illumination that reaches the capture points after illuminating the scene virtually from the center of the relay wall, with a phasor emitting waves with a 6 mm wavelength. Now we can apply the propagation models introduced in Sections 4.1 and 4.2 to these single-frequency data over the voxels defined on one or more reconstruction planes to obtain the phasors containing the information about the illumination that was reflected by the scene. In order to visualise this information, we display images containing heatmaps that show the square of the amplitude of each propagated phasor, which allows us to effectively see the reflections of the objects in the scene.

Single-frequency imaging. As seen in the previous 2D example, when applying single-frequency propagations without a pinhole using Equation 6, a different reconstruction plane is needed to see every object in the scene (Figure 14), and even then, the out-of-focus illumination reflected by other objects in the scene degrades the images greatly due to the shallow depth of field of this method.

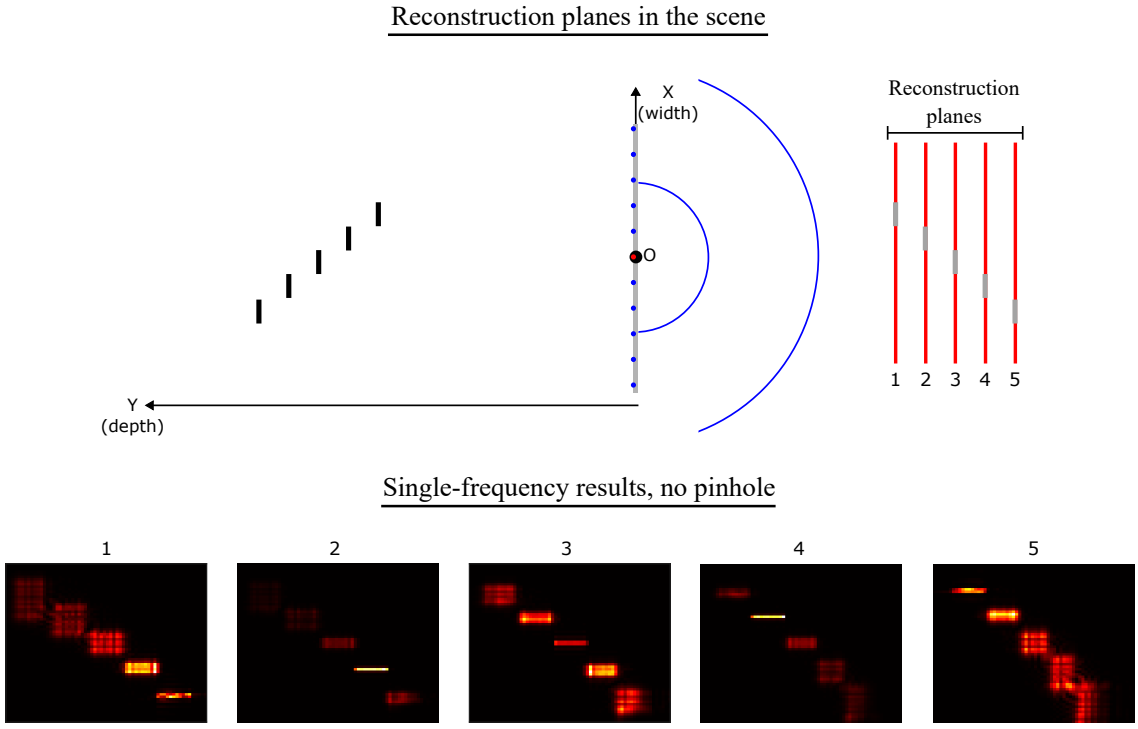

Figure 14: Top: Zenithal view of our virtual segments scene, and disposition of the reconstruction planes to obtain phasor fields results without a pinhole. Bottom: Images of the scene obtained using regular phasor field propagations to the pointed planes

Fixed pinhole imaging. In figure 15, top, we configure an experiment where we incorporate a virtual pinhole to filter the propagations employing Equation 8, we can focus our virtual camera at the central depth in the scene and leverage its depth of field to image all the objects with a single reconstruction plane. Again, if the reconstruction is made in this plane without filtering through the pinhole by applying Equation 6, the results are intelligible (Figure 15, bottom).

Although each individual plane reconstruction takes some more time to be computed with the use of a pinhole (3 times more, approximately), the extended depth of field provided by it allows us to reduce the number of reconstruction planes significantly, which ends up being faster and delivers the additional benefit of reducing out-of-focus artifacts. As practical proof, the computation of the 5 no-pinhole images was carried in 34.6 seconds in total, whereas with our pinhole method the single-plane reconstruction took 18.99 seconds, almost half the time.

