Optics EXPRESS

A calibration scheme for non-line-of-sight imaging setups

JONATHAN KLEIN,[1,*] MARTIN LAURENZIS,[2] MATTHIAS B. HULLIN,[1] AND JULIAN ISERINGHAUSEN[1,3]

[1] *Institute of Computer Science, University of Bonn, Endenicher Allee 19a, 53115 Bonn, Germany*
[2] *French-German Research Institute of Saint-Louis, 5 rue du Général Cassagnou, 68300 Saint-Louis, France*
[3] *Now at Google, USA*
[*] *kleinj@cs.uni-bonn.de*

Abstract: The recent years have given rise to a large number of techniques for "looking around corners", i.e., for reconstructing or tracking occluded objects from indirect light reflections off a wall. While the direct view of cameras is routinely calibrated in computer vision applications, the calibration of non-line-of-sight setups has so far relied on manual measurement of the most important dimensions (device positions, wall position and orientation, etc.). In this paper, we propose a method for calibrating time-of-flight-based non-line-of-sight imaging systems that relies on mirrors as known targets. A roughly determined initialization is refined in order to optimize for spatio-temporal consistency. Our system is general enough to be applicable to a variety of sensing scenarios ranging from single sources/detectors via scanning arrangements to large-scale arrays. It is robust towards bad initialization and the achieved accuracy is proportional to the depth resolution of the camera system.

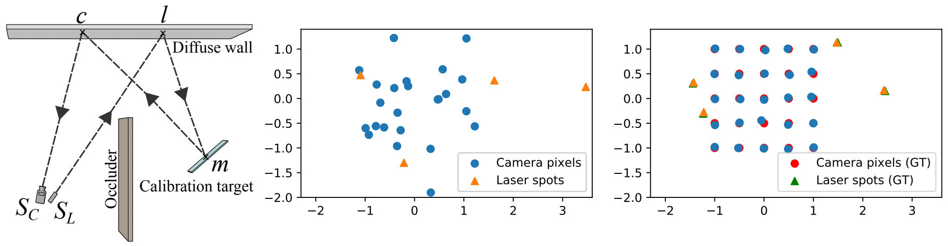
1. Introduction

The ability to "see" beyond the direct line of sight forms not only an intriguing academic problem but also has possible future applications ranging from emergency situations, where situational awareness about dangers and victims is key, to scientific scenarios, where microscopes supporting such techniques reveal hidden structures.

The recent years have produced a number of techniques that sense objects located "around a corner" by recording time-resolved optical impulse responses, where light that bounces off a directly visible wall enters the occluded part of the scene and thus gathers information about hidden objects; see Fig. 1(a) for a schematic illustration. The available operation modes [1–6] support not only object detection and tracking of components of the occluded scene but extend to the full reconstruction of 3D shape and texture. In general it is assumed that the entire geometry of the setup is known and only the hidden object is to be reconstructed. This implies that the capture must be preceded by a manual calibration: Positions and distances of devices and objects have to be measured with high accuracy, a task which is tedious and often results in imprecise results.

Here, we propose an automatic system for calibrating the geometry of non-line-of-sight sensing setups. Our scheme does not require any additional hardware other than a common, planar mirror which serves as the calibration target. As in traditional camera calibration, the target is recorded in different positions and orientations. Since the calibration scheme does not rely on the target being textured, and since only a temporal onset (rather than the full time-of-flight histogram) is used, our calibration scheme can be employed for all types of ultrafast sensors, including single-pixel sensing scenarios [7], randomly scattered measurement locations [8] as well as low-resolution imagers and even correlation time-of-flight sensors [9]. Additionally, task-specific constraints (e.g., pixel positions restricted to a scan line) are easily integrated in the method.

#398647 https://doi.org/10.1364/OE.398647
Journal © 2020 Received 9 Jun 2020; revised 16 Jul 2020; accepted 27 Jul 2020; published 9 Sep 2020
Corrected: 16 September 2020



a) Scene setup b) Initialization c) Calibration result

Fig. 1. (a) We propose a novel method for the geometric calibration of three-bounce non-line-of-sight setups using transient imaging hardware. Light travels from a laser S_L to a laser spot l located on the diffuse reflector wall. From there, it is reflected towards a calibration target m and back to a projected camera pixel c, finally reaching the camera S_C. We calibrate the setup using multiple images of a specular, planar mirror in different positions and orientations, analog to the procedure in classical 2D camera calibration. Instead of relying on known features on the calibration target, we use the time of flight of the full path from laser to camera to solve for the individual laser spot positions l and projected camera pixels c. (b) The optimization problem is non-convex but has very low initialization requirements (e.g. eyeballing). (c) Even in the presence of time-of-flight noise, our method reconstructs the setup geometry up to a very high precision. The ground truth values (shown in red and green) are barely visible under the reconstruction.

Our calibration scheme requires an initialization to warm-start the non-linear optimization problem. In contrast to a laborious measurement however, we rely only on a rough estimate of the setup's geometry: As long as the initial solution coarsely reflects the dimensions of the scene geometry, the method is robust even in the presence of time-of-flight noise.

Using an experimental measurement setup, we demonstrate that our scheme not only recovers relevant parameters to high accuracy, but that it also improves the outcome of non-line-of-sight (NLoS) reconstructions obtained using data from the setup.

2. Related work

The last decade gave rise to a comprehensive body of work on non-line-of-sight sensing, i.e., the estimation of targets hidden from direct view by means of light undergoing indirect diffuse scattering off directly visible proxy objects. While various lines of research are exploring the use of steady-state measurements in order to extend the direct line of sight [10–15], the majority of works remains focused on the use of time-resolved measurements (transient images).

A survey by Jarabo et al. provides a good overview of transient imaging [16]. Seminal works include the recovery of low-parameter geometry and reflectance models from transient measurements [2,17] as well as the first reconstruction of distinct shapes [1]. Since then, significant effort has been devoted to unlock novel sensor technologies and interferometric setups for transient imaging [5,18,19] while simultaneously improving the performance of the de-facto standard reconstruction technique, ellipsoidal error backprojection [1,20,21]. Recent additions to the non-line-of-sight reconstruction problem include the introduction of the confocal capture setting [4] as well as attempts to cast the problem into paradigms borrowed from wave optics and seismic tomography [22,23]. While most of these works rely on volumetric representations for the hidden target, other researchers have explored alternative, surface-driven representations as well [6,24,25]. These models typically lead to improved consistency of the solution with respect to a physically-based forward simulation of light transport, and they also naturally express effects like surface reflectance (BRDFs) or self-occlusion. Equipping volumetric representations with such surface-based characteristics to "guide" the reconstruction is possible, but comes at greatly increased implementation effort and computational cost [3,9]. Lately, there has also been some work introducing machine learning algorithms to NLoS reconstructions [15,26–28].

While details on setup calibration are often omitted in publications and the setup geometry is just assumed to be known, the reported calibration methods can be grouped in several categories. Instead of completely manual measurements (e.g. [5,11]), extending the setup by dedicated calibration hardware is a common approach. Buttafava et al. uses a web cam to estimate the 3D position of the visible laser spot, however the webcam itself is manually calibrated using a dot pattern [8]. La Manna et al. demonstrate NLoS reconstruction using a moving curtain as relay surface which is scanned by an additional SPAD camera to achieve real-time calibration [29]. Co-axial setups (where the position of the laser spot always coincides with the current camera pixel position) usually use precise galvometers, which provide accurate angle information. Together with the ability to measure the time-of-flight of the first reflection, the position of the currently observed point can directly be computed [4,23]. Speckle correlation based approaches (e.g. [10,27]) reconstruct the scene from non-transient measurements of a speckle pattern on the reflector wall and thus do not rely on a geometric calibration in the same way as transient approaches do. Machine-learning based methods that are trained on a static setup implicitly learn the setup geometry and are inherently calibration-free [26,28]. However, a such trained network cannot be transferred to new setups.

