A Quantitative Platform for Non-Line-of-Sight Imaging Problems

Jonathan Klein[1, 2]
kleinj@cs.uni-bonn.de
Martin Laurenzis[2]
martin.laurenzis@isl.eu
Dominik L. Michels[3]
dominik.michels@kaust.edu.sa
Matthias B. Hullin[1]
hullin@cs.uni-bonn.de

[1] University of Bonn
[2] French-German Research Institute of Saint-Louis (ISL)
[3] King Abdullah University of Science and Technology (KAUST)

Abstract

The computational sensing community has recently seen a surge of works on imaging beyond the direct line of sight. However, most of the reported results rely on drastically different measurement setups and algorithms, and are therefore hard to impossible to compare quantitatively. In this paper, we focus on an important class of approaches, namely those that aim to reconstruct scene properties from time-resolved optical impulse responses. We introduce a collection of reference data and quality metrics that are tailored to the most common use cases, and we define reconstruction challenges that we hope will aid the development and assessment of future methods.

1 Introduction

The challenge of imaging objects outside the direct line of sight is of great potential relevance in many applications and has fascinated scientists, engineers and the general public alike for many years. Recently, the introduction of computational sensing approaches has enabled researchers to "look around corners" and given the topic new momentum [17, 32].

Many published works aim at recovering various scene properties (room geometry, object shape and position, materials) from time-resolved measurements of indirect light reflections. However, the use of different measurement setups with different spatial and temporal resolution as well as the lack of standard targets and ground truth makes it hard to draw meaningful comparisons between the reconstruction algorithms used, to derive recommendations for future sensing designs, and to predict the performance of such designs under real-world conditions. In fact, to this day it remains unknown what the theoretical and practical limits of non-line-of-sight (NLoS) imaging are.

Here, we take a first step to fill this void by proposing a quantitative foundation that is designed to facilitate the development, characterization and comparison of non-line-of-sight reconstruction methods based on time-of-flight data. Our effort is threefold and comprises the following main contributions:

- a database of annotated synthetic time-resolved scene responses that reflects common reconstruction tasks in a hardware-independent manner,
- the development of task-specific error metrics to benchmark reconstruction results, and
- supporting software infrastructure, namely a code repository and an online service that hosts a selection of benchmarks and blind reconstruction challenges.

We hope that this quantitative platform will contribute to the consolidation of existing research efforts, aid the development of future reconstruction techniques, and serve the community as a device for adherence to, and documentation of, good scientific practice.

2 State of the art

We consider works that aim to circumvent the occlusion problem by using electromagnetic waves where the occluder becomes transparent, such as radio waves [1, 2, 8], or that exploit coherence properties of light, reconstructing objects using interferometry or speckle correlation [16] to be outside the scope of this paper. Instead, we focus on those that rely on geometric optics and classic radiative transfer. In the following, we provide an overview of devices and setups, scene layouts and reconstruction algorithms of these works and conclude the section with an attempt to unify the most relevant within our quantitative framework.

2.1 Scene setup: three diffuse bounces

An object that is located within the direct line of sight of a camera or a structured light source can be imaged either by direct observation or by probing it with a projector [29]. The challenge of "looking around corners" refers to settings where the target can neither be directly illuminated nor observed, and where reflections off other objects (reflectors) are the only remaining source of information.[1] The glossiness of these reflectors has a strong influence on the amount of information they transmit (see [15] for a detailed analysis of this trade space). The more mirror-like a surface is, the better it can be used to trivially observe the occluded region; on the other end of the scale are diffuse surfaces that completely destroy the directionality of light upon reflection.

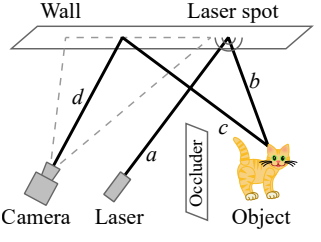


Figure 1: In the most common scenario of NLoS reconstruction, the path traveled by the light (laser-wall-object-wall-camera) consists of four segments *a*–*d* connected by three diffuse reflections.