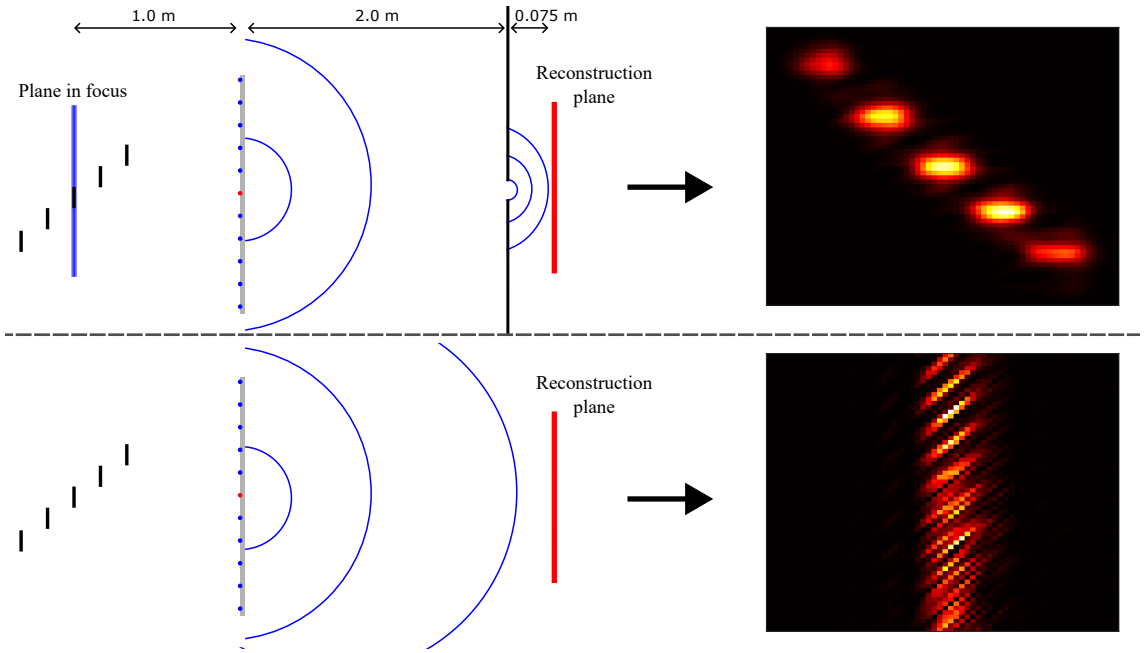

Figure 15: Top: Virtual fixed pinhole setup and image of the 3D segments scene that is obtained with it, over a single reconstruction plane. Bottom: The same reconstruction plane without the use of a pinhole leads to a completely out-of-focus image

We also show an example of the results that can be achieved by applying the full pipeline proposed in 4.3 (see Figure 16), which is meant to be applied when we have no prior knowledge of the hidden scene or the objects are placed along a wide range of depths: instead of applying our fixed pinhole method to a single reconstruction plane like in the previous examples, we compute several (two) of them, each one focused at a different depth in the scene, and merge the resultant images to obtain a single image that displays all the scene objects in focus. To this end, consider the previous segments scene but more spread along depth (the segments are now evenly separated between distances 0.8 and 1.2 metres from the relay wall), so that it cannot be captured in a single-plane image even with our fixed-pinhole method. In this case, we need at least two reconstruction planes focused at different depths to cover the full scene, which can be merged afterwards to display everything in focus. Each one of these can be obtained by following the same steps that were followed in the previous fixed-pinhole reconstruction: we start by defining the positions of all the elements in the scene (capture grid, illuminated point and virtual segment objects) and computing response of the scene to a single-frequency illumination phasor emitted from the center of the relay wall by accounting for all third-bounce propagations between the emission point, the scene and the capture grid. Then, the propagations to two reconstruction planes are computed one by one as before: by placing an occluder wall with a fixed-position pinhole at the double of distance than the plane we would like to focus on and the reconstruction plane behind it, and accumulating the contributions from visible capture points for every voxel in the planes (Equation 8). In this case, we use relay wall-pinhole distances of 1.8 and 2.2 and focal lengths of 0.07 and 0.08 to to focus our virtual camera at depths of 0.9 and 1.1 metres in the scene.

Like in most of NLOS imaging methods, the afterwards merging is performed by means of the max-projection approach discussed in 3.3, after a normalization of every individual image. It must be highlighted that with the use of any other NLOS imaging method and even with prior knowledge of the scene, at least 5 planes, one for every object at a different depth in the scene, would have been necessary to obtain the same merged image of an scene that is spread this much along depth.

