Transient Non-Line-of-Sight Imaging

Dissertation

zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Jonathan Klein

aus
Siegen, Deutschland

Bonn, Dezember, 2020

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Matthias Hullin

2. Gutachter: Prof. Dr. Wolfgang Heidrich

Tag der Promotion: 15. 04. 2021
Erscheinungsjahr: 2021

Contents

Abstract

Sight is perhaps the most important sense of the human species. But while it allows us to gain a near-instant understanding of our surrounding, it has a fundamental limitation: An object can be hidden from view if it is occluded by an obstacle such as a building or a car. To reveal it, techniques for *looking around a corner* are required, which became only recently available through the use of computational photography.

In most common setups, a laser is used to illuminate a diffuse wall in the visible part of the scene from where the light can bounce off towards the hidden object. From the object, the light is reflected back onto the wall in the visible part of the scene, where it can be detected by a camera. Typically, the camera is a transient imaging camera, which can temporally resolve the propagation of light through the scene when it is illuminated by a synchronized laser. This then allows the recording of temporal light profiles on the reflector wall. The diffuse reflections destroy most of the angular information of where the light was coming from, but leave the temporal offset (caused by the travel time of the photons) intact.

The measured signal is a three-dimensional transient image in which the hidden object is not directly visible. It does, however, encode information about the hidden object which can be used together with physically based models of indirect light transport to attempt a reconstruction of the hidden object. Such a reconstruction is a challenging task which becomes apparent in the limitations of today's system. It thus remains an active field of research that receives high interest from both academia and industry due to its many potential applications.

In this thesis we address some of the main limitations to help the field of indirect vision advance into product-ready technology. Our solutions are presented as a cumulative thesis consisting of three peer-reviewed publications:

In the first publication, we present a novel approach for real-time tracking of hidden objects. So far, setups have relied on expensive hardware and required lengthy reconstruction time. We argue, that sometimes it is more important to have real-time information about the position and movement of a target than a more precise three-dimensional reconstruction that takes minutes to obtain. Furthermore, the analysis-by-synthesis scheme that we use is extensible and works with different types of hardware including non-transient intensity cameras like webcams.

In the second publication, we present a comparison and evaluation platform for the multitude of reconstruction approaches that have been published in the previous years. The results from different research groups are usually coming from different hardware, scenes, reconstruction targets and setup scales. This makes results hard to compare, for example, when two camera system have very different signal-to-noise ratios or a scene is more chal-

lenging than another. In our benchmark, we provide a unified measurement data set that allows to run different reconstruction algorithms on the same input date and also domain specific evaluation metrics to compare the reconstruction results.

In the third publication, we present a flexible calibration algorithm that does not rely on any additional hardware. In order to estimate possible light interactions in the hidden part of the scene, knowledge about the directly visible part of the scene is used. Methods for capturing three-dimensional scenes are established but the requirement of additional hardware would increase the complexity of indirect vision systems even further. Our calibration method only facilitates an additional household-grade mirror which makes it especially suitable for the stage of lab testing in which most of the current research progress happens.

In conclusion, we present a range of contributions that partake in the global efforts of making indirect vision systems available as an additional corner stone of future vision tasks.

Zusammenfassung

Das Sehen ist die vielleicht wichtigste Sinneswahrnehmung des Menschen. Es erlaubt uns in Sekundenbruchteilen unsere Umgebung einschätzen zu können, hat dabei jedoch eine fundamentale Einschränkung: Ein Objekt kann nicht gesehen werden, wenn es durch ein anderes verdeckt wird. Um es dennoch sichtbar zu machen bedarf es Techniken zum *um-die-Ecke-sehen*, die erst kürzlich in dem Feld der Computational Photography (computergestützte Fotografie) entwickelt wurden.

Im üblichen Versuchsaufbau beleuchtet ein Laser eine diffus reflektierende Oberfläche (etwa eine Wand) in dem sichtbaren Teil der Szene. Von dort aus können Photonen in Richtung des verdeckten Objektes reflektiert werden, welches sie wieder zurück in Richtung Wand wirft, wo sie schließlich von einer Kamera wahrgenommen werden können. Typischerweise wird dazu eine *Transient-Imaging*-Kamera verwendet, die die Lichtausbreitung eines synchronisierten Lasers in der Szene zeitlich auflösen kann. Dadurch wird es möglich, ein Zeithistogramm der Lichtstärke auf der Reflektorwand zu messen. Die diffusen Reflektionen zerstören zwar den überwiegenden Teil der Winkelinformationen (aus welcher Richtung das Licht kam), die unterschiedlichen Flugzeiten unterschiedlicher Photonen bleiben jedoch erhalten.

Das gemessene Signal ist ein dreidimensionales, transientes Bild in dem das verdeckte Objekt nicht direkt zu erkennen ist. Es enthält aber kodierte Informationen über das Objekt, die unter Zuhilfenahme von physikalischen Modellen der indirekten Lichtausbreitung für eine Rekonstruktion verwendet werden können. Solch eine Rekonstruktion ist eine sehr herausfordernde Aufgabe was auch in den Limitierungen aktueller Systeme deutlich wird. Es bleibt daher ein aktives Forschungsfeld, das durch sein vielfältiges Anwendungspotential von großem Interesse sowohl für die akademische Welt als auch für die Industrie ist.

In dieser Doktorarbeit werden im Rahmen von drei begutachteten Veröffentlichungen einige der wichtigsten Einschränkungen adressiert und das Forschungsfeld der indirekten Sicht so ein Stück weiter in Richtung Produktreife gerückt:

In der ersten Veröffentlichung stellen wir ein System vor, dass verdeckte Objekte in Echtzeit nachverfolgen kann. Bisher war die Rekonstruktion verdeckter Objekte nur mit teurer Hardware und langen Rekonstruktionszeiten möglich. Wir argumentieren, dass es in manchen Situationen nützlicher ist, Echtzeitinformationen über die Position und Bewegungsrichtung eines Objektes zu kennen, anstatt minutenlang auf eine komplette dreidimensionale Rekonstruktion zu warten. Außerdem ist unser verwendetes Analyse-durch-Synthese-Verfahren vielseitig erweiterbar und kann mit vielen Hardwaretypen angewandt werden, darunter auch nicht-transiente Intensitätskameras wie Webcams.

In der zweiten Veröffentlichung stellen wir eine Evaluationsplattform vor, welche die

verschiedenen Rekonstruktionsansätze der letzten Jahre vergleichbar macht. Die Ergebnisse verschiedener Forschungsgruppen wurden üblicherweise auf sehr unterschiedlicher Hardware, Messaufbauten und Szenen erzielt. Dies macht sie kaum vergleichbar, da beispielsweise unterschiedliche Kameras stark unterschiedliches Rauschverhalten haben können, oder eine Szene komplexer ist, als eine andere. Unser Benchmark liefert einheitliche Messdaten die für eine Vielzahl an Rekonstruktionsalgorithmen als Eingabe dienen können und darüber hinaus angepasste Evaluationsmetriken mit denen die Rekonstruktionsergebnisse verglichen werden können.

In der dritten Veröffentlichung stellen wir ein neuartiges Kalibrierungsverfahren vor, das keine zusätzliche Hardware benötigt. Zur Rekonstruktion verdeckter Objekte wird häufig die Geometrie des sichtbaren Teils der Szene benötigt. Diese kann zwar mit etablierten Methoden erfasst werden, dazu wird jedoch zusätzlich Messhardware benötigt was den Aufbau indirekter Sichtsysteme weiter verteuert. Unser Kalibrierungsansatz benötigt hingegen lediglich einen handelsüblichen Spiegel und ist damit insbesondere für die Kalibrierung von Laboraufbauten geeignet.

Zusammengefasst stellen wir in dieser Arbeit eine Reihe an Beiträgen vor, die zu den weltweiten Anstrengungen, Indirekte-Sicht-Systeme in der Messtechnik zu etablieren, beitragen.

Acknowledgments

This work would not have been possible without the help and support of many people which should not remain unmentioned.

Firstly, I am thankful for my family and friends for providing me with a supportive environment that allowed me to accomplish a project of this scale.

My research at the University of Bonn was supervised by Prof. Dr. Matthias Hullin who invested countless hours in planning, guidance, feedback and practical help. For the most parts, my work was founded by the *French-German Research Institute of Saint-Louis* (ISL) in a fruitful collaboration that was initiated by Dr. Martin Laurenzis who also supervised my research stays in France with equal efforts. Additional funding was provided by the X-Rite Chair for Digital Material Appearance, the German Research Foundation (HU2273/2-1), and the European Research Council (ERC Starting Grant ECHO).

Many of my colleagues in Bonn have become dear friends to me and with many more I enjoyed fruitful conversations or paper collaborations. I do not dare to draw a line between them and instead only refer to the officially listed co-authors here.

For the most projects we performed experiments at the ISL, next to Dr. Martin Laurenzis I received plenty support from Emmanuel Bacher, Dr. Frank Christnacher, Dr. Yves Lutz, Nicolas Metzger, and Stephane Schertzer.