3. Method

A non-line-of-sight setup can be viewed as a high-dimensional function that maps parameters such as the setup geometry, the hidden object, reflective properties of various components, a background signal, the sensor model of the camera, and others to measurements. We distinguish radiometric parameters (that govern the amount of light being transported) and spatio-temporal parameters (that govern the time of flight). A first abstraction step drops camera and laser peculiarities and describes measurements as transient histograms, i.e., the time-resolved (on a pico- to nanosecond scale) intensity of light arriving at each sensor pixel. Commonly all participating reflectance functions (BRDF) are assumed to be Lambertian (with notable exceptions such as NLoS BRDF reconstruction [17] or retro-reflective objects [4]), and scenes are set up to minimize reflections from the background. With these assumptions only the scene geometry and the hidden object remain unknown.

With scene geometry measurements available, an analysis-by-synthesis approach can be employed to reconstruct the hidden geometry [6,11]. A setup calibration can be attempted in a similar fashion: Given a known hidden object (i.e. information like position, shape, size, and reflective properties that are required to compute light transport are known) the setup is inferred from measured transient data. The hidden object can be chosen freely (e.g. for diffuse objects a image formation model as presented in [11] could be used) but we propose to use simple planar mirrors, as available as common household object. As we will show in the following, this choice significantly simplifies the image formation model. This then leads to an easier-to-solve optimization problem (compared to general diffuse objects) that has far weaker requirements on its initialization due to its implicit constraints. Our approach jointly optimizes for setup geometry and mirror placement, which allows for a setup calibration with little manual measurements (that can be performed with reduced accuracy to acquire only a rough estimate) for initialization.

The mirrors can be placed in the visible and hidden part of the scene. Thus access to the hidden part is not strictly required, however it can lead to more robust calibration, if it is accessible.

3.1. Image formation model

Figure 1(a) gives a schematic illustration of an NLoS calibration setup: a sensor / laser light source setup on the left hand side which is separated from the mirror calibration target by an occluder. We denote the physical position of the camera and the laser with S_C and S_L respectively. As they are usually close to each other we define the shorthand notation $S = \{S_C, S_L\}$. In the classic three-bounce setup the signal is reflected from a planar wall. We denote the projected camera pixels on this wall with $c \in C$ and the (potentially multiple) laser spots with $l \in L$. The mirrors that replace the hidden object in our setup are denoted with $m \in M$.

Whether the pixels lie on a fixed grid (as for 2D image sensors), a single line (as for streak cameras) or are placed arbitrarily on the wall (as for scans with single-pixel detectors) matters only insomuch as that some cases allow for specialized parameterizations that can improve calibrations (see Section 3.3). Due to Helmholtz reciprocity the roles of L and C are always interchangeable in the following discussion. Most common NLoS setups assume that all $l \in L$ and $c \in C$ lie on the same plane, which is the case for a planar wall. However, our method is also applicable for general 3D points, which allows us to cover a wide variety of NLoS setups such as curved walls, or walls that are rough (in the scale of the hardware's temporal resolution).

Hidden objects have usually a complex shape and thus interreflections have to be taken into account. In contrast, the specular reflections on the mirrors we use as calibration targets allow for only a single, unique optical path $l \longrightarrow m \longrightarrow c$, connecting laser spot, mirror and projected pixel. Compared to classical transient rendering this means that no integration over the surface of the object is required, which allows for fast and noise-free computation. Our transient histograms only contain a single, sharp peak. We assume that those peaks can be retrieved in a hardware-specific pre-calibration step that handles effects such as background radiation or higher-order bounces (see Appendix A.1).

A complete measurement consists of a series of paths $P_{i,j,k} = S_L \longrightarrow l_i \longrightarrow m_j \longrightarrow c_k \longrightarrow S_C$ (we omit indices in unambiguous cases). We assume that those paths are measured individually (i.e. using only one mirror and illuminating only one laser spot at a time).

Each path is characterized by a time of flight and an intensity. The intensity depends on the BRDF of the diffuse wall and its normal vector, while the time of flight is independent of both. For our calibration we only rely on the time of flight. We thus neither need to assume nor to estimate any BRDFs or wall normals (however, the wall's surface normal can be estimated using the reconstructed 3D positions of laser spots and camera pixels).

For the time of flight computation we need to compute the length of a path $S_L \longrightarrow l \longrightarrow m \longrightarrow c \longrightarrow S_C$. Note that m is a plane while S_L, S_C, l, and c are points. Due to the specularity constraint of the mirror reflection there exists a unique point m^r on m at which the light is reflected. The length of the sub path $l \longrightarrow m^r \longrightarrow c$ is equal to the path length $l' \longrightarrow c$, where l' is the point l mirrored at m (see Fig. 2).

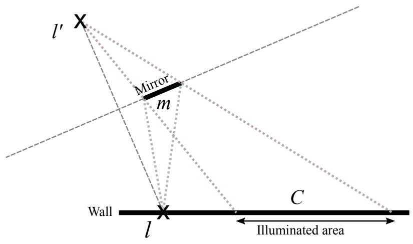


Fig. 2. To assess the optical path $l \longrightarrow m^r \longrightarrow c$, we use a similarity relation: The laser spot l illuminates the wall as if it was reflected on the mirror plane, resulting in a virtual light spot l'.

A mirror plane is represented in the Hesse normal form as normal vector n and scalar offset d. Then

$$l' = l - 2(n \cdot l + d)n \tag{1}$$

and the total path length is the sum of all path segments,

$$f(S, l, c, m) = \|l - S_L\| + \|c - l'\| + \|S_C - c\|\,. \tag{2}$$

While mathematical planes are infinite, real mirrors are usually not. If m^r does not lie on the physical mirror plane, c will not receive any signal (see Fig. 2). In this case, the path can simply be removed from the optimization (see Section 3.2).

3.2. Calibration

We optimize our scene setup model by minimizing the temporal differences between time-of-flight measurements t from the real setup and time-of-flight values computed from the current estimate of the setup. If all possible light paths are used there are a total of $\#L \cdot \#M \cdot \#C$ measurements. We solve

$$\underset{S,L,C,M}{\arg\min} \sum_{l\in L} \sum_{c\in C} \sum_{m\in M} \left\| f(S, l, c, m) - t_{l,c,m} \right\|^2 . \tag{3}$$

using a standard gradient descent algorithm (BFGS [30]).

A calibration is only unique up to a rigid transformation of the whole setup since a rigid transformation does not change any path lengths. We can therefore define the camera location S_C as the origin of the coordinate system and determine all other points relative to it. In general we consider the offset between the camera location S_C and the laser location S_L as a known feature of the hardware setup (Relative to the distance to the wall, the offset between S_C and S_L is usually small. In these cases the angle between S_C and S_L viewed from any c or l is marginal and the dominant factor is the total distance from the hardware to the wall).

The initialization is further discussed in Section 4. Due to the compact image formation model automatic differentiation can be used for gradient computation.