This leads to a canonical scene arrangement that has been prominently featured in most prior works [3, 5, 6, 7, 9, 10, 11, 15, 17, 18, 21, 25, 30, 31, 32, 35] and that is illustrated in Figure 1. The unknown target is located in front of a planar wall (or floor), and occluded from direct observation. Illuminating a spot on the wall with a collimated light source (laser) turns this spot into a small area light source which illuminates the target. A portion of the light received by the target, in turn, scatters back to the wall, from where it is reflected into a collimated detector or other imaging device. Eventually, the total path of light received

[1] This also excludes settings like the one described by Jin et al. [12] that employ a pinhole to indirectly image the hidden scene.

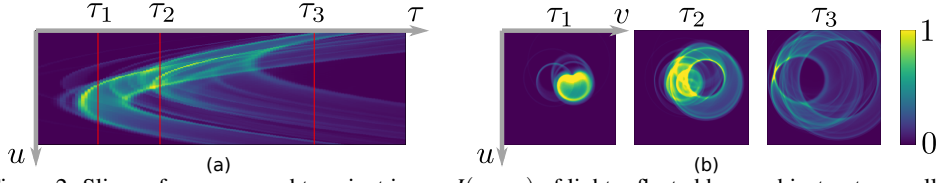


Figure 2: Slices of an unwarped transient image $I(u, v, \tau)$ of light reflected by an object onto a wall as illustrated in Figure 1. u–τ slices (a) resemble streak images, whereas u–v slices (b) can be interpreted as frames of a video of light in flight. (Range normalized for display.)

by the detector consists of four straight segments connected by three bounces. A common way of interpreting this setting is to assume that the geometry of capture setup and wall are known. Similar to treating the laser spot as a virtual area light source, the wall point observed by the detector pixel can be interpreted as a virtual omnidirectional detector [15]. Only considering the 2-segment light path between these virtual devices leads to a transform of the time axis that has been called *unwarping* by Velten et al. [33].

2.2 Space-time impulse response / Devices

The unavailability of a direct line of sight calls for alternative sources of information about the unknown target. Often, ultrafast light sources and time-resolved detectors are used to probe the temporal impulse response of the scene. This typically leads to the notion of a *transient image* $I(u, v, \tau)$, where u and v are the usual image coordinates and τ is the travel time of light (Figure 2). We refer the reader to a recent survey on this general topic [13]. Among the devices used for the purpose are fast photodiodes [17], streak tubes [32], gated image intensifiers [21] and avalanche photodetectors [5, 7, 9, 11, 25]. Although common sense dictates that a high spatial resolution requires a high temporal resolution of the measurement equipment, other researchers have also demonstrated the use of slower emitters and sensors for certain tasks. Examples include amplitude-modulated continuous-wave (AMCW) time-of-flight setups [10, 15, 30] or even entirely unmodulated intensity images [18].

2.3 Reconstruction tasks and algorithms

The non-line-of-sight sensing solutions reported in literature greatly vary in the number of degrees of freedom, ranging from object detection, identification and tracking [7, 15, 17, 18, 30] via characterization of room shapes [25] to the recovery of full 3D shapes [5, 10, 11, 32]. This also reflects in the variety of proposed algorithms, where we identify two main classes of approaches. The first class aims to explain the observed signal in terms of a more or less sophisticated forward model. For instance, researchers have proposed radiative transfer simulations based on oriented surface patches [17, 18, 25], derived a linearized light transport tensor [10, 11], and exploited additional geometric constraints to express light transport as a convolution of light cones [24]. The problem of reconstructing the scene s thus typically takes on the form of a regularized least-squares minimization of the difference between the measured and simulated images.

The second class of reconstruction algorithms are based on the backprojection principle, where intensity values in the space-time response "vote" for feasible object locations within a reconstruction volume. For each given sample, the manifold of such locations forms an ellipsoid in space [33].

We are not aware of any systematic investigation as to which of these approaches is best suited for a given reconstruction problem. For the backprojection technique, La Manna et al. compared different flavors (additive/multiplicative backprojection) as well as different iteration and filtering strategies [20].

3 Challenge design

On the highest level, the non-line-of-sight reconstruction problem addressed in our challenge is: given the transient image $I(u, v, \tau)$, what is the scene s? Here, s can stand for any scene properties that are of interest, like object or room shapes, object classes, object position and orientation, material reflectance, texture, and so on. In this section, we aim to unify the previously discussed work into our proposed evaluation benchmark.

3.1 Basic scene geometry

The huge variety of setups makes it hard to directly compare existing approaches and therefore calls for a unification. We propose a new, minimalistic setup that only consists of the key elements that are common in the previous setups, as shown in Figure 3. Our scene only consists of a light source, an object reflecting the light and the wall receiving the reflections. Currently all scenes contain only a single object, some examples of which can be seen in Figure 5a.