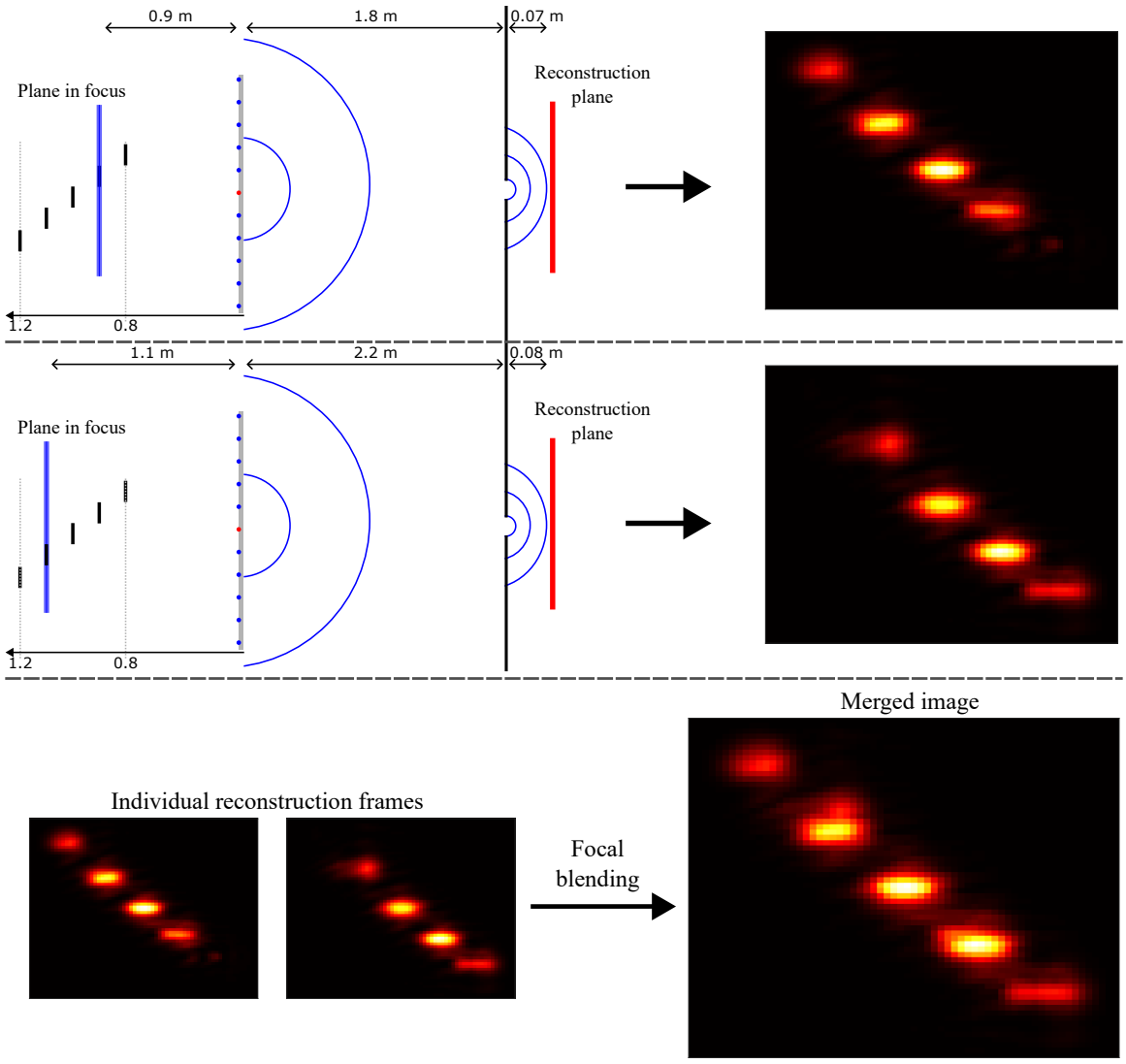

Figure 16: Top, center: by placing the virtual occluder at different depths beyond the relay wall, we are able to obtain in-focus images of the objects at different depths in the scene, with an extended depth of field. Bottom: the images obtained with our method can be merged with max-projection to obtain a final result and final result where all the scene is in focus

5.2.3 3D reconstructions - Actual datasets

Here we present the results that can be obtained by applying some of the techniques discussed in this work to official NLOS datasets that were generated using transient rendering engines. This time, we compare quality and efficiency of bare single and multi-frequency propagations to our multiple-position pinhole approach: the relevance of this method lies in the fact that the pinhole placed on the virtual occluder varies with the position of the voxel that is being imaged every time, to accumulate more relevant contributions from the capture grid. As a result, we are able to image the whole scenes with a single reconstruction plane.

We will be showing two scenes: one containing two vertical planes at different depths ("preoc-clusion") and another one with two spheres ("spheres"). In both cases, the objects are placed at two different depths at 0.4 meters from each other. In Figure 17, a rendered view of the scenes can be seen, accompanied by NLOS images of them that can be obtained using multi-frequency phasor fields implemented using an iterative backprojection-based solver as in the original paper by Liu et al. [LGLM+19], over a volumetric area composed of two image planes (where the scene objects are placed) plus a max-projection focal blending. Note that, in order to obtain these without prior knowledge of

the scene, plenty of reconstruction planes should be computed to cover all the hidden geometry.