3.3. Parameterization

In Eq. (3), we have $l, c \in \mathbb{R}^3$ and $m \in \mathbb{R}^4$ (represented in Hesse normal form). From this general case, specialized parameterizations $g : p \longrightarrow (S, L, C, M)$ can be derived. We implement two such parameterizations for common special cases. A suitable parameterization can decrease the degrees of freedom of the optimization (making it faster and more robust) and enforce certain constraints on the solution.

3.3.1. Planar walls

Most current non-line-of-sight reconstruction approaches assume planar walls (with exceptions such as [23,29]). After defining two basis vectors and an origin, each point on a planar wall can be described by $(u, v) \in \mathbb{R}^2$. As a calibration is only unique up to a rigid transformation we can define the wall plane as the X/Z plane. Then the only remaining parameter of the wall plane is the offset to our origin S. As the mirrors reside outside the plane, their parameterization remains unchanged.

3.3.2. Regular grids

On 2D camera sensors the individual pixels are usually arranged on a regular grid. This grid is projected into the scene along the view direction leading to strong constraints between the relative positions of the projected pixels. In the case of a planar wall this projection can be fully characterized by a homography that maps homogeneous 2D coordinates of the image sensor to 2D coordinates on the wall. Since 2D sensors usually contain hundreds or thousands of pixels, the reduction of degrees of freedom to a constant of 9 (8 for the homography plus 1 for the distance of the wall plane) is significant.

This parameterization can be further generalized by specifying a sensor pattern that is projected onto the wall. Figure 10 shows an example of a pattern where some dead pixels have been masked out. Such a pattern is assumed to be given and not part of the calibration process.

4. Method evaluation

A setup is characterized by a number of different parameters, some of which are easier to change than others. Fixed parameters include those defined by the hardware, e.g., the resolution of the image sensor (the number of camera pixels) and the accuracy of the time-of-flight information. Flexible parameters include the number of laser positions, the number of mirror positions and the quality of the initialization. It is important to understand how these parameters influence the calibration process to choose the best values in practical applications.

4.1. Evaluation setup

Our standard evaluation setup (shown in Fig. 3) consists of 25 camera pixels (arranged in a 5 × 5 grid), 8 laser spot positions and 40 mirror positions. During the evaluation a varying amount of the laser and mirror positions are used. The camera view frustum on the wall is 2 × 2 units and 4 units away from the camera and laser. The laser spot positions are arranged around the view frustum while the mirrors are placed in front of the wall. We use the default case of a planar wall for the majority of the evaluation. To mimic real calibration situations, we apply varying levels of noise to the ground truth geometry to resemble measurement uncertainties. This perturbed data is then used as the initialization for the optimization process, which helps us to assess what level of accuracy is required to successfully estimate the correct geometry. In particular, we apply measurement noise to the setup geometry using

- Gaussian noise with standard deviation of σ to pixel and laser spot positions,
- Gaussian noise with standard deviation of $\sigma/4$ to mirror normals and renormalize them,
- and Gaussian noise with standard deviation of σ to the mirror plane offsets.

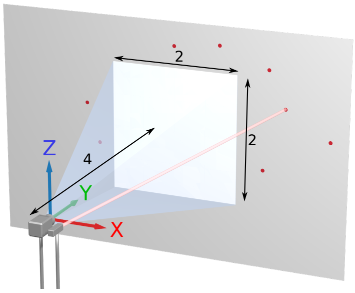


Fig. 3. Setup used for the synthetic evaluation. The camera and laser are in the origin, the red dots mark the laser spot positions. A total of 40 wall-facing mirrors (not shown here) are placed between camera and wall.

It should be noted, that the noise level for positions is measured in distance units while the noise level for normal vectors is measured in degrees, which makes them incomparable. The factor of $\sigma/4$ is used here as it results in similar disturbances for both for this setup.

Similarly, Gaussian noise in various levels is applied to the reference time-of-flight values t. Figure 1 shows the ground truth values along with an example initialization where spatial noise with a standard deviation of $\sigma = 0.5$ was applied. At this noise level not much of the original structure is preserved.

We characterize the quality of a calibration by the root-mean-square (RMS) error between the individual components. Mirror positions are not considered part of the calibration result and thus excluded from the metric. For two setups $P = \{S_1, l \in L_1, c \in C_1\}$ and $Q = \{S_2, l \in L_2, c \in C_2\}$ (e.g. a ground truth setup and a calibration result) we compute

$$\mathrm{RMS}(P, Q) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|P_i - Q_i\|_2^2}. \tag{4}$$

As mentioned before, the calibrated setup might be in a different coordinate system and naively applying Eq. (4) can result in high errors even for actually good result. Therefore we use the Kabsch algorithm [31] to determine an optimal rigid transformation that transforms a setup onto a reference, after which the RMS becomes meaningful. Since the RMS error has the same unit as the initialization noise σ, the two can directly be set into relation. For instance, the example in Fig. 1 uses 4 mirror positions and time-of-flight noise with a standard deviation of 0.02 was applied. It achieves a reconstruction error of 0.042 scene units.

4.2. Required measurements

For a robust optimization the ratio between the input and output dimensions is an important measure. The number of input dimensions of the optimization problem is defined by the amount of measurements (i.e., used paths), while the number of output dimensions depends on the parameterization. For the fully connected case (where all possible connections between lasers, mirrors and cameras are included) there are $\#L \cdot \#M \cdot \#C$ measurements (where the # denotes the number of elements in a set, e.g. $\#M$ is the number of mirrors). The output dimensions are:

- Default: $3 \cdot \#C + 3 \cdot \#L + 4 \cdot \#M$,
- Planar: $2 \cdot \#C + 2 \cdot \#L + 4 \cdot \#M + 1$,
- Grid: $2 \cdot \#L + 4 \cdot \#M + 9$.

The planar parameterization uses 2-dimensional points on the wall plane but has the wall distance as additional dimension. Similarly, the grid parameterization does not depend on the number of camera pixels, instead it always has 9 additional dimensions (8 for the homography plus 1 for the distance of the wall plane).

In Fig. 4 we compare the total number of measurements to the reconstruction error. To analyze the performance of different combinations of laser and mirror positions, we realize the same number of measurements with different combinations. All optimizations use the planar parameterization and are initialized with a noise level $\sigma \in [0, 0.5]$ and a time-of-flight noise level of 0.02.

The results show that the number of measurements alone says little about the structure of the problem, the same number of measurements may lead to severely different errors depending on the ratio between lasers and mirrors. As expected, the reconstruction improves when more measurements are used. More interestingly, it is also beneficial to have about as many laser positions as there are mirror positions: The more extreme the ratio between laser and mirror positions is (for a constant number of total measurements), the worse the results become. This is related to the fact, that the ratio between available measurements and number of variables in the optimization is maximized for equal amounts of laser and mirror positions. For practical applications, the reconstruction error should be close to or below the depth resolution of the

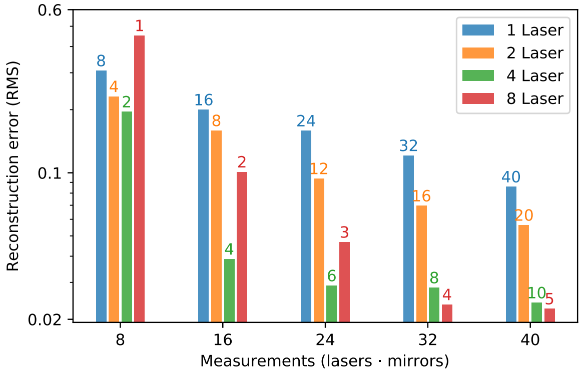


Fig. 4. Calibration performance depending on the total number of measurements. Each data point shows the mean of 100 individual optimizations. The numbers above the bars show the number of mirrors used for that data point.

camera. We conclude that 32 measurements using at least 4 laser positions are a lower bound for a sufficiently accurate reconstruction.