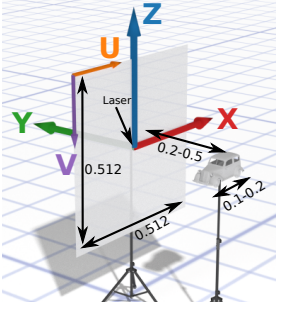


Figure 3: Our unified scene geometry.

Notably, this setup does not include an actual occluder or, in general, a scene surrounding the wall and the object. In previous publications it has been usually assumed that these elements do not interact with the light transport via occlusions or reflections and thus their existence is usually neglected.

The setup consists of a single laser spot, which is centered on the wall and forms the origin of the coordinate system. An array of observation points sample the backscatter arriving at the wall. Due to the reciprocity of the light transport, our data can also be used for methods assuming a single observation and multiple illumination points. We do not include the most general (five-dimensional) case with multiple observation and multiple illumination points, as capturing and storing such data would be intractable in practice. Attempts which require this more general transient images (such as [24]) only use a certain subset of them, but there is currently no agreement on a specific subset. When a new standard emerges, our database will be updated accordingly.

Scene objects are placed at different positions inside a volume in front of the wall such that their projection on the X/Z plane lies always completely inside the wall. This constellation can be considered a sweet spot for the reconstruction, although in practice, placing the laser spot within the detector's field of view would make the setup more prone to lens flare.

Almost all previous work assumes perfectly diffuse materials, which is why most of the objects in our benchmark are perfectly diffuse as well. To probe the limits of the diffusity assumption, some objects use a shiny metal material, based on the GGX model [34]. Since material reconstruction is not part of the benchmark in this first iteration, all material parameters are provided.

3.2 Data units and formats

Our image formation models does not contain any nonlinearities and thus the actual scale of the setup is irrelevant. The dimensions shown in Figure 3 are derived from an extensive analysis of the proportions of setups found in the literature (see supplementary). Expressing the temporal dimension in terms of the optical path length allows us to use the same arbitrary units for spatial dimensions and time of flight.

With the exception of [18] all considered hardware platforms use time-resolved data for the reconstruction, but the data format depends on the used hardware. We provide raw transient data in a generic format that can be converted to any kind of hardware format (including the intensity images used in [18] by integrating over the time dimension). An example of a transient image can be seen in Figure 2.

All our data are time unwarped (path segments a and d in Figure 1 are removed), but we provide a conversion tool in which a camera and laser position can be specified. The tool inverts the unwrapping, including a cropping and perspective transformation of the reflector wall. Additional corruption of the data by various noise sources is also possible (see the supplementary for detail). Contestants are encourage to use these tools to produce realistic raw data for their setups (including sensor response, additional lens distortion, conversion to camera data format, and other effects). The results on these data reflect how they behave in realistic setups but are not part of the competition as the different setups make them incomparable.

3.3 Transient image generation

We motivate the usage of synthetic renderings instead of real measurements in two ways: firstly, each hardware has its own limitations and no setup can capture the actual transient light transport directly. Secondly, ground truth data is required for the evaluation. Building and measuring a real scene will inevitably introduce certain errors and would also prevent the usage of the minimalistic setup shown in Figure 3. Thus synthetic renderings provide both high-quality transient images and high-quality ground truth data.

As rendering tool, we extended `pbrt-v3` [26], a state-of-the-art, multi-purpose global illumination renderer with special focus on physical accuracy, by tracing the path lengths and writing three-dimensional transient output. The correctness of the obtained images was verified as follows: We assume that the intensity images computed by the unmodified `pbrt-v3` are physically accurate, as one of its explicit design goals is physical accuracy, it has been around for many years and its open source code has been studied by hundreds of scientists worldwide. We integrated transient images over the temporal domain and successfully compared it to the intensity rendering of the same scene, meaning that each transient pixel has the correct total amount of light. We checked the correct temporal distribution by rendering test scenes with sharp temporal responses whose time offsets are easily measurable. Lastly, the importance sampling was successfully verified by rendering the same scene with enabled and disabled importance sampling to almost full convergence and comparing the results.

Rendering noise-free images with global illumination is computationally expensive and the additional third dimension of transient images reinforces this problem drastically. The images are rendered with a spatial resolution of 256×256 pixels and a temporal resolution of 1600.