My research stay at KAUST was supervised by Prof. Dr. Wolfgang Heidrich and Prof. Dr. Dominik Michels where I also enjoyed working together with Yuanhao Wang, Dr. Qiang Fu and Dr. Thomas Auzinger (remotely from IST Austria). The KAUST baseline funding from the Visual Computing Center also covered the use of the Shaheen super computer which was used for synthetic data generation.

Finally, I am thankful for Prof. Dr. Andreas Kolb and his group who supervised my Bachelor and Master Thesis at the University of Siegen and paved the way to my doctoral degree.

CHAPTER 1

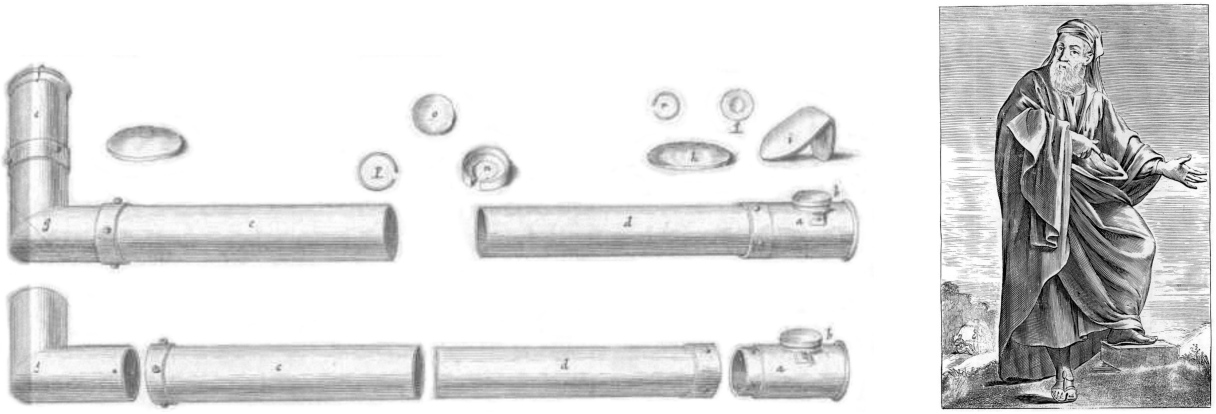
Introduction

In the field of computer vision we teach machines to perceive their surroundings through optical measurements. While traditional digital photography merely aims at recreating a two-dimensional image, computer vision puts a strong emphasis on scene understanding and the extraction of higher level information which are used in a vast amount of applications: 3D scanning enables robots to navigate their environment, medical imaging uncovers fractures and abnormal tissue, objects are augmented with bar codes or QR markers to allow fast and reliable scanning, artificial intelligence algorithms recognize objects and distinguish faces, optical guidance system steer missiles, and even science fiction-like technologies such as autonomous driving are enabled largely by computer vision.

In this thesis we present research work centered around the idea of *non-line-of-sight* (NLoS) imaging. Traditional camera systems perform *line-of-sight* imaging in which the image contains only the parts of the scene that are directly visible to the camera; i.e. the scene points that can be connected with the camera by a straight, uninterrupted line. This results in a two-dimensional projection of the three-dimensional scene, a process through which only a small part of the information contained in the scene is preserved. Everything that lies outside the field-of-view of the camera or that is occluded by some other object will not be visible in the image.

The human eye shares the same limitation and efforts to overcome it can be dated back as long as 8000 years to the oldest known use of mirrors which were found in Neolithic settlements in modern-day Turkey [Eno06]. Several thousand years later mirrors and lenses were then used to build periscopes (from the Greek *Peri* (around) and *skopein* (seeing)), to look around objects in a NLoS fashion (see Figure 1.1a). They were first described in Hevelius' *Selenographia sive lunae descriptio* in 1647; however, he used the term *Polemoskop* for it [Hev47, p. 26].

While periscopes can only be used to look behind occluders that are relatively close (closer than the periscope is long), this limitation was overcome by the introduction of modern day NLoS imaging. Figure 1.2 shows a traffic scenario in which the driver cannot see the cyclist since a house is occluding the view. However, the car is equipped with a NLoS imaging system that aims a laser beam at a second house. From there the light is reflected into the hidden part of the scene (where the cyclist is), and back onto the wall where it can be observed by a camera. It is important to note that the whole system resides inside the



(a) Periscope (b) Empedocles

Figure 1.1: (a) The various parts of a periscope, as sketched by Johannes Hevelius in 1647. The light is guided through the tubes by mirrors, adding an offset to the line of sight and thus allowing to look around objects. (public domain, taken from [Hev47]); (b) Greek philosopher Empedocles (c. 494 – c. 434 BC), who reasoned about the nature of sight (public domain, taken from [Sta55])

car, the house acts as an *accidental* reflector and is viewed as part of the scene and not the imaging system.

Many NLoS imaging systems are based on transient imaging, a technique that aims to resolve the time delay of a light pulse traveling through a scene. In Figure 1.2 the car thus measures the length of the light path traveling from the car to the house, the cyclist, the house again, and back to the car. When different points on the wall are used, this path length changes as well and the hidden scene is reconstructed from this information.

This active imaging approach was already foreshadowed by the Greek pre-Socratic philosopher Empedocles (see Figure 1.1b). In his work *On Nature* (written in verse) he describes the eye as containing fire from which the vision emerges [Bur08, p. 252, 287]. This theory was superseded in Alhazen's *Book of Optics* by the modern view of light traveling from a source towards the eye [Has21], but is surprisingly still held by many laymen today [Win+02]. In active imaging, both principles are combined to extend the capabilities of traditional systems.

The term NLoS imaging, as used in this thesis, describes approaches that attempt to view around occluders (like a periscope). Approaches to look through occluders exist as well (for example in medical x-rays), but are not in the main scope of this thesis.

The field of NLoS imaging is progressing rapidly and there are multiple areas of application in which such systems would be beneficial. With the advent of autonomous driving, traffic safety in scenarios such as the one depicted in Figure 1.2 is increased if the car can recognize the cyclist, and initiate the braking process earlier. In search and rescue missions a NLoS imaging device could be used to find trapped people and rescue them. Similarly, there are many applications in remote sensing and threat prevention. Further down the line, it could also become feasible to utilize NLoS imaging for medical applications such as endoscopy.

Today's limitations in NLoS imaging arise from some principal challenges: The diffuse

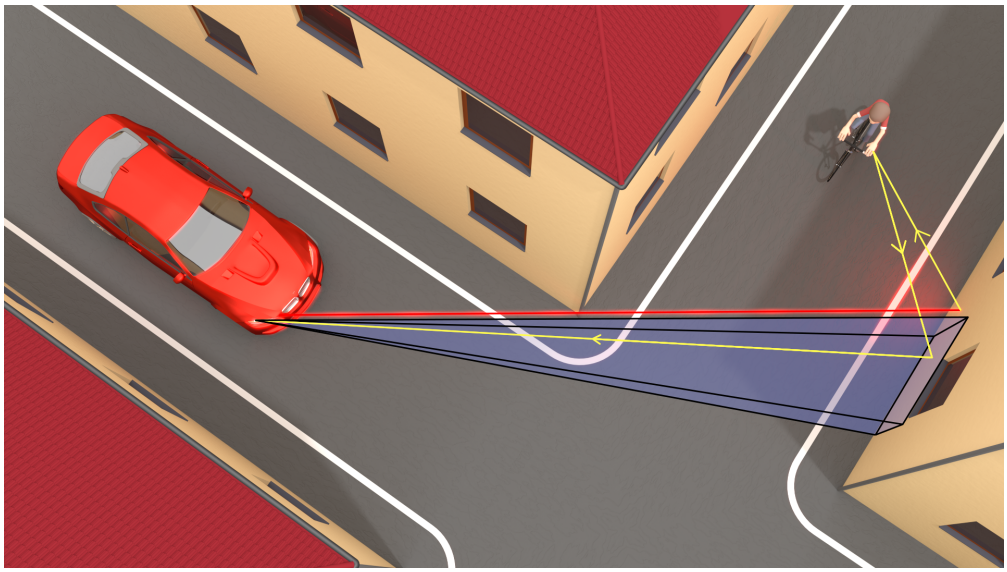


Figure 1.2: Non-line-of-sight imaging. The car is equipped with a laser (red) and a camera (blue field of view). A house wall is used to bounce light to and from the hidden part of the scene (yellow arrows) and thus the driver can 'see' the cyclist around the corner.

reflections destroy angular information and greatly weaken the intensity of the returned signal which in turn increases noise. In order to reconstruct anything at all strong assumptions need to be made, such as assuming the shape of the setup is known, the hidden scene is empty except for the object, the scene is static, materials are known, and the overall setup size is limited. Most work presented so far is therefore much closer to an in-lab proof of concept, rather than an usable product.