Figure 5 shows the same data set as in Fig. 4, but this time decoded in terms of its dependence on the initialization error. We find that within generous bounds the initialization has no effect on the convergence of the optimization; the RMS error primarily depends on the number of measurements involved.

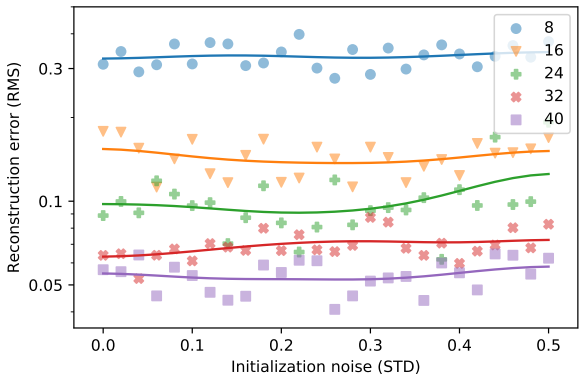


Fig. 5. Reconstruction error for various levels of initialization noise. The data is the same as shown in Fig. 4, averaged over all laser/mirror combinations for a specific number of measurements (shown in different colors / markers).

Figure 6 shows the limits of the allowed initialization error. Even for high values some optimization runs still converge to the correct result, but there are no guarantees and it cannot be considered a safe initialization. Up to a certain threshold close to 1 units, the distribution of reconstruction errors is strongly centered at low RMSE, indicating an accurate result (note the log-log scale). Once this threshold is crossed, the optimization does not converge anymore and exhibits a sudden drop in quality.

We can transfer these insights to form an important rule with respect to the calibration of real setups: We cannot rely on arbitrary initialization values but indeed require a rough knowledge of

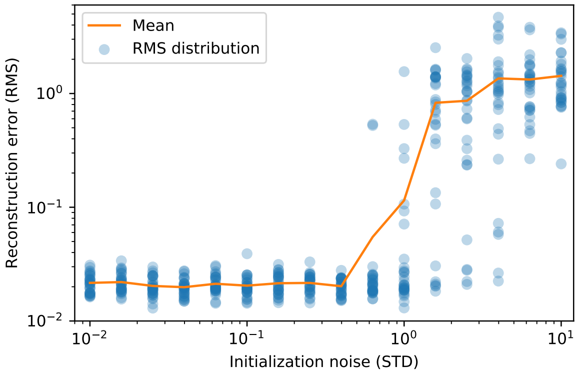


Fig. 6. Reconstruction success depending on initialization noise. Time-of-flight noise is fixed at 0.02. The blue distribution of the individual optimization results gives an better intuition than the orange mean value - the results split in two distinct clusters for increased noise. Note that both axes are in log scale.

the geometry. Still, even a rough estimate is sufficient to yield a very accurate calibration, which might not even require the use of measuring tapes and rulers.

Figure 7 shows the reconstruction error in dependency of the time-of-flight noise and the parameterization. To generate the data the standard setup with 5 mirrors and 6 laser positions is initialized with a random noise value between 0 and 0.5. We find that there is an approximately linear relationship between the uncertainty of the time-of-flight data and the reconstruction error.

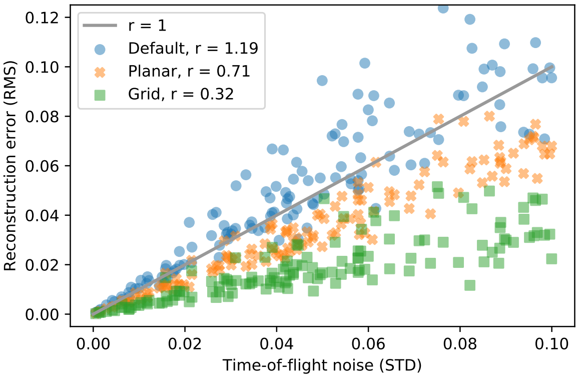


Fig. 7. Calibration error depending on the time-of-flight noise and parameterization. The r-values show the slope of a linear fit for each parameterization.

The default parameterization supports arbitrarily shaped walls such as the curved wall shown in Fig. 8.

Our main findings of this analysis is that for a sufficient amount of measurements, a wide area of safe initialization exists. For optimal results, an equal amount of laser points and mirrors should be used, equally distributed in the scene (but not in a symmetric pattern, which would yield equal values for some measurements). If a setup uses only a single camera pixel or a single laser position for object reconstruction, calibration results can be improved by adding additional camera/laser positions for the calibration and later discard the calibration results of these additional points.

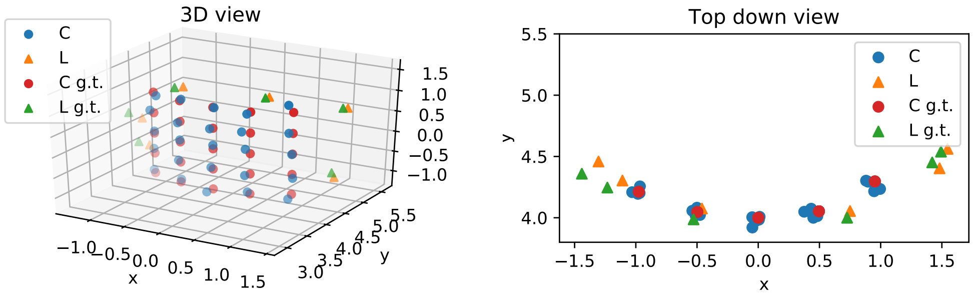


Fig. 8. Example calibration of a curved wall. The setup consists of 6 lasers and 6 mirrors, the initialization noise is 0.5, the time-of-flight noise is 0.1. The RMS error of the calibration is 0.099.

Additional evaluations of the impact of the setup geometry can be found in Appendix C.

4.3. Implementation and runtime

In our prototype Eq. (3) is implemented purely in Python, the optimization of Eq. (3) is performed using the BFGS algorithm from the `scipy.optimize` package with gradients compute by `autograd`. On typical setups, the optimization runs for about 2 minutes on desktop hardware, with unoptimized code.

For large calibration problems with a high number of laser and camera positions, the number of unknowns can be significantly reduced when the planar or grid parameterization is used. For such highly overdetermined problems a significant amount of connections (laser→mirror→camera paths) can be omitted as additional equations in the optimization problem to improve performance.

5. Experimental results

We evaluate the performance of our calibration procedure in a NLoS experiment, and examine the impact of calibration on NLoS reconstruction. In addition to measured data we also use significantly less noisy synthetic time-of-flight data to repeat the evaluation on the same setup geometry in order to emulate additional capture hardware.

The setup is shown in Fig. 9. It uses a total of 7 different laser spot and 7 mirror positions (as described in Section 4.2 this ratio is efficient), and 26 × 29 camera pixels. The pixels are arranged in a regular layout which enables the use of the grid parameterization. The reflector wall is 6.6 m away from the camera, the field-of-view on the wall measures 1.35 m × 1.35 m. The reconstruction target is a house shape which measures 69.5 cm × 54 cm. The mirror measures 80 cm × 100 cm.