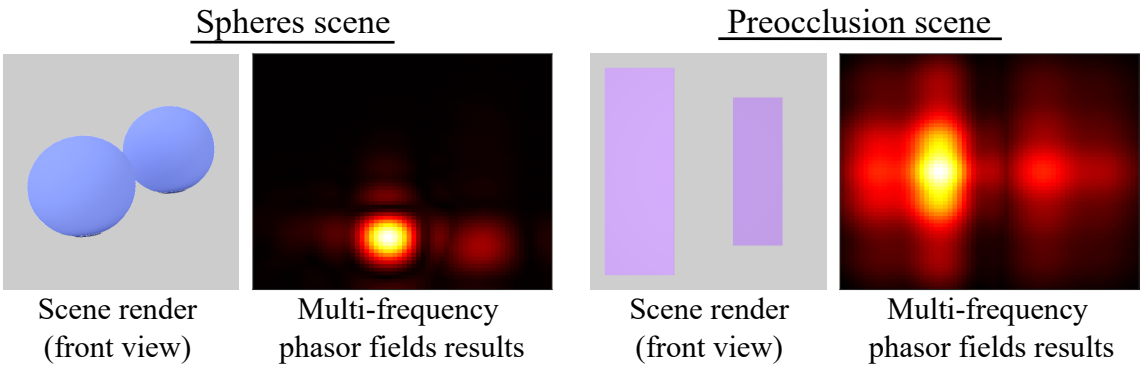

Figure 17: NLOS scenes chosen to show the performance of our method and references of the results that can be obtained with more computationally expensive methods

These datasets contain the whole temporal measurements captured after illuminating the scene with a light pulse emitted to the center of the relay wall, and therefore provide the information relative to all virtual frequencies. In order to obtain the phasors in a sole frequency, one must filter the signals with a single-frequency illumination function, as elaborated in Section 3.2. For the results of this part, we will be working with the information carried by computational waves with a wavelength of 13 cm. After the filtering, the RSD propagators presented in Sections 4.1 and 4.2 can be applied to them to obtain the desired scene reconstructions. Note that in this case, the capture grid is known beforehand, and it is composed of 256x256 points evenly distributed over a 1x1 metres relay wall, with the illuminated point in its center.

Single-frequency imaging. Once more, we compute firstly the scene images using Equation 6 (no pinhole), for which the reconstruction planes must be placed at the depths of the scene where we want to display the objects in focus, mirrored with respect to the relay wall as in the previous experiments. In figure 18, we show this setup with the "spheres" scene, where the planes in focus correspond to the front part of the objects.

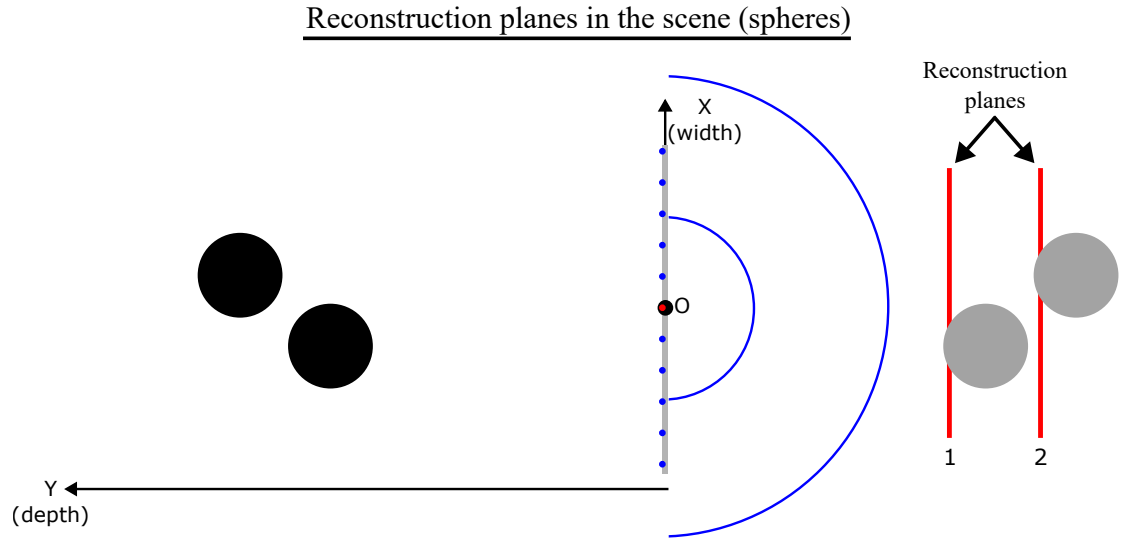

Figure 18: In order to obtain images of the spheres in the scene using single-frequency propagations with no pinhole, we place the reconstruction frames behind the relay wall, at the same distance as the objects in the scene