1.1 Contributions

Some of these main challenges of NLoS imaging are addressed by our work:

- Especially in their early days, NLoS imaging systems have often been slow and expensive. Real-time information of hidden objects is an obvious requirement for many applications such as autonomous driving and using cheap and widespread components as hardware basis helps widespread usage. In J. Klein et al. 2016 we present a NLoS imaging system that can track objects in real time with consumer-grade hardware [Kle+16].

- To understand the signal characteristics of NLoS setups, synthetic transient renderings are an invaluable tool. Apart from judging whether a certain approximation of the light transport is suitable for a given application, synthetic renderings are also used as training data for machine learning based approaches.

  In J. Klein et al. 2018 we extended a physically based steady state renderer by a transient component to render highly realistic NLoS images [Kle+18]. In J. Klein et al. 2016 we use a less accurate but much faster renderer for inverse problem solving [Kle+16]. Finally, in J. Klein et al. 2016 we provide an in depth analysis of transient images to gain understanding of how light propagates through a scene [KLH16].

- Ongoing research has lead to a wealth of different setups and reconstruction methods. Since new methods are often evaluated on hardware available to a particular research group and custom-built evaluation scenes, direct comparison of reconstruction results from different groups is often impossible. We therefore provide a test suite of universal input data and evaluation metrics in J. Klein et al. 2018 [Kle+18]. With these, results from different reconstruction algorithms are directly comparable and are listed in an online benchmark.

- NLoS imaging algorithms often require geometrical knowledge of the visible part of the scene for calibration. Although extensive research and well established methods for the scanning of line-of-sight scenes exist, naively adding those to a NLoS imaging system usually requires additional hardware. In J. Klein et al. 2020 we thus develop a calibration method that does not require any additional hardware and can be plugged in into a large variety of existing solutions [Kle+20]. Such kind of automatic calibration is helpful to transfer systems from lab environments to real-life situations.

1.2 List of publications

Here we give a list of all publications the author has contributed to during his PhD studies.

1.2.1 Publications in this thesis

The main part of this cumulative thesis is formed by these peer-reviewed, first author publications:

- Jonathan Klein, Martin Laurenzis, Matthias B. Hullin, and Julian Iseringhausen. "A Calibration Scheme for Non-Line-of-Sight Imaging Setups" In: *Optics Express* (2020) [Kle+20]

- Jonathan Klein, Martin Laurenzis, Dominik L. Michels, and Matthias B. Hullin. "A Quantitative Platform for Non-Line-of-Sight Imaging Problems" In: *British Machine Vision Conference (BMVC)* (2018) [Kle+18]

- Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B. Hullin. "Tracking objects outside the line of sight using 2D intensity images" In: *Scientific Reports* (2016) [Kle+16]

1.2.2 Additional publications on non-line-of-sight imaging

The following additional publications on NLoS imaging are either co-authored, not full papers or invited publications:

- Martin Laurenzis, Jonathan Klein, Emmanuel Bacher, and Stephane Schertzer. "Approaches to solve inverse problems for optical sensing around corners" In: *SPIE Security + defense: Emerging Imaging and Sensing Technologies for Security and Defence IV* (2019) [Lau+19]

- Jonathan Klein, Martin Laurenzis, and Matthias B. Hullin. "Wenn eine Wand kein Hindernis mehr ist" In: *photonik* (2017) [KLH17]

- Jonathan Klein, Christoph Peters, Martin Laurenzis, and Matthias B. Hullin. "Non-line-of-sight MoCap" In: *ACM SIGGRAPH Emerging Technologies* (2017) [Kle+17b]

- Martin Laurenzis, Jonathan Klein, and Frank Christnacher. "Transient light imaging laser radar with advanced sensing capabilities: reconstruction of arbitrary light in flight path and sensing around a corner" In: *SPIE Laser Radar Technology and Applications* (2017) [LKC17]

- Martin Laurenzis, Andreas Velten, and Jonathan Klein. "Dual-mode optical sensing: three-dimensional imaging and seeing around a corner" In: *SPIE Optical Engineering* (2017) [LVK17]

- Jonathan Klein, Martin Laurenzis, and Matthias B. Hullin. "Transient Imaging for Real-Time Tracking Around a Corner" In: *SPIE Electro-Optical Remote Sensing* (2016) [KLH16]

- Martin Laurenzis, Frank Christnacher, Jonathan Klein, Matthias B. Hullin, and Andreas Velten. "Study of single photon counting for non-line-of-sight vision" In: *SPIE* (2015) [Lau+15a]

- Martin Laurenzis, Jonathan Klein, Emmanuel Bacher, and Nicolas Metzger. "Multiple-return single-photon counting of light in flight and sensing of non-line-of-sight objects at shortwave infrared wavelengths" In: *Optics Letters* (2015) [Lau+15b]

1.2.3 Additional publications on transient imaging

The following co-authored papers were published the related field of transient imaging:

- Martin Laurenzis, Jonathan Klein, Emmanuel Bacher, Nicolas Metzger, and Frank Christnacher. "Sensing and reconstruction of arbitrary light-in-flight paths by a relativistic imaging approach" In: *SPIE* (2016) [Lau+16]

- Martin Laurenzis, Jonathan Klein, and Emmanuel Bacher. "Relativistic effects in imaging of light in flight with arbitrary paths" In: *Optics Letters* (2016) [LKB16]

- Shuochen Su, Felix Heide, Robin Swanson, Jonathan Klein, Clara Callenberg, Matthias B. Hullin, and Wolfgang Heidrich. "Material Classification Using Raw Time-of-Flight Measurements" In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016) [Su+16]

- Christoph Peters, Jonathan Klein, Matthias B. Hullin, and Reinhard Klein. "Solving Trigonometric Moment Problems for Fast Transient Imaging" In: *ACM Transactions on Graphics (SIGGRAPH Asia)* (2015) [Pet+15]

1.2.4 Other publications

In an extension of his main research focus, the author also published in other fields (where J. Klein et al. 2020 is not yet peer-reviewed):

- Jonathan Klein, Sören Pirk, and Dominik L. Michels. "Domain Adaptation with Morphologic Segmentation" In: *arXiv preprint* (2020) [KPM20]
- Elena Trunz, Sebastian Merzbach, Jonathan Klein, Thomas Schulze, Michael Weinmann, and Reinhard Klein. "Inverse Procedural Modeling of Knitwear" In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019) [Tru+19]
- Jonathan Klein, Stefan Hartmann, Michael Weinmann, and Dominik L. Michels. "Multi-Scale Terrain Texturing using Generative Adversarial Networks" In: *IEEE Conference on Image and Vision Computing New Zealand (IVCNZ)* (2017) [Kle+17a]

1.3 Thesis outline

The remainder of this thesis is organized as follows:

In Chapter 2 the theoretical background on NLoS imaging and transient imaging including relevant hardware types, image formation model, synthetic data generation and back projection based reconstruction are explained. Chapter 3 gives an detailed overview of the historical development and current state-of-the-art in NLoS imaging, including a discussion of problems related to NLoS imaging. In Chapter 4 we present our method for real-time tracking of hidden objects using consumer-grade hardware. In Chapter 5 we present our NLoS reconstruction benchmark and describe available scenes and evaluation metrics. In Chapter 6 we present our method for the calibration of the line-of-sight part of the scene that does not rely on any additional hardware. Finally, Chapter 7 discusses our research in the light of the development that took place after its publication including the impact our work already had and possible extensions of it.

CHAPTER 2

Background

This chapter discusses the basic idea of *non-line-of-sight* (NLoS) imaging as well as the theoretical and practical foundations of it. It starts with a brief discussion of various other methods in the more general field of indirect vision. Next, the idea *transient imaging* (capturing the propagation of light through a scene) is introduced, including an overview of various hardware setups for it. Transient imaging is then used as a fundamental building block for *transient NLoS imaging*, on which this thesis focuses.

The term *NLoS imaging* is not used consistently throughout the literature and a number of variations and alternatives have been proposed over the years. These include (in chronological order): *Looking around corners* [Kir+09], *non-line-of-sight imaging* [Pan+11], *sensing hidden objects* [Gup+12], *diffuse mirrors* [Hei+14], and *occluded imaging* [Kad+16]. However, NLoS imaging has been established as the most popular one and is thus used throughout this thesis.