For a detailed description of the data formats and renderer implementation, we refer to the supplementary material.

3.4 Submission

The datasets for our reconstruction benchmark are available through a web frontend at https://nlos.cs.uni-bonn.de/. The functionality of the system is inspired by existing offerings, in particular the known two-view and multi-view stereo reconstruction challenges. Besides the transient datasets, users can download an SDK with functions for data handling, error metrics and a base-line reconstruction algorithm, namely ellipsoidal backprojection.

Users can create an anonymous account to upload their reconstructions and have them scored against the ground truth. The scores are time-stamped and can be submitted to the leaderboard (in anonymized or de-anonymized form), where they are compared to the scores of other contestants.

4 Scenes

We present a set of challenges, each tailored to a specific problem in non-line-of-sight imaging, and introduce appropriate metrics for their evaluation. A complete list of all scenes is found in the supplementary.

Apart from the four challenges presented here, our platform is open for future extensions ranging from material reconstruction [24], non-rigid pose estimation (like tracking of articulated human motion) to complex scenes with many detailed objects.

4.1 Materials

Our dataset contains models with two different materials. Non-line-of-sight imaging literature commonly assumes that the hidden scene is perfectly diffuse. We thus use a diffuse material (with an albedo of 0.8) for most objects. To reflect real-world situations, we "pollute" our database with roughly 25% objects that are made of a non-diffuse material, namely `pbrt`'s default *metal* material which implements Walter et al.'s GGX model [34]. The material parameters `k`= 3.63 and `eta`= 0.216 represent copper at a wavelength of 650nm. They are kept constant throughout the whole benchmark. While this additional variation is not sufficient to include the reconstruction of material parameters in the challenge, it probes how well different reconstruction algorithms handle different materials, or how much they are influenced by the invalid assumption of a diffuse world.

4.2 Geometry reconstruction

The goal of this challenge is to reconstruct the object's geometry from a single transient image as illustrated in Figure 4a. For this, sixteen different object types with varying complexity are provided.

In order to evaluate the results, ground truth mesh and reconstructed mesh have to be compared. There exist a wide variety of classical metrics for mesh comparison employing measurements of surface distance and curvature [27, 28] or volume [22]. A global comparison between two meshes can be achieved using an error metric based on the Hausdorff distance [19]. However, there is no uniquely best metric and an appropriate choice depends on the specific scenario.

Since we are dealing with opaque objects and thus the reflected light does not carry any information about its inside, the application of a surface metric is a natural choice. More

precisely, we chose to compare triangle meshes, as they are a widely used and easily processable surface representation. We define our metric as follows. Let $\mathcal{M} \subset \mathbb{R}^3$ be a mesh described by its triangulation $\mathcal{T} \subset \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$. Consider a triangle $t := (v_0^t, v_1^t, v_2^t) \in \mathcal{T}$ with its corner vertices v_i^t. Its center is given by $c^t = (v_0^t + v_1^t + v_2^t)/3$ and its area by $A^t = \|(v_1^t - v_0^t) \times (v_2^t - v_0^t)\|/2$ in which $\|\cdot\|$ denotes the Euclidean norm. The asymmetric distance from a triangle mesh $\mathcal{M}_0$ to another triangle mesh $\mathcal{M}_1$ is then given by

$$d(\mathcal{M}_0, \mathcal{M}_1) = \sum_{t_0 \in \mathcal{M}_0} \frac{A^{t_0}}{A^{\mathcal{M}_0}} \min_{t_1 \in \mathcal{M}_1} \|c^{t_0} - c^{t_1}\| \tag{1}$$

with $A^{\mathcal{M}_0} = \sum_{t_0 \in \mathcal{M}_0} A^{t_0}$, and the symmetric distance by

$$D(\mathcal{M}_0, \mathcal{M}_1) = \max\left(d(\mathcal{M}_0, \mathcal{M}_1), d(\mathcal{M}_1, \mathcal{M}_0)\right). \tag{2}$$

Essentially, the average distance per surface area is computed. With $\mathcal{G}$ as the ground truth mesh and $\mathcal{R}$ as the reconstructed mesh, we store both distances $d(\mathcal{R}, \mathcal{G})$ and $d(\mathcal{G}, \mathcal{R})$ as they represent different quality indicators. For example, $d(\mathcal{R}, \mathcal{G}) = 0$ is reached if only a single point is reconstructed correctly while $d(\mathcal{G}, \mathcal{R}) = 0$ is reached when the reconstruction contains the whole volume. Other properties of this metric are:

- Neutrality of treatment due to area-weighting: every part of the mesh is of the same importance, standard operations like subdivision are handled appropriately.
- Robustness to incompleteness and overcompleteness: if $\mathcal{M}_1 \subset \mathcal{M}_0$, the superfluous parts of $\mathcal{M}_0$ would not have a good match and thus increase $d(\mathcal{M}_0, \mathcal{M}_1)$. Likewise $\mathcal{M}_0 \subset \mathcal{M}_1$ is handled by $D(\mathcal{M}_0, \mathcal{M}_1)$. Superfluous geometry far away from the mesh receives a stronger penalty.

The reflected signals contains mostly information about the front side of the object. Therefore also only the front sides of objects are considered in the evaluation by filtering out triangles that face away from the wall.

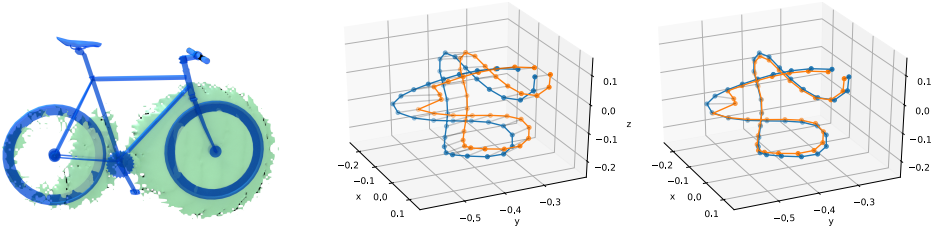
Some proposed algorithms reconstruct occupancy or probability volumes. Such volumetric representations can be converted into triangle meshes using an implementation of Marching Cubes [23] that is provided as part of the SDK. Contestants concerned about the triangulation quality are encouraged to use a different implementation.

4.3 Position and orientation tracking

In object tracking the goal is to reconstruct the position and orientation of an object for each frame resulting in a full trajectory reconstruction, see Figure 4c. For that, different objects with known and unknown geometries are provided.

For each object there are four different animation tracks: i) object moves along the three main axes, ii) object rotates around the three main axes, iii) object moves along a complex path, and iv) object moves along a complex path and adopts its orientation. For each object, individual paths are used.

Animation tracks are limited to 40 frames to keep the database size manageable, where each frame consists of a position (the objects center of mass) and an orientation. Two paths P



(a) (b) (c)

Figure 4: (a): Exemplary geometry reconstruction. The basic shape of the bike object (blue) is recognizable in the reconstruction (green), however details like saddle, pedals and handlebar are missing. (b–c): Trajectory reconstruction. (b): The ground truth trajectory is shown in blue, the reconstructed in orange. (c): After subtracting a constant offset, the trajectories are close together, except for two outliers.

and P' with $P = (p_0, p_1, \ldots, p_{n-1})^\mathsf{T}$ with $p_i = (p_i^x, p_i^y, p_i^z) \in \mathbb{R}^3$ are compared by computing the root-mean-square (RMS) error:

$$\|P - P'\|_{\text{pos}} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1} \left\|p_i - p_i'\right\|^2}. \tag{3}$$

Since this metric penalizes outliers, contestants are encouraged to apply appropriate outlier detection and removal, e.g., by comparing the result to neighboring time frames. Furthermore, the computed centers of mass of the objects might be biased, so a least-squares optimal constant offset x between P and P' is computed.

For the evaluation, the minimal path distance $S = \|(P - x) - P'\|_{\text{pos}}$, the length of the offset $\|x\|$ and the completeness (the number of reconstructed frames divided by the total number of frames) are evaluated.