The ground truth setup geometry that is used for the evaluation is obtained by manually measuring the position of each mirror, laser spot, camera view frustum corner, and the position of the hidden object using a measuring tape. The shape of the house is given by the SVG file from which it was manufactured.

5.1. Calibration results

Our hardware setup consists of a PrincetonLightwave InGaAs Geiger-mode avalanche photodiode camera and a Keopsys pulsed Er-doped fiber laser. The camera has a spatial resolution of 32 × 32 pixels; however, some pixels are defective, which reduces the effective resolution to 26 × 29 pixels. The temporal bin width is 250 ps (7.495 cm at the speed of light) and each measurement consists of 200,000 individual binary frames captured in about 4 seconds. The laser emits light

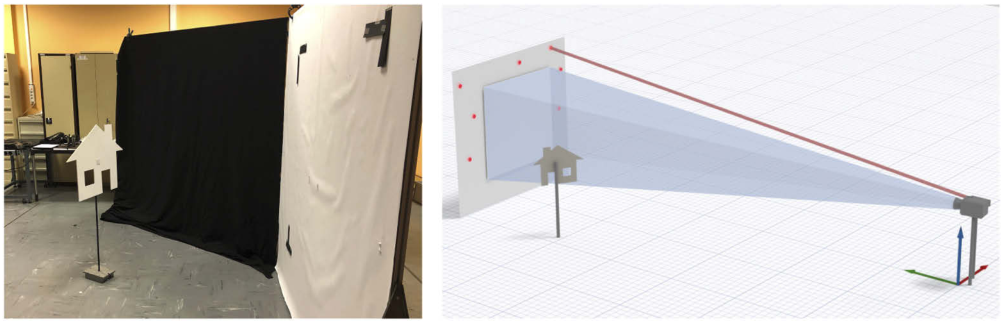


Fig. 9. Photograph and schematic of our experimental setup. The reconstruction target is a house shape outside the field-of-view of the camera. The red spots on the wall show the 7 laser spot positions that are used.

at a wavelength of 1.55 μm and has a pulse length of 500 ps. The transient histograms retrieved from the camera are converted into discrete time-of-flight values by fitting a Gaussian function to the main peak (see Appendix A.1). The house shape is made of white-painted plywood.

We measure the camera's field of view using a moving marker on the wall and observing it in the cameras live image (where the pixel size projected onto the wall is 4.2 cm × 4.2 cm). The 7 spot positions of the near-infrared laser were measured using an IR detector card. We estimate that these measurements are accurate up to 1–2 cm, which should be considered when interpreting the calibration results. The signal offset between camera and laser (which results in a time-of-flight offset) is calibrated by placing a planar calibration target in front of the setup at several known distances. A household-grade mirror is mounted on a tripod which we place at 7 different locations in the scene. The mirror planes were initialized by measuring the position of the tripod over the floor and assuming that the plane normal faces towards the geometric mean of the camera and laser points. Although being a rough estimate, this approach proved sufficient.

Measurements are also affected by scattering (e.g. when the laser spot is close to the view frustum or the laser beam crosses it and hits tiny particles in the air) resulting in invalid values. As our proposed method uses a flexible list of $l \rightarrow m \rightarrow c$ paths, we can automatically detect and remove invalid paths from the optimization (see Appendix A.2 for details on the detection).

Figure 10 shows the calibration results. We evaluate a series of different initialization noise values (see Section 4.1), namely 10, 20, 35, and 50 cm. For each noise level, two different initializations are shown. Note that since the grid parameterization is used, noise is applied to the corners of the view frustum instead of individual pixels (since their layout is given by the sensor pattern). In real applications, the worst case (σ = 50 cm) would correspond to a rough initialization obtained with just a sense of proportion and without any measuring devices.

As seen in the previous evaluation, the calibration usually either converges to a good solution or not at all. For successful calibrations we achieve a typical RMS error of 3–4 cm on this setup. Considering the poor temporal resolution of the setup, these results are consistent with our findings in Section 4.

Additional to the measured time-of-flight data we also use synthetic data representing a more advanced hardware setup. This data is created using the same model as in the inverse optimization. We use identical conditions including removing the same pixels and using the same subset of connections as for the real measurements. We apply noise to the time-of-flight data as described in Section 4 with σ = 0.5 cm. The results are shown in Fig. 10. As expected from the significantly lower noise level, the calibration results are about an order of magnitude better than for the measured data.

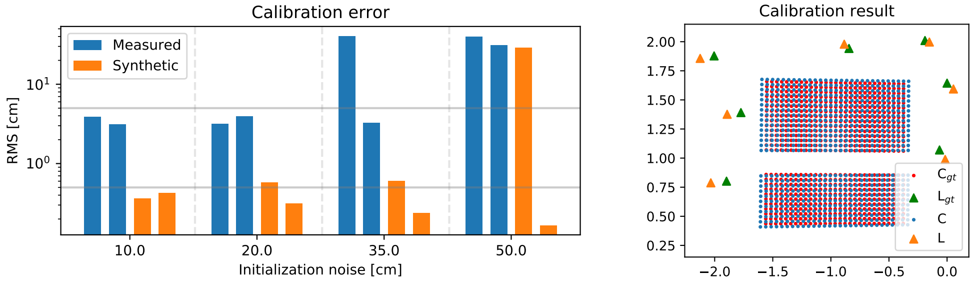


Fig. 10. Calibration quality on the experimental setup. Left: The RMS is computed according to Eq. (4). For each noise level, two initializations are created which are shared by the evaluations on both the measured and synthetic time-of-flight data. The gray lines show the 0.5 cm and 5 cm error. Right: Comparison between a typical calibration result and measured positions on measured time-of-flight data for an initialization noise of 35 cm. The RMS is 3.27 cm. Some rows and columns with dead pixels were removed, resulting in visible gaps.

5.2. Reconstruction results

For object reconstruction, we use the phaser-field backprojection algorithm described in Liu et al. [22]. Since properties like the resolution, the noise level, and general intensity vary between the measured and synthetic data, the reconstruction parameters must be fine-tuned individually (in Fig. 11 parameters are different for each row, but constant within a row). Details about the reconstruction parameters are found in Appendix B.

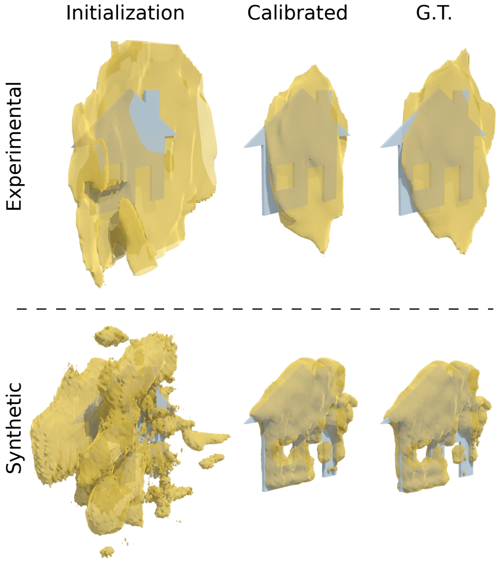


Fig. 11. Object reconstructions obtained from different setups: *Initialization*: the ground truth setup perturbed with a noise level $\sigma = 10$ cm, *Calibrated*: the setup obtained by the presented calibration method, *G.T.*: the ground truth / measured setup. Note that in both cases the calibrated reconstruction closely resembles the ground truth reconstruction, which implies that the calibration was successful.