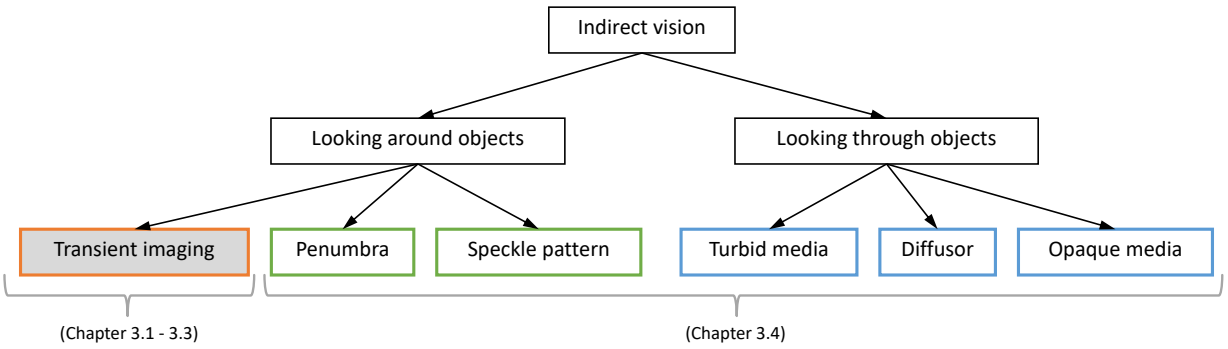


Figure 2.1: Taxonomy of indirect vision methods. This thesis is mostly concerned with transient imaging-based looking around objects. The distinction of methods follows different dimensions: temporal shape of measurement (orange), spatial shape of measurement (green), and occluder transparency (blue).

2.1 Taxonomy of indirect vision

NLoS imaging is a part of the larger field of indirect vision which covers all types of methods to image objects that are in some sense not directly visible. In almost all cases active illumination (a user-controlled light source) is used to achieve more control over the measurements. A taxonomy of the field is shown in Figure 2.1.

When the direct line of sight is blocked, attempts can be made to either look through the occluder or around it. In the first case, different types of occluders require different types of reconstruction methods: Turbid media such as fog or muddy water will scatter the active illumination and create large amounts of stray light. While some light might still travel from the camera to the object and back, this signal is overlayed with the noise of the reflections of the turbid medium. Laurenzis et al. model this as a superposition of two signals and use time-gated imaging to separate them [Lau+12]. However, this approach requires a sufficiently long mean free path length in the turbid medium.

For diffusors such as frosted glass, there exist (almost) no free paths from the object to the camera and all of the signal is distorted. If the diffusor is reasonably thin, the distortion is similar to a diffuse reflection where all angular information is lost but some spatial variation still remains. This case is therefore similar to NLoS imaging where the diffusor takes the role of the relay wall and similar algorithms can be applied to reconstruct the object (see Section 3.4 for more details).

Finally, some opaque objects like cinder blocks are transparent in non-visible wavelengths (such as radio waves) and imaging through them becomes possible using an appropriate wavelength [KM17]. This approach is probably most familiar to the layman from medical X-rays.

For imaging around occluders, a large variety of setups and methods have been proposed. In most setups an optical signal from the hidden object reaches the camera by bouncing off of a diffuse reflector (such as the house in Figure 1.2) which by the nature of diffuse reflection will destroy all angular information (i.e. from which direction the light arrived before reflection).

Apart from some methods that work on pure intensity data (such as our own work in Chapter 4), the lost angular information is replaced by some other type of information which can be used for reconstruction. When the light source emits coherent light, speckle patterns form on the reflector that depend on the relative phase differences and encode geometric information about the scene that can be used for reconstruction. Similarly, occluders in the hidden scene make objects only visible from parts of the reflector wall and thus measuring multiple positions of the penumbra (half shadow) gives clues about where the light is coming from. Both the usage of coherent light and the addition of occluders in the scene therefore create additional spatial detail in the measured signal. Some examples of these two approaches are discussed in Section 3.4.

Lastly, the most popular choice and focus of this thesis is the use of transient imaging which adds an additional temporal dimension in the scale of the speed of light to the intensity measurements. Since light can arrive at the same point on the reflector from many locations at different distances, this temporal dimension is measured as a transient histogram which describes how much light arrived at each point in time. Although the spatial dimension of

the signal is used here as well (as the transient histograms are different at different spatial measurement positions), the spatial resolution is usually significantly lower than for speckle-based or occlusion-based approaches.

In the rest of this thesis, NLoS imaging stands for imaging around occluders using transient imaging, if not stated otherwise.

2.2 Transient imaging

Transient imaging is an umbrella term for various methods to measure the arrival time of light at the sensor.

Since the speed of light is finite, a scene is not immediately illuminated when a light source is turned on. Rather, different objects are illuminated in the order of their distance to the source as the light travels through the scene. The time scale for this process is very small (about 1 ns per 30 cm) and it is usually ignored in normal imaging applications. In transient imaging, however, the goal is to record and make use of this effect.

Commonly cited as the first transient image is the work of Abramson, who used a holographic approach to record a wavefront that is partially reflected by a mirror [Abr78]. More recently, Velten et al. used a streak camera and a femtosecond laser to record a high resolution video of the propagation of light through a plastic bottle [Vel+13].

Since normal light sources (like candles or ceiling lights) emit a continuous stream of photons, different path lengths can not be distinguished for these sources. Thus transient imaging commonly relies on a light source that is synchronized with the detector and those two form a single active imaging system.

Figure 2.2 shows two examples of transient images. In both, the light source illuminates the scene with a short pulse and is placed close to the position of the camera, which causes the car in the second scene to cast a visible shadow. The transient image can then be thought as a video of light propagating through the scene.

In the corner scene, the light propagates down the walls towards the corner (Figure 2.2c), which is the point in the scene that is furthest away from the camera. In d) through f) reflections between the walls are visible; they have a distinct delay due to the longer path length. In g) and h) only higher order reflections are visible, and the scene exhibits some sort of afterglow close to the intersections of the walls. This is also visible in the histogram in b), where the intensity fall-off is much longer than the rise in the beginning.

In the second scene, a complex car model is added which drastically increases the amount of interreflections. Notably the shadow of the car (best seen in the intensity image in a)) occurs in the primary wavefront (seen in f) and g)) but not in the indirect reflections in h).

For scenes where the light source does not coincide with the camera, apparent velocities of the wave do not necessarily correspond to the speed of light but are rather superluminal or subluminal [LKB16]. This optical illusion often makes the understanding of transient images prone to misinterpretation.

A more exhaustive treatment of recent advances in transient images can be found in Jarabo et al. 2017 [Jar+17].

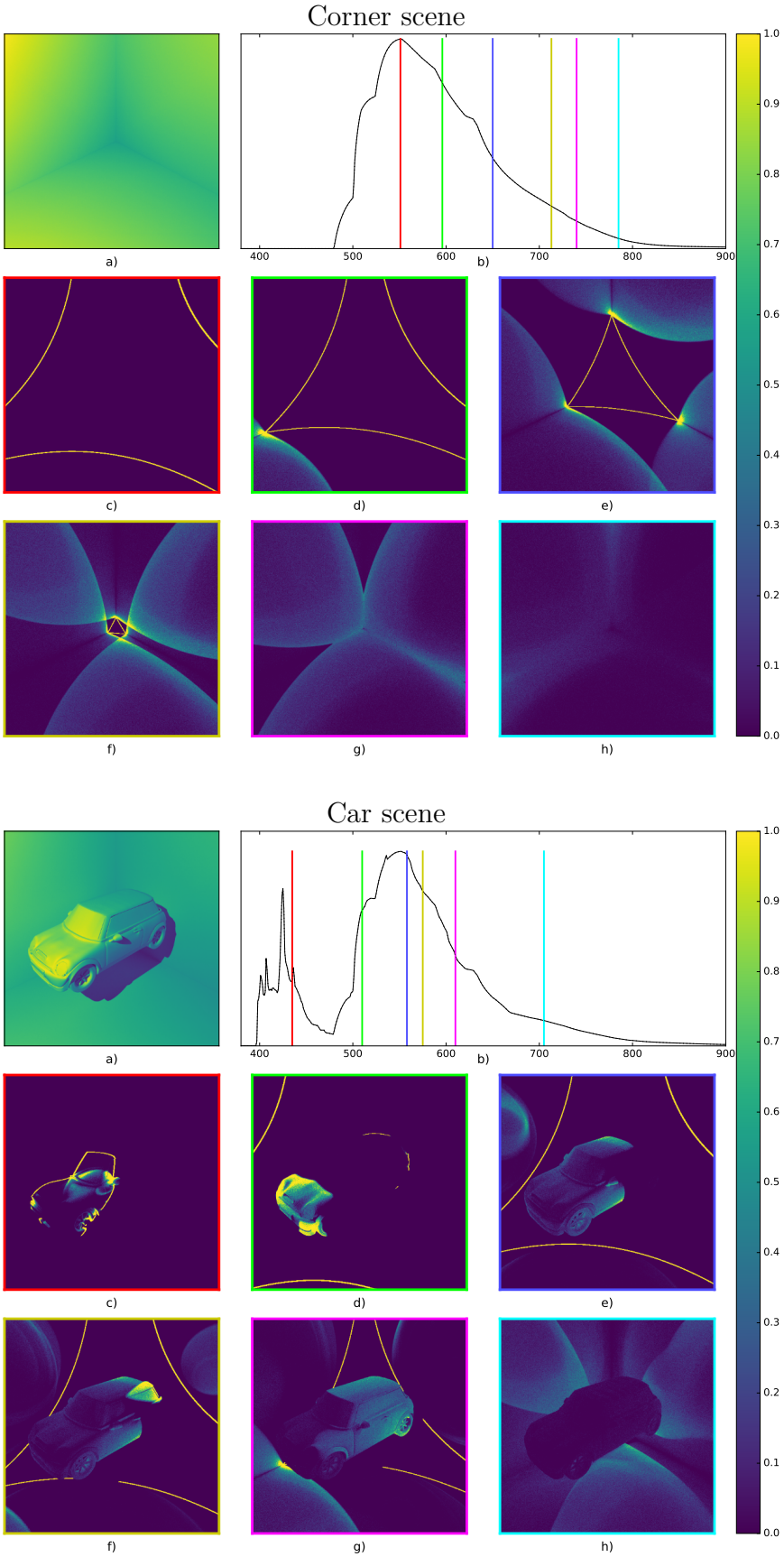


Figure 2.2: Transient images of two scenes. a) Intensity image of the scene. b) Transient histogram of the scene with frame markers. c)-h) 6 frames from the transient image at times marked in the histogram. (Figure previously published in [KLH16])

2.2.1 Hardware

There are a number of hardware platforms available which can be used to measure transient images. Important characteristics are temporal resolution, capture speed, form factor, noise level, and price.