Orientations are treated in a similar fashion: given orientations q and q' in quaternion representation, the difference is computed by the unit quaternion dot product metric

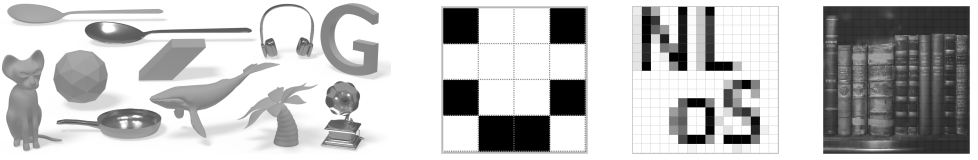
$$\|q - q'\|_{\text{quat}} = 1 - \left|\langle q, q' \rangle\right| \in [0, 1],$$

where $|\langle \cdot, \cdot \rangle|$ denotes the absolute value of the dot product between the four components of the quaternions [12]. Defining the original orientation of the object is not as straightforward as defining its origin as its center of mass. Therefore the initial orientation for the first frame of each animation is given, and thus only $n - 1$ frames are evaluated. With this metric, the orientation reconstruction accuracy is evaluated analogously to the path distance, including the completeness score.

4.4 Classification

The goal of the classification challenge is to accurately determine the type of an object. For that we provide a classification data set which consists of eleven known models, see Figure 5a. Each model is rendered at various positions and orientations inside the usual volume. The goal is to decide for each scene, which of the objects is shown. This challenge is expected to be the easiest as the possible output has a very limited range.

Classification results are evaluated in a confusion matrix using the harmonic average of precision and recall (F_1 score). In general, fuzzy classification is used; algorithms that do a



(a) (b) 4×4 (c) 16×16 (d) 128×128

Figure 5: a) Classification data set. Overview of the eleven different models used for the classification challenge. b-d) Examples textures from the texture reconstruction challenge. The textures have different resolutions and different color depths.

hard classification consequently restrict their weights to 0 and 1. If no solution is provided for a certain frame, identical weights for all classes are assumed.

4.5 Texture reconstruction

For the texture reconstruction challenge, a known, planar geometry is set up in parallel to the reflector at a specified position. It has varying textures which have to be reconstructed. They are split in three classes with increasing resolution and color depth (4×4 pixels in black and white, 16×16 pixels in 5 gray values, and 128×128 pixels in 256 gray values). Examples of the different classes can be seen in Figure 5.

Non-line-of-sight texture reconstruction has some unique characteristics that need to be taken into account, when a comparison metric is defined. Although no publication so far directly tackled the problem of texture reconstruction, a few have reconstructed flat letters, a problem that is similar in nature. Based on these results we expect reconstructed texture to cover the low-frequency content better than the high-frequency details. Thus we propose a multi-scale approach which compares different frequency bands independently.

Given an $n\times n$-pixel texture $T\in[0,1]^{n\times n}$ in which n is a power of 2, we compute a Laplacian pyramid by iteratively blurring and downsampling T, and storing the differences between the steps in the individual pyramid layers [4]. This essentially decomposes the image into its different frequency components.

Let $T_n, T_{n/2}, \ldots, T_1$ respectively $T'_n, T'_{n/2}, \ldots, T'_1$ denote the individual pyramid layers. The differences between each layer are computed by applying the Frobenius norm $\|\cdot\|_\mathrm{F}$ onto the texture difference $T_n - T'_n$ and normalizing the result by n^2. This allows for quality measurement on different scales, e.g., by taking the square root of the average squared per-pixel differences

$$\sqrt{\frac{\|T_n - T'_n\|_\mathrm{F}^2}{n^2}}, \sqrt{\frac{\|T_{n/2} - T'_{n/2}\|_\mathrm{F}^2}{(n/2)^2}}, \ldots, |T_1 - T'_1|. \tag{4}$$

Next to these quality indicators on each scale, its (uniformly weighted) squared average value

$$\|T - T'\|_\mathrm{img} = \sqrt{\frac{\sum_{i=0}^{\log_2(n)} \|T_{n/2^i} - T'_{n/2^i}\|_\mathrm{F}^2 / (n/2^i)^2}{\log_2(n)+1}} \tag{5}$$

is used as a quality metric.

5 Reconstruction results

With the exception of Arellano et al. [3], reconstruction code is not available, making the comparison challenging. We therefore seed our reconstruction challenge with the de-facto standard reconstruction method, ellipsoidal backprojection, using the implementation of [3]. Its generality and ubiquity (as well as the lack of general-purpose alternatives) makes backprojection a natural baseline for current and future work. Although the method itself can only be used for geometry reconstruction, we implemented a straightforward extension for position tracking (where the object position is defined as center of mass of the reconstructed volume). We imagine that adaptations to the other challenges can be developed as well, but consider this to be beyond the scope of this benchmark.