The results that are achieved by applying this approach to the presented scene are displayed in

Figure 19, top. At the bottom, we show the images obtained with the "preocclusion" scene analogously, by focusing our virtual camera at the planes where the two surfaces lie in the scene. Like it could be seen in the scenes simulated by us, the narrow depth of field of this method causes the inclusion of out-of-focus illumination reflected by objects that are present at a different depth that the one we focused at, degrading the parts of the image that are indeed in focus.

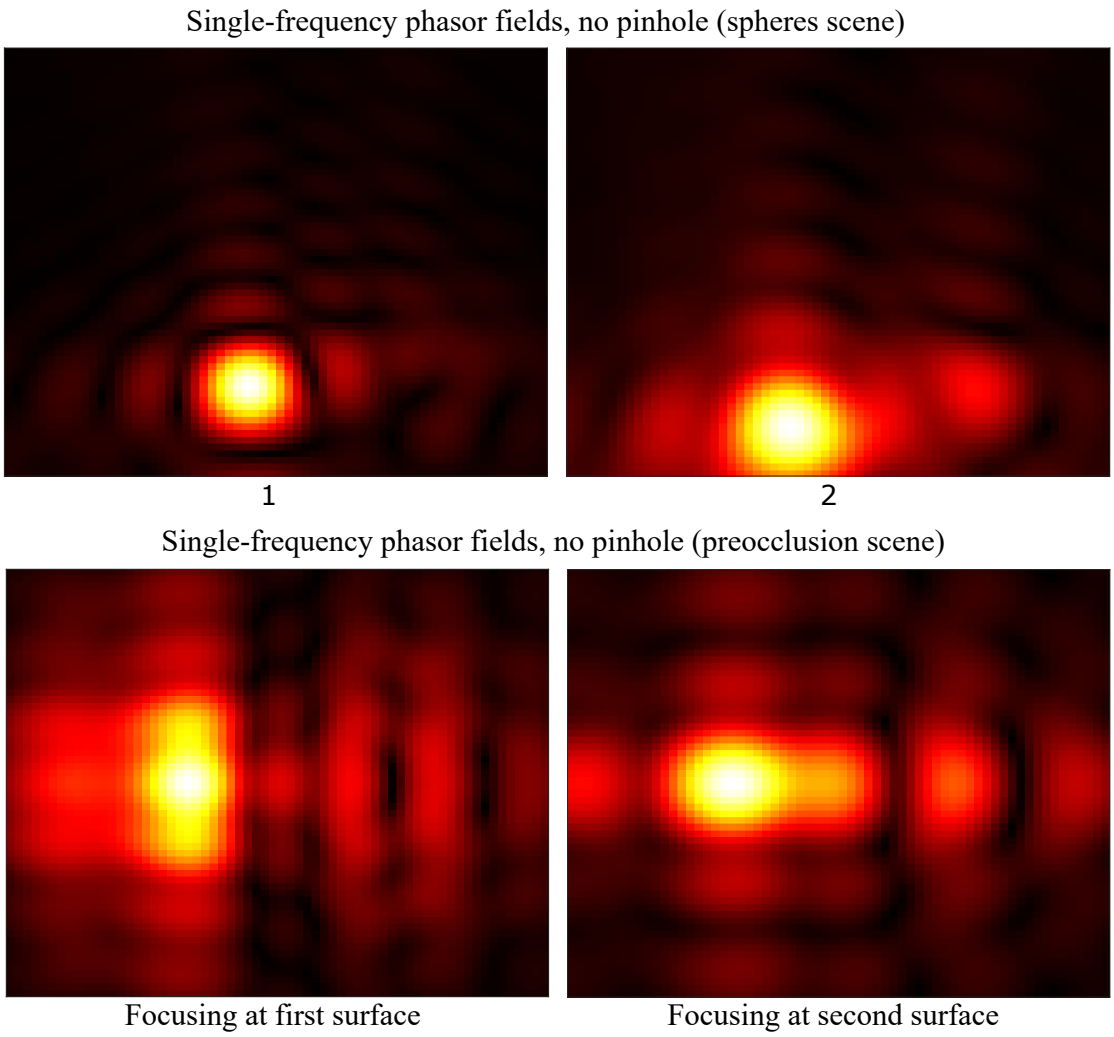

Figure 19: Top: Images obtainable by applying single-frequency propagations to the "spheres" scene without the use of a pinhole, by focusing at the front parts of the objects. Bottom: Pictures of the "preocclusion" scene, gotten by applying the same procedure to the planes where the surfaces lie in the scene

Movable pinhole imaging. We show how the results improve greatly with our movable pinhole method, by applying Equation 8 to pinholes whose position changes depending on the imaged voxel, as shown in Figure 20. Experimentally, we have been able to obtain single-plane images of both scenes where the elements at different depths are visible by placing the occluder wall at any distance d from the relay wall, provided that the reconstruction plane is located at this same distance beyond it. Since different values of d do not lead to different pictures, we are able to image the hidden scene with a single reconstruction frame and no prior information about the depths at which the objects are placed. As for the size of the pinholes, the best results have been achieved by using a diameter of 1.1 times the propagation wavelength.