The synthetic time-of-flight data for the reconstruction cannot be computed with the same approach as for the synthetic calibration since the scene now contains a diffuse object. Therefore we use the transient renderer presented by Iseringhausen et al. [6] which computes the required transient histograms. We set the binning to 0.5 cm, similar to the time-of-flight noise of the synthetic calibration. Additionally we apply shot noise to the transient histograms using a poisson distribution (where the maximal transient pixel intensity is around 1500).

For a quantitative evaluation we use the NLoS mesh distance metric introduced by Klein et al. [32]. It computes the precision (minimal distance to the reference from each point of the reconstruction) and completeness (minimal distance to the reconstruction from each point of the reference) of the reconstruction. Note that in this metric a reconstruction consisting of a single point on the reference surface would have perfect precision but bad completeness score, while a reconstruction consisting of all possible points would have perfect completeness but bad precision score. Thus, there is in some sense a trade-off between both scores which is why the maximum is taken as combined score.

The results are shown in Fig. 12, while Fig. 11 shows reconstruction renderings as qualitative comparison. We evaluate only a calibration with 10 cm initialization noise, as all converged calibrations have essentially the same quality (see Fig. 10). In this case the initial setup (before calibration) can be interpreted as a previously measured setup geometry that is improved through calibration rather than a coarse initialization (which would be obviously unsuitable for any reconstructions) for a first-time setup geometry estimation.

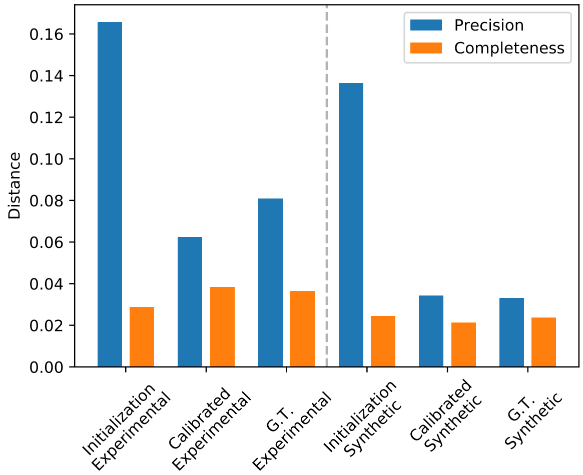


Fig. 12. Bi-directional distance between reconstructions and reference obtained from different setups: *Initialization*: the ground truth setup perturbed with a noise level $\sigma = 10$ cm, *Calibrated*: the setup obtained by the presented calibration method, *G.T.*: the ground truth / measured setup. Smaller values are better, the combined score is the maximum of both (here always the precision). After the calibration, the combined distance is significantly lower.

We make the following observations:

- As expected, the higher temporal resolution and lower noise levels of the synthetic case leads to significantly improved reconstruction results.
- Even for the experimental data where the house shape is not easily recognizable in the reconstructed shape, the shape from the calibrated setup looks much more similar to the shape from the hand measured (ground truth) setup than to the shape from the initialization

setup. This shows that the calibration itself works well, even when the house shape cannot be properly reconstructed.

- Thus, setup calibration is less sensitive to noise than object reconstruction.
- On experimental data, the calibrated setup actually leads to slightly better reconstructions than the hand-measured ground truth setup. As described in Section 5.1 the measured setup has some uncertainties which could be corrected by the calibration (similar to how the initial setup is improved), but the improvement is also close to the general noise level.

6. Conclusion

Our proposed method for non-line-of-sight setup calibration is demonstrated to robustly optimize real-world setups. Despite being a non-convex problem we show that a generous convergence basin exists around the global minimum which results in low requirements of the initialization. While completely arbitrary initialization is not sufficient, a rough estimate that does not necessarily rely on the use of measuring tapes and rulers is sufficient for good results. Additionally, roughly the same number of laser points and mirror positions should be used. The achieved accuracy depends on the depth resolution of the setup, but setup specific parameterizations can be used to enforce constraints and increase the accuracy. As the mirror target results in a single sharp peak in the signal, we do not rely on hardware being able to record full transient histograms. This makes our method applicable on a wide variety of hardware including amplitude-modulated continuous-wave lidars. The ability to calibrate also non-planar walls could enable non-line-of-sight imaging applications in everyday situations.

There are various ways in which our method could be extended in future work. When multiple mirrors are placed in the scene at the same time instead of being measured one-by-one, the mapping between measured peaks and physical mirrors becomes and additional optimization problem. Solving this would allow for faster calibrations.

Although the calibration problem could be reformulated and extended to better support co-axial setups, this might not be worth the effort, since co-axial setups are in general easier to calibrate (see Section 2.).

Additionally the mirror that acts as calibration target could be augmented with a calibration pattern that is then projected onto the wall. This would allow to capture additional information which could possibly be used to improve results. Similarly, including also the intensity of paths could allow to formulate additional constraints on the wall normal.

Appendix

A. Importing SPAD data

As our proposed method works purely on time-of-flight data, each hardware setup requires a pre-processing step to convert sensor data to time-of-flight values. In the following we detail this process for the hardware used in the evaluation in Section 5.

A.1. Distance extraction

For our measurements we use a PrincetonLightwave InGaAs Geiger-mode avalanche photodiode camera where each pixel contains a counter that stops when the first photon is detected. By varying the diode voltage the probability of a photon detection can be controlled and a full transient histogram can be recorded. As the existence of early photons reduces the probability of the detection of later photons, these histograms do not directly correspond to light intensities. However, since our method uses only time-of-flight values and no intensity vales, this effect can safely be ignored.

The pixel counters are synchronized with the laser pulse, but setup-specific features such as cable length between the two devices require an offset calibration. We perform this by placing a flat calibration target at 3 known positions in front of the setup and fitting the offset of a linear function (the gradient is known through the bin width) to the measurements.

Figure 13 shows an example of a pixel histogram. Due to the close proximity of the laser spot to the camera view frustum the histogram contains lens flare artifacts which manifest as a peak at the distance of the wall to the setup. The second peak in the histogram is light reflected by the mirror, our actual signal. The peak shape is widened by the pulse duration of the laser.

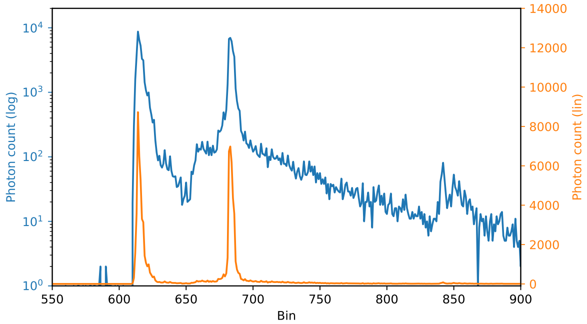
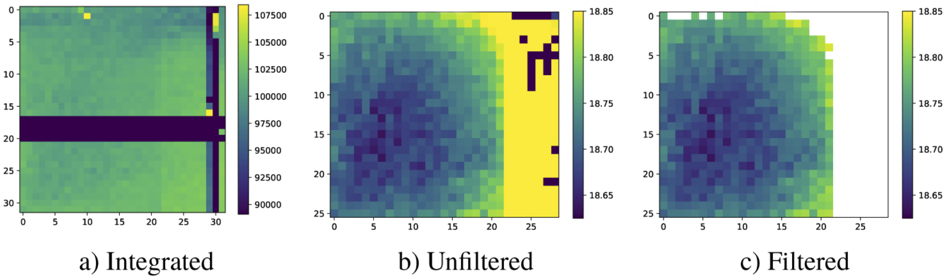


Fig. 13. Histogram recorded by a single camera pixel. Both scales show the same data. The two peaks are well visible in the linear scale (orange). Scattering in the scene produces some background noise after the primary peak, which is visible in the logarithmic scale (blue).