Results of the geometry reconstruction and object tracking are shown in Figure 4. Exact numbers for each scene are found in the supplementary material.

6 Discussion and outlook

In this paper we introduced methodology and a data foundation for a first reconstruction benchmark for non-line-of-sight imaging. The research in the field so far resembles a collection of isolated data points, most of them with promising and inspiring results but without strong links to other pieces of work. Of course, in light of the diversity of tasks, scales and devices, all a database like ours can ever hope to provide must be a compromise. Nevertheless, we hope that this work can act as a seed for a continuing effort to draw quantitative connections between past and future efforts that will further unify the field.

As the research advances, we plan to constantly update the database with new reconstruction problems and realistic data (e.g. light scattered from the scene background that needs to be filtered out by contestants). We also hope that more researchers will be willing to share their reconstruction code in order to build an open source repository of reconstruction algorithms.

The database, the submission system, and all other material is available on our website at https://nlos.cs.uni-bonn.de/.

References

[1] Fadel Adib and Dina Katabi. See through walls with wifi! *SIGCOMM Comput. Commun. Rev.*, 43(4):75–86, August 2013. ISSN 0146-4833. doi: 10.1145/2534169.2486039. URL http://doi.acm.org/10.1145/2534169.2486039.

[2] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 34(6): 219:1–219:13, October 2015. ISSN 0730-0301. doi: 10.1145/2816795.2818072. URL http://doi.acm.org/10.1145/2816795.2818072.

[3] Victor Arellano, Diego Gutierrez, and Adrian Jarabo. Fast back-projection for non-line of sight reconstruction. *Opt. Express*, 25(10):11574–11583, May 2017. doi: 10.1364/OE.25.011574. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-25-10-11574.

[4] Peter J. Burt and Edward H. Adelson. Readings in computer vision: Issues, problems, principles, and paradigms. chapter The Laplacian Pyramid As a Compact Image Code, pages 671–679. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987. ISBN 0-934613-33-8. URL http://dl.acm.org/citation.cfm?id=33517.33571.

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

[6] Piergiorgio Caramazza, Alessandro Boccolini, Daniel Buschek, Matthias Hullin, Catherine Higham, Robert Henderson, Roderick Murray-Smith, and Daniele Faccio. Neural network identification of people hidden from view with a single-pixel, single-photon detector. *arXiv preprint arXiv:1709.07244*, 2017.

[7] S. Chan, R.E. Warburton, G. Gariepy, Y. Altmann, S. McLaughlin, J. Leach, and D. Faccio. Fast tracking of hidden objects with single-pixel detectors. *Electronics Letters*, 53(15):1005–1008, 7 2017. ISSN 0013-5194. doi: 10.1049/el.2017.0993.

[8] David D. Ferris, Jr. and Nicholas C. Currie. Survey of current technologies for through-the-wall surveillance (tws). *Proc. SPIE*, 3577:62–72, 1999. doi: 10.1117/12.336988. URL http://dx.doi.org/10.1117/12.336988.

[9] Genevieve Gariepy, Francesco Tonolini, Robert Henderson, Jonathan Leach, and Daniele Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10(1), 2016.

[10] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B. Hullin. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[11] Felix Heide, Matthew O'Toole, Kai Zhang, David B. Lindell, Steven Diamond, and Gordon Wetzstein. Robust non-line-of-sight imaging with single photon detectors. *CoRR*, abs/1711.07134, 2017. URL http://arxiv.org/abs/1711.07134.

[12] Du Q. Huynh. Metrics for 3d rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35(2):155–164, 2009. URL http://dx.doi.org/10.1007/s10851-009-0161-2.

[13] Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics*, 1(1):65 – 79, 2017. ISSN 2468-502X. doi: https://doi.org/10.1016/j.visinf.2017.01.008. URL http://www.sciencedirect.com/science/article/pii/S2468502X17300104.

[14] Chenfei Jin, Zitong Song, Siqi Zhang, Jianhua Zhai, and Yuan Zhao. Look through a small hole using three laser scatterings. *Optics Letters*, November 2014.

[15] Achuta Kadambi, Hang Zhao, Boxin Shi, and Ramesh Raskar. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics*, 35(2):15:1–15:12, March 2016. ISSN 0730-0301. doi: 10.1145/2836164. URL http://doi.acm.org/10.1145/2836164.

[16] 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, 2014.