Time-gated cameras

Time-gated cameras are a class of hardware setups that consist of a pulsed laser and a time-gated detector. The detector has a shutter speed in the nano-second range which can be implemented with different hardware components such as Kerr-cells [Kal+93], ICCDs [BH04], or EBCCDs [Wil+95]. The pulse length of the laser and the shutter of the detector define the temporal resolution of the system. Since light is filtered depending on its traveled distance, this imaging modality is commonly called *range-gated imaging*.

Range-gated imaging has been studied since the 1960s [Gil66; SS69]. A prominent use case is imaging through turbid media (such as submarines in muddy water or a car driving through fog), where the range gate can be used to remove stray light from particles floating in front of the object of interest. This drastically increases contrast and can extend the visible range by a factor of 2 to 3 [LV14].

In default operation mode, the recorded image is not so much time-resolved as time-cropped. But by sweeping the timing of the gate over a series of measurements a time resolved image can be retrieved after some post-processing [BH04; And06; LCM07].

Streak cameras

Streak cameras record images with one spatial and one temporal dimension on a two-dimensional image sensor by smearing a one-dimensional line over the sensor over the course of the capture time [Ham08]. The scene is illuminated by an ultra-short pulse of light. Inside the camera, incoming photons hit a photocathode and release electrons which are deflected by a time-varying electric field so that early electrons go towards the top of the sensor and late photons towards the bottom. Although this process allows only the capturing of a single scan line, the achieved temporal resolution is in general higher than for most other systems. Full two-dimensional images can be recorded by sweeping the scanline across the scene and combining the individual measurements. This, however, requires careful calibration and can increase the overall capture time drastically, depending on the desired resolution.

Streak cameras were the first cameras used for NLoS imaging [Nai+11; Pan+11; Vel+12]. However, due to their high price and slow operation speed, alternatives were soon considered.

AMCW Lidar

Amplitude-modulated continuous-wave (AMCW) lidars (LIght Detection And Ranging) illuminate the scene with a continuous wave whose amplitude is modulated in the scale of the scene. On the sensor, the incoming light is correlated with the reference modulation signal from the light source. By some post-processing, the phase of the incoming light can be determined.

Correlation time-of-flight (ToF) sensors (such as the Photonic Mixer Device (PMD) are a lidar technology that allows for cheap multi-pixel sensors with up to 40,000 pixels for independent, parallel measurements [Sch+97]. They are mostly used for range imaging (e.g. in the Microsoft Kinect v2 camera) since they cannot directly measure a full transient image. However, when multiple measurements of different frequencies are combined, individual paths can be distinguished [Kad+13; KBC13; Dor+11] or even full transient images computed [Hei+13; Pet+15].

Although some NLoS related work using AMCW cameras has been presented [Hei+14], the difficulties of measuring transient images directly makes them a somewhat unpopular choice. As widely used consumer hardware they are however small and cheap.

SPAD

Single-photon-avalanche-diodes (SPAD) contain photo diodes in which a single photon triggers an electron avalanche [Zap+07]. While conceptually similar to avalanche photo diodes, they are operated with a voltage above the reverse-bias breakdown voltage. In this so called *Geiger-mode*, the response of the diode is no longer linear but exponential which leads to an extremely high sensitivity and the ability to record individual photons. During capture, the scene is illuminated by an ultra short light pulse from a laser which is synchronized with a time counter in the pixel. The counter starts when the laser is triggered and stops as soon as the first photon is detected and therefore records the time of flight. Since photons are only detected with a certain probability, repeating the measurement yields full transient histograms. However, for each measurement each pixel can only detect a single photon and after that remains blind for the rest of the measurement time. This means that later-arriving photons are shadowed by early ones and ultimately results in the histogram not being proportional to the actual intensity values [HGJ17]. With frame rates of several hundred thousands, transient images with reasonable noise level can be captured within seconds.

SPADS are commonly available as single-pixel sensors [RGH09], 1D sensors [BBC17], or 2D sensors [Bur+14]; however, their resolution is significantly lower than that of traditional CCD sensors. First used by Buttafava et al., they quickly became a popular tool for NLoS imaging.

Single pixel SPADs are nowadays cheap enough to be found in various consumer hardware, where they are usually used as proximity sensor [STM20].

Other hardware

Apart from transient cameras there are also other hardware setups that can measure some form of time resolved data. The basic principle of AMCW lidars is inspired by radar systems (RAdio Detection And Ranging) which were developed during the second world war and remain one of the most common range measurement systems to this day [Ric14]. Similarly, in sonar systems (SOund NAvigation Ranging) acoustic waves are used for distance measurement, especially in under water scenarios where the attenuation is weaker [Uri83]. However, these systems measure the time difference of discrete pulses, rather than the phase shift between continuous waves.

2.2.2 Simulation

Synthetic transient images can be created through simulation. This offers several advantages such as a precise control over the scene, absence of camera noise (if not explicitly modeled), and cheaper costs since no hardware is required.

Since transient images are a superset of conventional intensity images, they can be computed with similar algorithms. Such physically-based rendering algorithms are well-researched and ongoing research is now mostly focused on performance improvements or special cases.

As a prominent example, path tracing is a versatile rendering algorithm to solve the *rendering equation* [Kaj86], and is readily extended to output transient images by keeping track of the path lengths (see Section 2.3.1). However, the additional temporal dimension imposes several difficulties. Not only does it increase the output size and thus linearly scales the amount of samples required, but is also hard to sample directly [PVG19].

Transient rendering was first presented by A. Smith, Skorupski, and Davis [SSD08] and subsequently improved by various other groups [Jar+14; PVG19; SC14]. Some works also include the simulation of camera sensor to produce realistic, synthetic measurements [Kel+07; LHK15; MNK13].

In our own work, we extend *pbrt* (a physically based ray tracer developed by Pharr, Jakob, and Humphreys [PJH16]) by a transient component and implement a new importance sampling strategy, specifically tailored to NLoS imaging setups (see Chapter 5).

Simulation of transient images also enables inverse problem solving. In many cases, the constrained geometry of the setup can be exploited to derive more efficient renderers. For example, full path-tracing is not required if there is only a very constrained set of possible paths. Also full photo realism is often not required to find a suitable solution, which allows for further speed-up. An example of this can be seen in our own work, see Chapter 4.

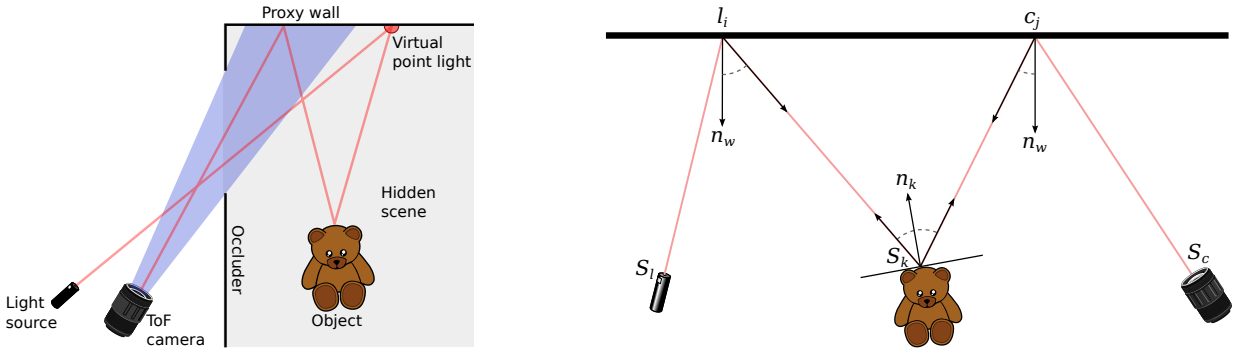
2.3 Transient non-line-of-sight imaging

Here we cover the basic principles of transient NLoS imaging. The various extensions that have been presented over the last years are discussed in the next chapter.