To extract the location of the return with sub-bin resolution, we fit a Gaussian function to the data . Despite this procedure, the overall accuracy is still limited by the camera noise. We employ an iterative scheme, where peaks are located using the Python package `scipy.optimize` and subtracted from the data to find additional peaks in the next iteration. Finally, the fractional bin numbers of the peak locations are converted to time-of-flight values by applying the linear mapping determined in the offset calibration.

Unfortunately some rows and columns in our sensor are broken and contain invalid values. Figure 14(a) shows a raw image of the camera, integrated over time. The dead rows and colums are removed before further processing. In addition to the dead rows in the middle, the first two rows are removed as well, as they contain a single invalid pixel each. This leads to the pixel mask seen in the main paper in Fig. 10.



a) Integrated b) Unfiltered c) Filtered

Fig. 14. (a) A SPAD measurement integrated over time. Some rows in the lower middle and some columns on the right contain invalid data and should be removed. (b) Invalid rows and columns are removed (hence the reduced spatial extent of 29×26 pixels), but some pixels are still invalid. (c) Result after filtering.

A.2. Selecting valid measurements

Apart from the dead pixels most measurements contain additional invalid pixels. The most common cause is that the reflection from the mirror does not cover the whole camera view frustum. We therefore compute a valid pixel mask for each measurement and reject each pixel marked as invalid.

The peak detection finds the highest peak first, so the peaks are sorted by their time delay. The first peak is then the direct reflection while the second peak is our actual signal which is later used for the calibration. Valid pixels are all pixels which fulfill all of the following criteria:

- The relative amplitude of the peaks should not differ by more than 20%. As the absolute intensity can vary drastically for pixels of the same measurement, an criterion on absolute peak amplitudes is less robust.
- The signal peak is at most 20 bins wide. If there are no clear two peak in the signal, the fitting can return a degenerated peak that is extremely wide.
- The first peak is approximately at the distance of the wall (620 bins). We expect a direct reflection from the wall and thus verify it. Note that this test is related to our hardware setup and not the calibration method itself. Knowledge of the wall position is not required for calibration.
- The second peak should have a minimum distance to the first peak (15 bins). This ensures that two actually distinct peaks are detected.

These criteria are rather conservative but robustly remove any outliers. Figure 14 shows the results on a measurement where the mirror reflection did only cover the left part of the view frustum. The mask successfully removes all invalid pixels on the right, however some probably good pixels in the top left are also removed. Since we only aim to reconstruct the overall sensor projection and not individual pixel positions, these holes don't significantly influence the end result.

B. Object reconstruction

We reconstruct the hidden objects in Section 5.2 using the phasor-field virtual wave optics algorithm by Liu et al. [22]. The parameters for the object reconstructions are kept as similar as possible, however the different data sources necessitate some parameter changes.

Due to the lower temporal resolution of the experimental data, a lower wave number is used, which smooths out some noise artifacts without removing true geometry features (experimental: 3, synthetic: 11). Similarly, as the intensity values are different different thresholds are used to convert the density cloud into a surface (experimental: 0.5, synthetic: 0.05).

In the SPAD sensor, early arriving photons can shadow the detection of later arriving ones. For pixel-histograms with a strong first peak (see Fig. 13), the second peak will be lower, even if the same number of photons arrive. Since only distances and not intensities are used for the calibration, this effect can be ignored, however for the backprojection it is advantageous to normalize the intensities of the secondary peak to equalize pixel importance. Since in our setup all pixels are illuminated quite homogeneously, a simple normalization approach yields good results.

C. Setup Geometry

In the following we analyze the influence of the setup geometry on the calibration success. Since in the most general case each laser position, camera pixel and mirror adds 3 degrees of freedom, the effect of their placement is hard to evaluate exhaustively. Instead we evaluate two particularly interesting cases, the influence of the angle between camera and wall and constraining mirror placement to the visible part of the scene.

C.1. Camera angle

We further analyze the impact of the camera angle with respect to the reflector wall using the synthetic setup as described in Section 4.1.

The camera position is rotated around the Z-axis with the rotation origin as the center of the reflector wall (see Fig. 3) in angles between 0° (view direction normal to the wall, as in Fig. 3) and 45° to the right. At each step the camera is oriented such that the center pixel always faces the rotation center.

The projected pixel pattern on the wall is distorted by this rotation: Pixels on the right side are squeezed together, while pixels on the left side are pulled apart. This changes not only the pixel center positions, but also their projected area. To account for this, the hardware agnostic model from Section 4.1 is extended by a pixel model that scales the time-of-flight noise according to the pixel size. This noise scaling is set to the relative difference between the distance of the projected center pixel to the camera and the distance of each other projected pixel to the camera. In practice this means that for the 45° case the time-of-flight noise for the most spread-out pixel is scaled by about 1.41, while the time-of-flight noise for the most squeezed pixel is scaled by about 0.82.

For this evaluation a time-of-flight noise of 0.05 and an initialization noise of 0.2 is used. The setup furthermore uses 8 lasers and 5 mirrors as well as the planar parameterization.

The results are shown in Fig. 15. For each of the 10 steps 16 random instances where calibrated. We find that the distortion from the camera rotation slightly worsens the results, however the effect seems almost negligible. When the noise scaling is turned of, the results have a similar pattern but have an overall lower RMS error (even for no rotation the projected center pixel is the closest to the camera, thus the overall noise scale is >1). Therefore the slight decrease of the RMS is caused mainly by the distorted pixel centers and not just the additional noise from the increased pixel area.

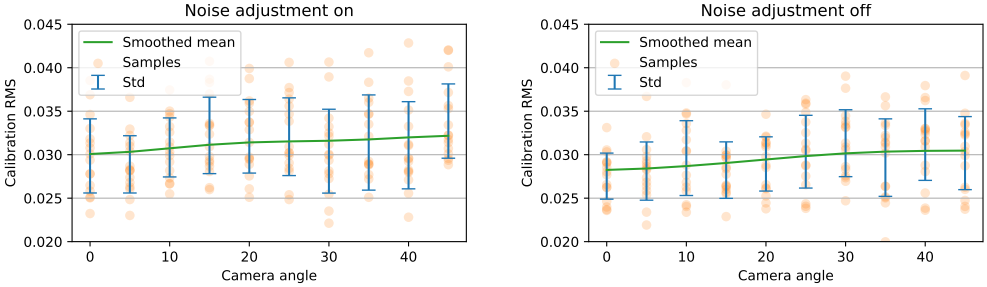


Fig. 15. Calibration error with respect to the angle between camera and reflector wall. *Left*: The time-of-flight noise is adjusted to the pixel size. *Right*: All pixels have the same mean noise.

C.2. Constrained mirror placement

In usage scenarios outside the laboratory the hidden scene might not be accessible. Therefore we perform a comparison between a setup with mirrors only in the visible part of the scene (defined here as having a positive X component in the coordinate system of Fig. 3) and a setup with free mirror placement.

The setup is based on the synthetic setup from Section 4.1. The time-of-flight noise is set to 0.05, an initialization noise to 0.2. 8 laser positions and 6 mirrors are used; in the free mirror placement case 3 are placed in the hidden part of the scene and 3 are placed in the visible part of the scene. The camera is rotated by 30° (as described in Section 1) which is usually required when an occluder is present in the scene. For both cases the calibration was performed 16 times with different initializations.