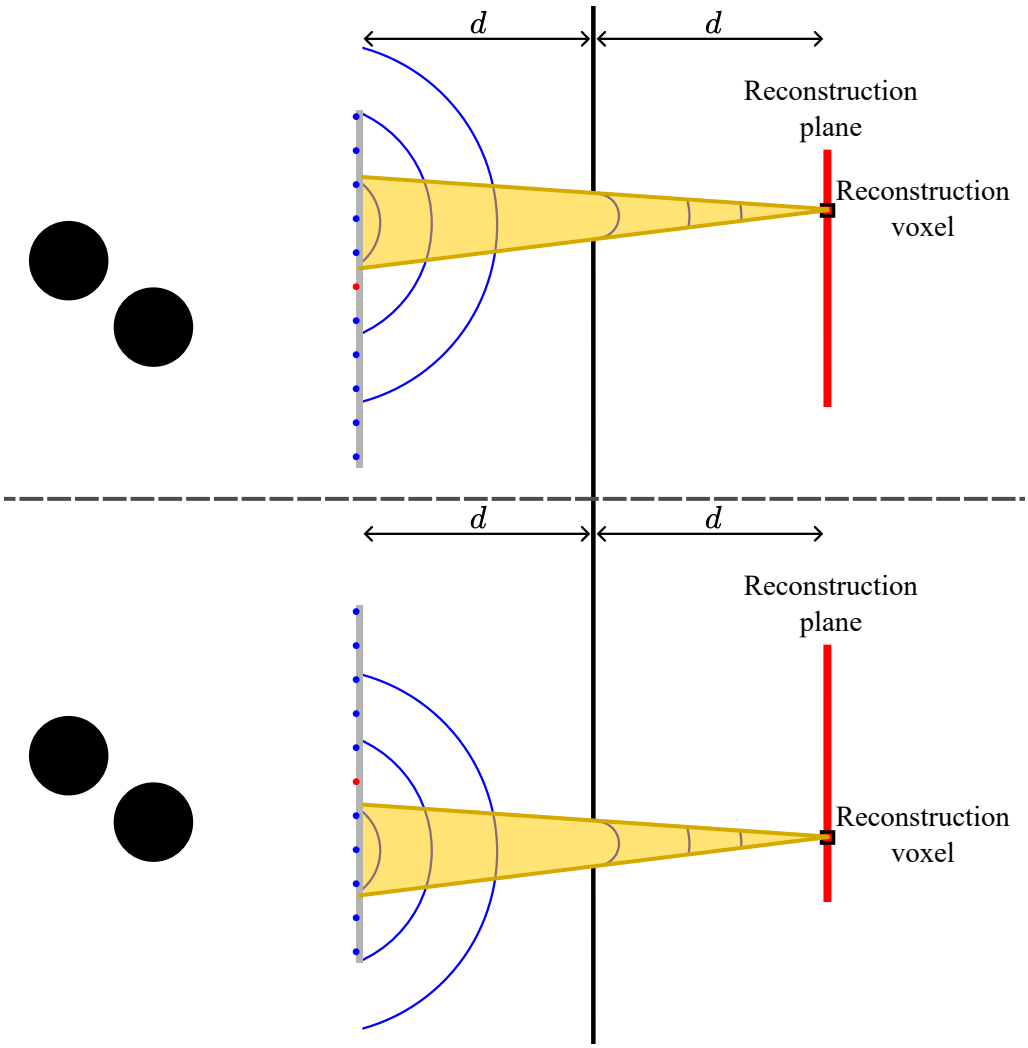

Figure 20: Top view of the virtual setup involving movable pinholes employed to image the spheres scene. In these reconstructions, only the phasors corresponding to capture points that are right in front of the reconstructed voxel are propagated

As it can be appreciated in Figure 21, our method grants an extended depth of field that makes elements placed at considerably distant depths visible with only one reconstruction plane, as opposed to bare propagations. It also displays parts of the objects that are not visible even with the phasor fields based multi-frequency approach (lateral parts of the spheres and outer zones of the planes) due to the missing cone problem. This is a phenomenon due to which certain surfaces cannot be reconstructed using any NLOS imaging method: the specular behaviour of the computational light waves that are used causes the reflections of some objects not to reach the capture grid at the relay wall, which affects the areas of these objects that are locally planar and whose specular direction with respect to the illuminated point in the relay wall does not point back to the capture grid.