The idea of using a relay surface (most commonly a diffuse wall) to reflect light to and from the hidden scene was first presented by Kirmani et al. [Kir+09] and is commonly called 3-bounce setup (since the light is reflected by the wall, the hidden object and the wall again). It is depicted in Figure 2.3a and is used almost ubiquitously for transient NLoS imaging.

Since the wall is commonly assumed to be diffuse, angular information is destroyed through the reflection. By using time-resolved (transient) measurements, the lack of angular information is in some sense offset by additional temporal information and a reconstruction can be attempted.

The 3-bounce setup makes a couple of assumptions, for example there are many more potential light paths than the one depicted in Figure 2.3a. After introducing notation we will discuss how the light transport in NLoS scenarios can be modeled and how the hidden scene can be reconstructed from transient measurements.



(a) NLoS 3-bounce setup

(b) Identifiers

Figure 2.3: a) Light travels from the laser to the wall, is reflected into the hidden scene, bounces off of the hidden object, and is finally reflected by the wall again towards the camera. b) Nomenclature of the position and normal vector along the light path.

Notation

Figure 2.3b shows only a single light path from the light source (usually a laser) to the camera.

We denote the physical location of the laser and the camera with S_l and S_c. We then call the position of the laser spot on the wall a *laser point* l, and the projection of a single camera pixel a *camera point* c. Since a single light path is in practice not sufficient for reconstruction [Ped+17], tools like galvanometers are often used to scan multiple points on the wall with the laser. The camera also observes either multiple positions concurrently (if it contains a pixel array) or a similar scanning technique is use for single-pixel cameras. We then have a whole set of laser and wall points, denoted as $l_i \in L$ and $c_j \in C$. The wall usually has a constant normal denoted n_w, if not, indices are added accordingly. The hidden object is at position S. Usually, the object is not a single point but a manifold that is approximated by a finite number of points. We call these points S_k and their normal vectors n_k.

By definition, L and C are part of the visible scene. Given a calibration of the hardware position relative to the wall, the distances between laser and wall $\overline{S_l l_i}$ and the distances between wall and camera $\overline{c_j S_c}$ can be calculated and removed from the transient measurements. These *inner* distances are then independent from the hardware position which simplifies later reconstruction. Since no information is lost in the process, by default all measurements are normalized in this fashion.

2.3.1 Image formation model

Light propagation in general scenes is a well-researched topic and its simulation is the main focus of computer graphics research. The two most common ways to describe light transport are the ray model and the wave model. The former is faster to compute but less general and does not model effects such as diffraction which is only possible using the wave model.

For the ray model, a very general description of light propagation through a scene is given by the *rendering equation* [Kaj86].

In the following, we will derive the NLoS image formation model from general light

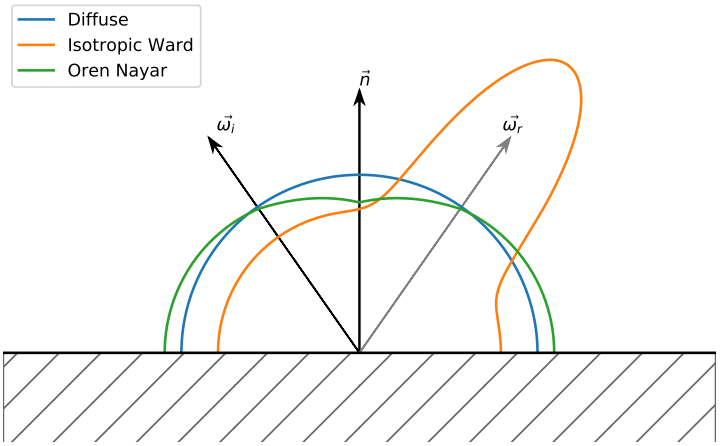


Figure 2.4: Various BRDF models plotted for a fixed incident vector $\vec{\omega_i}$ and normal vector $\vec{n}$. The gray vector $\vec{\omega_r}$ shows the direction of a perfectly specular reflection. *Blue:* A perfectly diffuse (Lambertian) model. *Orange:* The isotropic Ward model [War92] with strong specular component. *Green:* The Oren-Nayar model [ON94] which exhibits a slower fall-off for flat angles, compared to the diffuse BRDF.

transport theory. This gives not only a mathematical formulation to more efficiently handle NLoS light transport but also makes the various assumptions and approximations clear. The light transport model consists of a local component (the material model) and a global component (described by the *rendering equation*).

Material models

Real-world materials influence the way objects reflect light. In computer graphics this effect is often modeled by so called *bidirectional reflectance distribution functions* (BRDF) [PJH16, p. 348]. A BRDF describes the fraction of light that is reflected from an incoming direction ω_i towards an outgoing direction ω_r and is thus a four-dimensional function (with two angles per direction). Commonly it is denoted as $f(\omega_i, \omega_r)$.

A number of variations exist, for example the function can also depend on the location, or the wavelength, or even model subsurface scattering. These extensions increase the number of dimensions further and thus the simpler models are more commonly used.

Figure 2.4 shows radial plots of 3 common BRDFs. Note that in the plots ω_i and ω_r lie in the same plane and ω_i is fixed, which results in the plots being one-dimensional. For physically based BRDFs the total amount of reflected light should be equal to the total amount of incoming light, if the material does not absorb light (and convert it into heat) or emit light itself. Therefore materials with a specular component such as the Ward BRDF in Figure 2.4 must reflect less light in directions outside the specular highlight compared to a perfectly diffuse BRDF.

The rendering equation

The *rendering equation* describes global light transport in a scene. First presented by Kajiya [Kaj86] (and in a similar form by Immel et al. [ICG86]) it computes the radiance (the

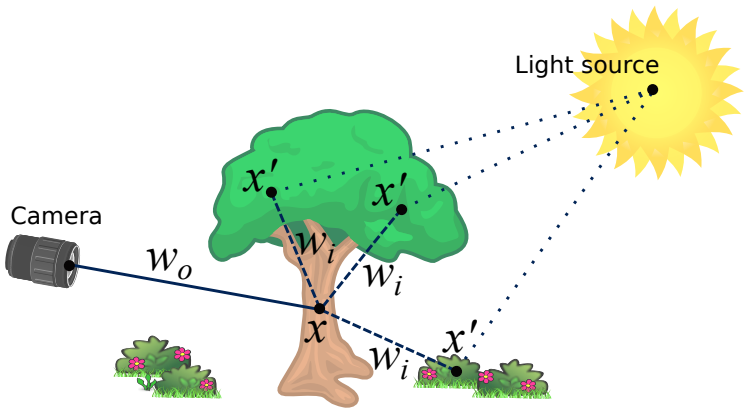


Figure 2.5: Illustration of the rendering equation. The light reflected by point x towards the camera depends on the light reflected by all other points x' in the scene and not just the sun itself.

power per unit projected area per unit solid angle) that a scene point emits or reflects into a certain direction:

$$L_o\left(x, \vec{w_o}\right) = L_e\left(x, \vec{w_o}\right) + \int_A g\left(x, x'\right) \cdot f_x\left(x, \vec{w_o}, \vec{w_i}\right) \cdot L_i\left(x, \vec{w_i}\right) \cdot \frac{\cos\theta_o \cdot \cos\theta_i \mathrm{d}A'}{\left\|x' - x\right\|^2} \tag{2.1}$$

Here

- $L_o\left(x, \vec{w_o}\right)$ is the outgoing radiance at a point x in direction $\vec{w_o}$,
- L_e is the radiance emitted from the surface,
- L_i is the incident radiance from another scene point,
- x and x' are points on surfaces in the scene,
- $g\left(x, x'\right)$ is the geometry term that models potential occlusion between x and x',
- θ is the angle between the normal vector at point x and $\vec{w}$, and
- f_x is the BRDF at point x.

Figure 2.5 shows the role of the points x and x' and the directions $\vec{w_o}$ and $\vec{w_i}$. In a local illumination model, L_o would only depend on the surface of x and the light source, (e.g.. the sun). However, this ignores the influence of other parts of the scene. In the figure, the point x is in the shade but receives light reflected from other leaves and nearby objects (denoted as x'). These receive direct radiance from the sun, but also from other parts of the scene. In general, the radiance of every point in a scene depends on every other point in the scene, with infinite recursion.

The geometry term g is used to model occlusion in the scene. If x' is not visible from x, g is 0, otherwise it is 1. The incoming radiance at x also depends on the distance to x' (the *inverse-square law* [PPP93, p. 12]) and the relative orientations of the surfaces (the *Lambert's cosine law* [PPP93, p. 13]).

Computing L_o results in an infinite recursion which must be numerically approximated for all non-trivial scenes. Algorithms like *ray tracing* and *path tracing* can be seen as approximate solvers for the rendering equation [PJH16, p. 12].