The results are shown in Fig. 16. We find that in this setup the resulting calibration error is about 25% higher if only the visible part of the scene can be used for mirror placement. We conclude that free mirror placement is an advantage but not a necessity for our method to work.

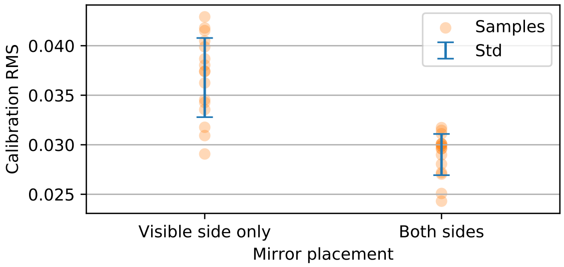


Fig. 16. Calibration results mirror placement constrained to the visible part of the scene and mirror placement in the visible and hidden part of the scene. In both cases 6 mirrors are used.

D. Code

The Python code containing our calibration framework as well as some examples is available Code 1 [33].

Funding

European Research Council (ECHO); Deutsche Forschungsgemeinschaft (HU-2273/2-1).

Acknowledgements

The authors thank Sebastian Werner for many helpful discussions.

Disclosures

The authors declare no conflicts of interest.

References

1. 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," Nat. Commun. **3**(1), 745 (2012).
2. A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using transient imaging," *IEEE International Conference on Computer Vision (ICCV)* pp. 159–166 (2009).
3. F. Heide, M. O'Toole, K. Zang, D. B. Lindell, S. Diamond, and G. Wetzstein, "Non-line-of-sight imaging with partial occluders and surface normals," ACM Trans. Graph. **38**(3), 1–10 (2019).
4. M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," Nature **555**(7696), 338–341 (2018).
5. G. Gariepy, N. Krstajic, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio, "Single-photon sensitive light-in-flight imaging," Nat. Commun. **6**(1), 6021 (2015).
6. J. Iseringhausen and M. B. Hullin, "Non-line-of-sight reconstruction using efficient transient rendering," ACM Trans. Graph. **39**(1), 1–14 (2020).
7. G. Musarra, A. Lyons, E. Conca, Y. Altmann, F. Villa, F. Zappa, M. J. Padgett, and D. Faccio, "Non-line-of-sight three-dimensional imaging with a single-pixel camera," Phys. Rev. Appl. **12**(1), 011002 (2019).

8. M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, "Non-line-of-sight imaging using a time-gated single photon avalanche diode," Opt. Express **23**(16), 20997–21011 (2015).
9. F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2014).
10. O. Katz, E. Small, and Y. Silberberg, "Looking around corners and through thin turbid layers in real time with scattered incoherent light," Nat. Photonics **6**(8), 549–553 (2012).
11. J. Klein, C. Peters, J. Martín, M. Laurenzis, and M. B. Hullin, "Tracking objects outside the line of sight using 2d intensity images," Sci. Rep. **6**(1), 32491 (2016).
12. K. L. Bouman, V. Ye, A. B. Yedidia, F. Durand, G. W. Wornell, A. Torralba, and W. T. Freeman, "Turning corners into cameras: Principles and methods," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 2270–2278 (2017).
13. C. Thrampoulidis, G. Shulkind, F. Xu, W. T. Freeman, J. H. Shapiro, A. Torralba, F. N. C. Wong, and G. W. Wornell, "Exploiting occlusion in non-line-of-sight active imaging," IEEE Trans. Comput. Imaging **4**(3), 419–431 (2018).
14. S. W. Seidel, Y. Ma, J. Murray-Bruce, C. Saunders, W. T. Freeman, C. C. Yu, and V. K. Goyal, "Corner occluder computational periscopy: Estimating a hidden scene from a single photograph," *IEEE International Conference on Computational Photography (ICCP)* pp. 1–9 (2019).
15. W. Chen, S. Daneau, F. Mannan, and F. Heide, "Steady-state non-line-of-sight imaging," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).
16. A. Jarabo, B. Masia, J. Marco, and D. Gutierrez, "Recent advances in transient imaging: A computer graphics and vision perspective," Visual Informatics **1**(1), 65–79 (2017).
17. N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala, "Single view reflectance capture using multiplexed scattering and time-of-flight imaging," ACM Trans. Graph. **30**(6), 1–10 (2011).
18. F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," ACM Trans. Graph. **32**(4), 1–10 (2013).
19. I. Gkioulekas, A. Levin, F. Durand, and T. Zickler, "Micron-scale light transport decomposition using interferometry," ACM Trans. Graph. **34**(4), 1–14 (2015).
20. M. Laurenzis and A. Velten, "Nonline-of-sight laser gated viewing of scattered photons," Opt. Eng. **53**(2), 023102 (2014).
21. V. Arellano, D. Gutierrez, and A. Jarabo, "Fast back-projection for non-line of sight reconstruction," Opt. Express **25**(10), 11574–11583 (2017).
22. X. Liu, I. Guillén, M. La Manna, J. H. Nam, S. A. Reza, T. H. Le, A. Jarabo, D. Gutierrez, and A. Velten, "Non-line-of-sight imaging using phasor-field virtual wave optics," Nature **572**(7771), 620–623 (2019).
23. D. B. Lindell, G. Wetzstein, and M. O'Toole, "Wave-based non-line-of-sight imaging using fast f-k migration," ACM Trans. Graph. **38**(4), 1–13 (2019).
24. A. K. Pediredla, M. Buttafava, A. Tosi, O. Cossairt, and A. Veeraraghavan, "Reconstructing rooms using photon echoes: A plane based model and reconstruction algorithm for looking around the corner," *IEEE International Conference on Computational Photography (ICCP)* pp. 1–12 (2017).
25. C.-Y. Tsai, A. C. Sankaranarayanan, and I. Gkioulekas, "Beyond volumetric albedo–a surface optimization framework for non-line-of-sight imaging," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 1545–1555 (2019).
26. J. G. Chopite, M. B. Hullin, M. Wand, and J. Iseringhausen, "Deep non-line-of-sight reconstruction," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).
27. C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Baraniuk, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," Optica **7**(1), 63–71 (2020).
28. P. Caramazza, A. Boccolini, D. Buschek, M. Hullin, C. F. Higham, R. Henderson, R. Murray-Smith, and D. Faccio, "Neural network identification of people hidden from view with a single-pixel, single-photon detector," Sci. Rep. **8**(1), 11945 (2018).
29. M. La Manna, J.-H. Nam, S. Azer Reza, and A. Velten, "Non-line-of-sight-imaging using dynamic relay surfaces," Opt. Express **28**(4), 5331–5339 (2020).
30. H. W. Tang and X. Z. Qin, "*Practical methods of optimization*," Dalian University of Technology Press, Dalian pp. 138–149 (2004).
31. W. Kabsch, "A solution for the best rotation to relate two sets of vectors," Acta Cryst A **32**(5), 922–923 (1976).
32. J. Klein, M. Laurenzis, D. L. Michels, and M. B. Hullin, "A quantitative platform for non-line-of-sight imaging problems," *British Machine Vision Conference (BMVC)* (2018).
33. J. Klein, M. Laurenzis, M. B. Hullin, and J. Iseringhausen, "A calibration scheme for non-line-of-sight imaging setups - supplementary code," figshare (2020). https://doi.org/10.6084/m9.figshare.12738641.