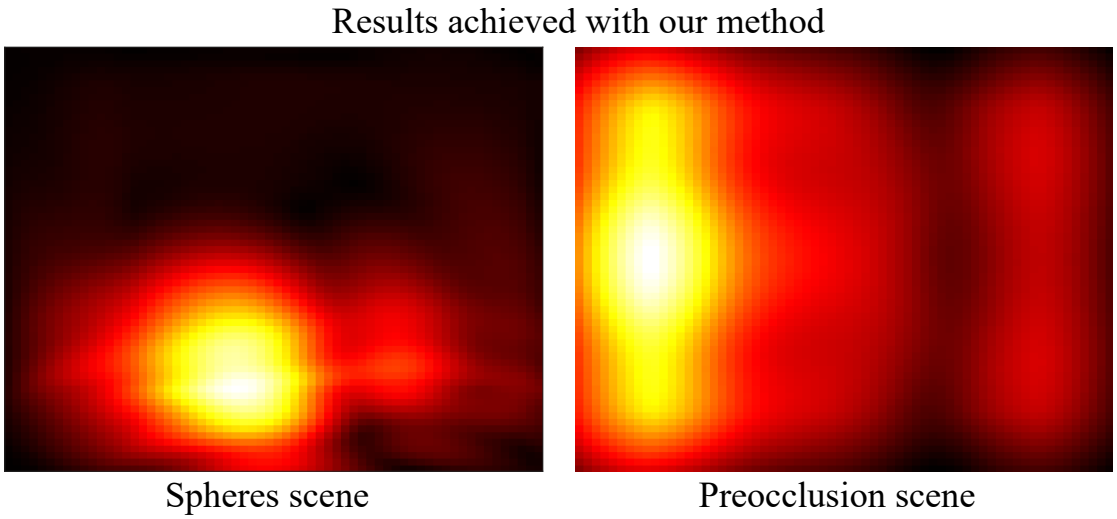

Figure 21: Single-plane reconstructions of the same scenes using the proposed multiple-position pinhole approach, with no previous knowledge of the position of the objects in the scene

Regarding the reconstruction times, each image obtained with our method took 22.90 seconds to compute, while each pair of reconstructions without employing virtual pinholes required 14.58. As for the latter, we only showed the planes where the elements are present, but in practice and with no prior knowledge of the scene, one would have to compute a dense focal stack of images, which in the end turns out much more costly: only 4 reconstruction planes would already surpass the reconstruction time with our method. As for the multi-frequency results shown in Figure 17, the execution times with the implemented backprojection algorithm turn out 1.5 orders of magnitude longer than our single-frequency pinhole-based method.

6 Conclusions and future work

In this work, we proposed a novel efficient phasor-based method for NLOS imaging that combines single-frequency virtual wave propagation and computational pinholes. This approach allows us to leverage the efficiency of single-frequency reconstructions while addressing its shallow depth of field, which constitutes their main issue. The extended depth of field achieved with our method makes it possible to image complete hidden scenes with less reconstruction frames, which turns out even more efficient with the additional benefit of dealing with out-of-focus artifacts.

Part of our methodology consisted in studying state-of-the-art formulations as well as their direct applications in actual imaging setups. This has allowed us not only to understand and implement them, but also to go beyond and find a new solution to improve current results:

- By understanding how computational waves propagate through space, we have been able to emulate their behaviour to create a single-frequency NLOS data simulator. This can serve as a proof of concept to test not only the proposed virtual pinholes, but any single-frequency method that requires experimentation and freedom to modify all the parameters in a test hidden scene.
- As elaborated throughout this report, our main contribution lies in incorporating virtual pinholes to single-frequency wave-based methods during reconstruction time. We presented and implemented two different approaches that increase the depth of field in every NLOS image, without entailing a high computational cost, allowing us to obtain cheaper and faster reconstructions.
- We compiled a set of scene reconstructions, both generated by us and from existing datasets, that display the advantages of the proposed methods against regular single and multi-frequency wave propagations. We improve the quality of regular single-frequency reconstructions with similar or shorter execution times, besides reducing considerably the reconstruction times with respect to multi-frequency reconstructions obtained with iterative backprojection.

To sum it up, our virtual pinholes methods address several relevant issues in current NLOS imaging: efficiency in the reconstructions and out-of-focus illumination, all of them limitant problems regarding practical and inexpensive setups. We hope the work presented here serves as a basis for future developments regarding these limitations to make the most of this computational imaging area.