While being very general, the rendering equation still has a couple of limitations, some of which are:

- As a ray model, wave effects such as diffraction are not modeled.
- It assumes free space between two surface points and does not cover volumetric scattering [PJH16, p. 671].
- It only describes light transport and not imaging hardware. Thus simulating the output of an actual measurement setup requires additional hardware models.

NLoS light propagation

Being a subset of general scenes, the light transport in NLoS imaging scenes is described by the rendering equation. With the typical assumptions of the 3-bounce setup however, the transport model becomes rather restricted which allows for a much simpler formulation.

Figure 2.3b show a single light path in the 3-bounce setup. Here, the incident radiance at a camera position S_c is computed as

$$\begin{aligned} L\left(c, \overrightarrow{cS_c}\right) = L_e\left(S_l, \overrightarrow{S_l l}\right) \cdot \sum_k & \cos\left(\overrightarrow{cS_k}, n_w\right) \cdot f_w\left(\overrightarrow{cS_k}, \overrightarrow{cS_c}\right) \\ & \cdot \cos\left(\overrightarrow{S_k c}, n_k\right) \cdot \cos\left(\overrightarrow{S_k l}, n_k\right) \cdot f_{S_k}\left(\overrightarrow{S_k l}, \overrightarrow{S_k c}\right) \\ & \cdot \cos\left(\overrightarrow{lS_k}, n_w\right) \cdot f_w\left(\overrightarrow{lS_l}, \overrightarrow{lS_k}\right) \\ & \cdot \frac{1}{\left\|\overrightarrow{lS_k}\right\|^2} \cdot \frac{1}{\left\|\overrightarrow{S_k c}\right\|^2}. \end{aligned} \tag{2.2}$$

Equation 2.2 is derived from Equation 2.1 through a number of assumptions:

- Only one laser point l is illuminated at the same time. The laser is focused on l and all light emitted at S_l is reflected at l. Therefore l can be thought of as the unique, point-shaped light source in the scene.
- Since the laser beam is focused, the distance falloff does not apply to $\overline{S_l l}$.
- The projected area of the camera pixel is the point c. In reality the camera pixel would integrate over a small area around the projected pixel center c and the pixel read-out would be the average of this area. Additionally the distance falloff and the projected pixel area cancel each other out and the wall has the same brightness for every camera distance.
- The scene consists only of the reflector wall and the hidden object. The physical laser and camera do not interact with the light propagation but only act as light source and sink. There are no other walls, no floor, and no ceiling.

- The object is discretized in a set of discrete points S. (This turns the integral into a finite sum.)
- Surface points that are not visible from l or c are excluded from S.
- There are no interreflections on the object (i.e. there is no light path $\overline{S_a S_b}$. Either because the object is convex, or because they are ignored.
- The distance between the object and the wall is sufficiently large such that higher order reflections (wall $\rightarrow$ object $\rightarrow$ wall $\rightarrow$ object $\rightarrow$ wall) have a very small contribution and can be ignored.

These rather strong assumptions break the infinite recursion of Equation 2.1 and leave only a finite sum over the object's surface for each combination of l and c. If the object is point-shaped, there is only a unique path with a unique length which allows for straightforward position detection (see Section 2.3.2).

For real measurements, none of the above assumptions are perfectly met. Sometimes experiments can be designed to satisfy them better (e.g. by covering the floor and back walls of the scene with black material to avoid background scattering) and some algorithms rely less on these assumptions (such as machine learning-based approaches, see Section 3.2) but otherwise these discrepancies will deteriorate reconstruction results.

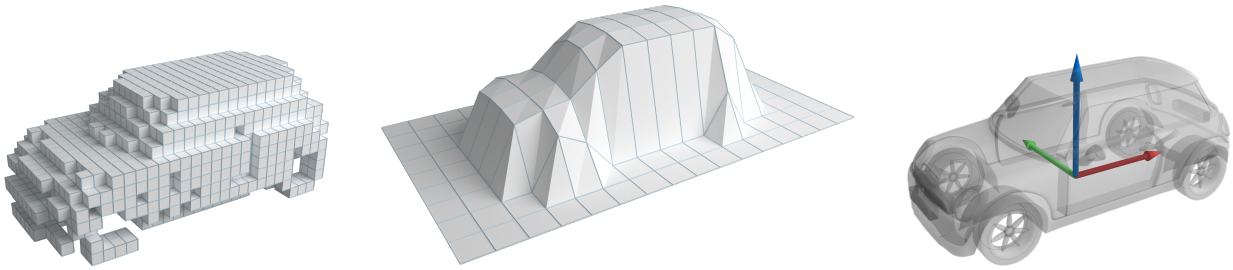
2.3.2 Reconstruction

Over the years a diversity of NLoS reconstruction methods have been proposed. Here we discuss some theoretical foundations and review a simple reconstruction approach to build some intuitive foundations before more algorithms and their differences are discussed in Chapter 3.

All reconstruction algorithms take the transient images as input but can have different forms of output. Depending on the parameterization of the output, different types and amounts of detail are reconstructed. Figure 2.6 shows an example of a scene containing a car that is represented in three different ways: A voxel volume, a height map, and a rigid transformation (position and orientation) of an object with known shape. While the first two are different instances of full geometry reconstruction, the later would occur in object tracking methods.

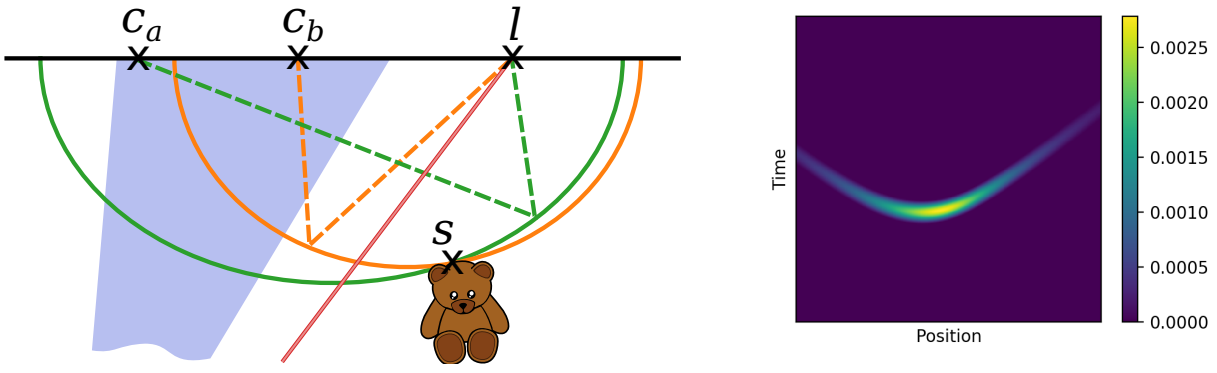
Different parameterizations have different amounts of degrees of freedom (DoF) and as a rule of thumb the more DoFs are to be determined, the harder the reconstruction problem becomes. In practice the choice often depends on the application. In traffic scenarios it might be more important to know just the position and velocity of a hidden vehicle than its shape, while other applications such as remote observation might require full geometry reconstruction for object identification. The parameterization might also contain other scene properties such as materials.

Some exotic parameterizations are used as well, for example Tsai et al. parameterize the surface of a known object to add finer surface detail [TSG19].



(a) Voxel (b) Height map (c) Rigid transformation

Figure 2.6: Different levels of reconstruction with varying amounts of degrees of freedom (DoF). a) Voxel grid, > 10.000 DoF. b) Height map, > 1000 DoF. c) Object position, 3 DoF. (Figure previously published in [KLH16])



(a) Ellipsoidal back projection (b) Transient image

Figure 2.7: Back projection principle. a) All possible reflection points for a path with constant length between l and c lie on an ellipsoid. This reflection point is the intersection of all ellipsoids. b) Transient image of a NLoS scene, showing light arriving at different parts of the wall at different times from a single reflection point. Finite resolution smears out the parabola-shaped line.

Back projection

The baseline algorithm for NLoS reconstruction is ellipsoidal back projection. Originally presented by Velten et al. [Vel+13] it has been used and improved in many publications throughout the last years (see Chapter 3). Due to its simplicity it is well suited to give an intuitive understanding of how the temporal information from transient images can be used for NLoS reconstruction.

The basic principle is shown in Figure 2.7a. Given transient measurements from a setup such as shown in Figure 2.3a, the light of each normalized measurement C_i and L_j originated from somewhere on an ellipsoid with the focal points C_i and L_j (since by definition an ellipsoid is the set of all points that result in the same travel time). For an ideal point-shaped object and perfectly precise measurements, the ellipsoids for any combination of C and L intersect at the same point, which then is the position of the object. In practice, temporal and spatial resolution is limited and for each camera point light arrives at slightly different times as shown in the transient image in Figure 2.7b. Each non-zero pixel corresponds than to an ellipsoid and the hidden scene is discretized into a voxel grid in which each ellipsoid is back-projected, adding up their contributions. Multiple laser positions can be used by back-projecting all corresponding transient images into the same voxel grid. The voxel grid is interpreted as a probability map with the object located at its maximum. By thresholding the probability map it is also possible to retrieve the approximate object geometry. Since effects like interreflections or occlusion are not handled by the model, the geometry reconstruction quality is limited.