[17] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar. Looking around the corner using transient imaging. In *Proc. ICCV*, pages 159–166, 2009.

[18] Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B. Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific Reports*, 6(32491), August 2016. doi: doi:10.1038/srep32491. URL http://www.nature.com/articles/srep32491.

[19] Reinhard Klein, Gunther Liebich, and Wolfgang Straßer. Mesh reduction with error control. In *Proceedings of the 7th Conference on Visualization '96*, VIS '96, pages 311–318, Los Alamitos, CA, USA, 1996. IEEE Computer Society Press. ISBN 0-89791-864-9. URL http://dl.acm.org/citation.cfm?id=244979.245624.

[20] Marco La Manna, Fiona Kine, Eric Breitbach, Jonathan Jackson, and Andreas Velten. Error backprojection algorithms for non-line-of-sight imaging. Technical Report TR1850, University of Wisconsin-Madison, 2017. http://digital.library.wisc.edu/1793/76968.

[21] Martin Laurenzis and Andreas Velten. Nonline-of-sight laser gated viewing of scattered photons. *Optical Engineering*, 53(2), 2014. doi: 10.1117/1.OE.53.2.023102.

[22] Peter Lindstrom and Greg Turk. Fast and memory efficient polygonal simplification. In *Proceedings of the Conference on Visualization '98*, VIS '98, pages 279–286, Los Alamitos, CA, USA, 1998. IEEE Computer Society Press. ISBN 1-58113-106-2. URL http://dl.acm.org/citation.cfm?id=288216.288288.

[23] William Lorensen and Harvey Cline. Marching Cubes: A high resolution 3D surface construction algorithm. In *Proc. 14th Annual Conf. on Computer Graphics and Interactive Techniques*, SIGGRAPH '87, pages 163–169. ACM, 1987. ISBN 0-89791-227-6. doi: 10.1145/37401.37422. URL http://doi.acm.org/10.1145/37401.37422.

[24] Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 2018. doi: http://dx.doi.org/10.1038/nature25489.

[25] Adithya Kumar Pediredla, Mauro Buttafava, Alberto Tosi, Oliver Cossairt, and Ashok Veeraraghavan. Reconstructing rooms using photon echoes: A plane based model and reconstruction algorithm for looking around the corner. In *Computational Photography (ICCP), 2017 IEEE International Conference on*, pages 1–12. IEEE, 2017.

[26] M. Pharr, W. Jakob, and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. Elsevier Science, 3rd edition, 2016. ISBN 9780128007099. URL https://books.google.de/books?id=iNMVBQAAQBAJ.

[27] Rémi Ronfard and Jarek Rossignac. Full-range approximation of triangulated polyhedra. *Comput. Graph. Forum*, 15:67–76, 1996.

[28] William J. Schroeder, Jonathan A. Zarge, and William E. Lorensen. Decimation of triangle meshes. *SIGGRAPH Comput. Graph.*, 26(2):65–70, July 1992. ISSN 0097-8930. doi: 10.1145/142920.134010. URL http://doi.acm.org/10.1145/142920.134010.

[29] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen R Marschner, Mark Horowitz, Marc Levoy, and Hendrik Lensch. Dual photography. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 745–755. ACM, 2005.

[30] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. Computational imaging with multi-camera time-of-flight systems. *ACM Trans. Graph. (SIGGRAPH)*, 2016.

[31] Chia-Yin Tsai, Kiriakos N. Kutulakos, Srinivasa G. Narasimhan, and Aswin C. Sankaranarayanan. The geometry of first-returning photons for non-line-of-sight imaging. *IEEE Intl. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017.

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

[33] 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 Trans. Graph.*, 32(4):44:1–44:8, July 2013. ISSN 0730-0301. doi: 10.1145/2461912.2461928. URL http://doi.acm.org/10.1145/2461912.2461928.

[34] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. Microfacet models for refraction through rough surfaces. *Eurographics Symposium on Rendering*, 2007.

[35] Ryan E. Warburton, Susan Chan, Genevieve Gariepy, Yoann Altmann, Steve McLaughlin, Jonathan Leach, and Daniele Faccio. Real-time tracking of hidden objects with single-pixel detectors. In *Imaging and Applied Optics 2016*, page IT4E.2. Optical Society of America, 2016. doi: 10.1364/ISA.2016.IT4E.2. URL http://www.osapublishing.org/abstract.cfm?URI=ISA-2016-IT4E.2.