References

[ADLT22] Sameer Agarwal, Timothy Duff, Max Lieblich, and Rekha R. Thomas. An atlas for the pinhole camera. *Foundations of Computational Mathematics*, 24(1):227–277, September 2022.

[AGJ17] Victor Arellano, Diego Gutierrez, and Adrian Jarabo. Fast back-projection for non-line of sight reconstruction. *Opt. Express*, 25(10):11574–11583, May 2017.

[BZT+15] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Opt. Express*, 23(16):20997–21011, Aug 2015.

[CdGB+23] Ruizhi Cao, Frederic de Goumoens, Baptiste Blochet, Jian Xu, and Changhuei Yang. High-resolution non-line-of-sight imaging employing active focusing. August 2023.

[CR99] Dr. Subhasis Chaudhuri and Dr. A. N. Rajagopalan. Depth from defocus: A real aperture imaging approach. In *Springer: New York*, 1999.

[FVW20] Daniele Faccio, Andreas Velten, and Gordon Wetzstein. Non-line-of-sight imaging. *Nature Reviews Physics*, 2(6):318–327, May 2020.

[GMO+19] Miguel Galindo, Julio Marco, Matthew O'Toole, Gordon Wetzstein, Diego Gutiérrez, and Adrián Jarabo. A dataset for benchmarking time-resolved non-line-of-sight imaging. pages 1–2, 07 2019.

[KHFG14] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature Photonics*, 8(10):784–790, August 2014.

[LBV20] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems. *Nature Communications*, 11:1645, 04 2020.

[LGLM+19] 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*, pages 1–4, 2019.

[LLGM23] Pablo Luesia Lahoz, Diego Gutierrez, and Adolfo Muñoz. Zone plate virtual lenses for memory-constrained nlos imaging. *Jornada de Jóvenes Investigadores del I3A*, 11, jul. 2023.

[LSW+13] Xing Lin, Jinli Suo, Gordon Wetzstein, Qionghai Dai, and Ramesh Raskar. Coded focal stack photography. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9, 2013.

[LWO19] David B. Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast f–k migration. *ACM Trans. Graph. (SIGGRAPH)*, 38(4):116, 2019.

[Mas04] Barry R. Masters. *Confocal Laser Scanning Microscopy*, pages 895–947. Springer US, New York, NY, 2004.

[MJN+21] Julio Marco, Adrian Jarabo, Ji Hyun Nam, Xiaochun Liu, Miguel Angel Cosculluela, Andreas Velten, and Diego Gutierrez. Virtual light transport matrices for non-line-of-sight imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2440–2449, October 2021.

[MSS+19] Tomohiro Maeda, Guy Satat, Tristan Swedish, Lagnojita Sinha, and Ramesh Raskar. Recent advances in imaging around corners, 2019.

[OLW18] Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Confocal Non-Line-of-Sight Imaging Based on the Light-Cone Transform. *Nature*, 2018.

[Pad08] S.W. Paddock. *Confocal Microscopy: Methods and Protocols*. Methods in Molecular Biology. Humana Press, 2008.

[Ren09] E. Renner. *Pinhole Photography: From Historic Technique to Digital Application*. Alternative Process Photography Series. Focal Press, 2009.

[RMBV19] Syed Azer Reza, Marco La Manna, Sebastian Bauer, and Andreas Velten. Wave-like properties of phasor fields: Experimental demonstrations, 2019.

[RSM+23] Diego Royo, Talha Sultan, Adolfo Muñoz, Khadijeh Masumnia-Bisheh, Eric Brandt, Diego Gutierrez, Andreas Velten, and Julio Marco. Virtual mirrors: Non-line-of-sight imaging beyond the third bounce. *ACM Transactions on Graphics*, 42(4), 2023.

[SA77] Colin Sheppard and A.Choudhury. Image formation in the scanning microscope. *Journal of Modern Optics*, 24:1051–1073, 10 1977.

[SHN+16] Guy Satat, Barmak Heshmat, Nikhil Naik, Albert Redo-Sanchez, and Ramesh Raskar. Advances in ultrafast optics and imaging applications. page 98350Q, 05 2016.

[SPT10] Seyfollah Soleimani, Wilfried Philips, and Linda Tessens. Image fusion using blur estimation. pages 4397 – 4400, 10 2010.

[VWG+12] 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*, 3(1):1–8, 2012.

[WLL23] Kaiyi Wu, Saiyu Luo, and Li Li. Non-line-of-sight reconstruction based on filtered back projection. In Weibiao Chen, Minghui Hong, Jianrong Qiu, Pu Wang, Jianqiang Zhu, and Tiancai Zhang, editors, *Eighteenth National Conference on Laser Technology and Optoelectronics*, volume 12792, page 127920E. International Society for Optics and Photonics, SPIE, 2023.