2.3.3 Challenges

So far, NLoS imaging is still an emerging technology with no available end-user products. It has almost exclusively been demonstrated in lab environments (with some exceptions [Sch+20; LWO19]) and the development is slowed by some intrinsic challenges:

- Diffuse reflections destroy angular information. However, this 'limitation' is what sparked the whole field of research in the first place and the goal is therefore to circumvent it (by finding algorithms that do not require angular information) rather than lifting it (by simply using mirrors instead of diffuse walls).

- Due to multiple diffuse reflections, light levels are very low. Equation 2.2 shows that if the hidden object is at a distance d from the wall, the light intensity is proportional to d^{-4}; a significant reduction. Using eye-safe infrared light the illumination power can be increased to some extent, but in practice the scene size is still limited. While low light levels do not directly destroy the signal, they increase the noise substantially and also make it hard to even differentiate the NLoS signal from background light.

- Although transient imaging hardware is becoming more widespread (and even smart phones come equipped with single SPAD pixels [STM20]), transient imaging hardware is still more expensive and less accessible than traditional cameras, especially when high temporal or spatial resolution is required.

- In order to perform the reconstruction, the visible part of the setup must be known. This adds a certain calibration overhead that is not present in traditional imaging.

CHAPTER 3

Related work

NLoS imaging research developed from a simple idea into a diverse field of research. Efforts have been made to improve on the various dimensions of the problem such as reconstruction quality, capture and reconstruction speed, setup constraints, and more. Current methods weight a compromise among these aspects by improving one dimension and sacrificing performance in another, or they allow to solve a special case particularly well.

Nowadays the field of NLoS imaging consists of a large number of individual publications with a varying degree of impact. This chapter aims to briefly discuss what we feel are the most impactful. To address the various dimensions, efforts were made to group similar publications and explore the historical development in each. The selection is furthermore strongly focused on NLoS methods using transient images (see Figure 2.1). While some publications outside this focus are briefly mentioned, a more thorough overview can be found in Maeda et al. [Mae+19].

NLoS system dimensions

Figure 3.1 shows an overview of the various dimensions (grouped into performance characteristics and implementation aspects) that describe NLoS systems. While the performance characteristics are of interest for practical applications, the implementation aspects are useful to describe the academical development.

Reconstructed information Different applications require different types of information from a scene. As described in Section 2.3.2, some are easier to reconstruct than others.

Supported materials The image formation model inside the reconstruction algorithm often dictates what material types can be reconstructed. The most common case are diffuse objects.

Scene complexity The supported scene complexity imposes strong constraints on possible use cases. A common assumption for tracking is that the scene consists of a single, point-shaped object. Planar scenes yield simpler reconstruction problems, while partial occlusion (where some scene points are only visible for specific c_i and l_j combinations) is especially hard.

Practicability Most setups have only been demonstrated in small (room sized) indoor environments. Uncontrolled environments (especially outdoor) yield a lot of noise

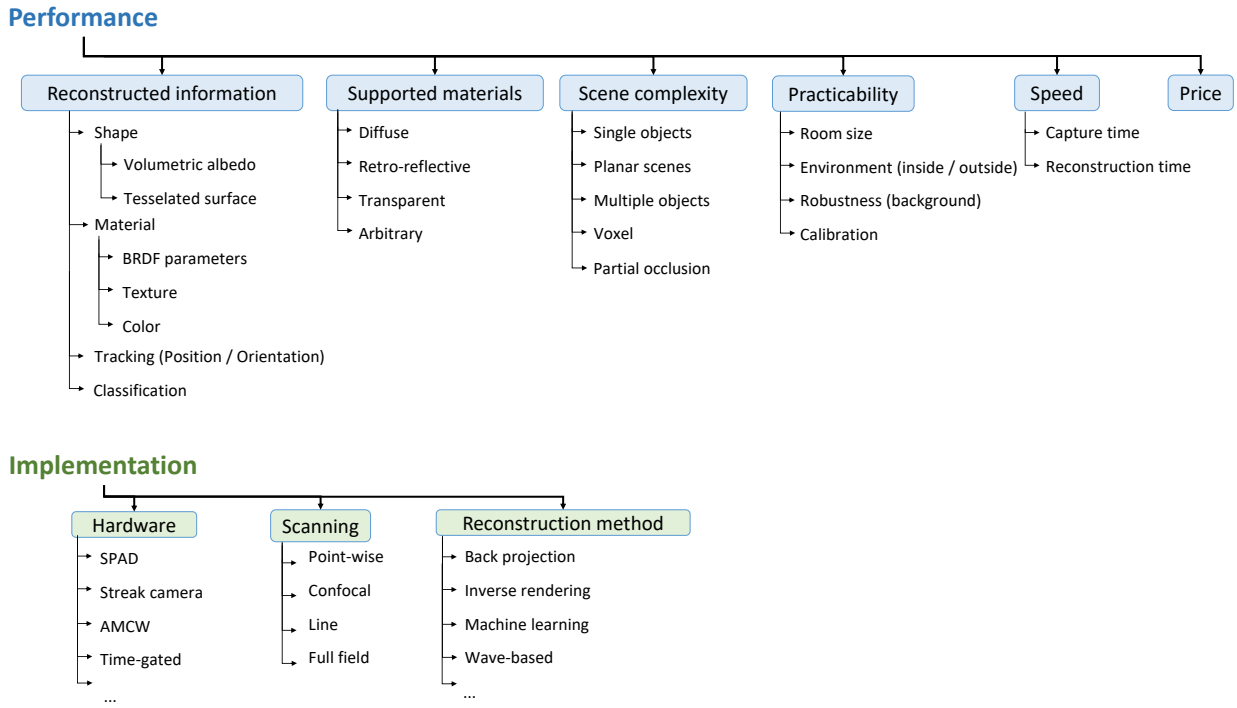


Figure 3.1: Different dimensions to categorize NLoS imaging Systems.

from ambient light. Setups also often have to be calibrated towards specific scenes, making them less portable.

Speed High measurement speed allows capturing dynamic scenes while high reconstruction speed allows real-time observation.

Price If off-the-shelf components are used, setups are more suitable for consumer products.

Hardware Various hardware platforms for measuring transient images are available (see Section 2.2.1). They can greatly differ in spatial and temporal resolution and impact speed and price of the setup.

Scanning The general model in Section 2.3 describes measurements as sets of camera and laser points. These can be measured in different ways, where scan line or full field measurements are more time efficient than single point scanning (often using galvanometers). However, the later is a requirement for confocal measurements.

Reconstruction method The reconstruction algorithm presents the underlying idea of an setup and is usually the main contribution of a new publication.

3.1 Historical foundation

The idea of NLoS imaging using transient imaging was first introduced by Kirmani et al. in 2009 [Kir+09]. They estimate pair-wise distances of individual patches that scatter light onto each other from transient measurements. In an extension, some of these patches can be hidden from both light source and detector, which renders it the first NLoS problem.

Pandharkar et al. first introduced the 3-bounce setup as shown in Figure 2.3a where there is a clear distinction between the visible side in which the sensing setup is located and, separated by an occluding wall, a hidden side in which the object of interest resides [Pan+11].

They demonstrate tracking of a moving object in a cluttered environment (which is filtered out by taking difference images) and an estimation of the object size from the focus width of the signal.

Naik et al. used a NLoS setup with known geometry to estimate the material of a hidden object by fitting a three-dimensional BRDF model to streak camera measurements [Nai+11]. The algorithm can handle multiple patches with different materials as well.

In 2012, Velten et al. published the first method to reconstruct the full geometry of a hidden object [Vel+12]. Using the 3-bounce setup the measurements of a streak camera are back-projected into a voxel volume. By applying spatial filtering, the geometry is revealed. Even though the setup is slow and expensive, the amount of reconstructed details was remarkable and likely sparked much of the later research.

These early approaches established the idea of NLoS imaging as an active research area and also already cover the most common types of information to be reconstructed: Object shape, position, and material. With the exception of Kirmani et al. (who uses an oscilloscope), all these approaches use streak cameras which offer high temporal resolution but are in general expensive and slow.

Capture hardware variety

A variety of hardware platforms for transient images exists (see Section 2.2.1) on which NLoS imaging was subsequently ported and evaluated.

Laurenzis et al. demonstrate that a range-gated imaging system can be used in place of a streak camera [LV14]. The physical setup and reconstruction algorithm largely follows Velten et al., while introducing a new filtering method for the post-processing step.

Heide et al. use a setup based on AMCW Lidar [Hei+14]. Since it can not measure full transient histograms, the reconstruction is performed by solving a linear equation system. The hidden scene is modeled as a height field and multiple measurements with different modulation frequencies are taken. Exploiting the linearity of light propagation and some sparsity priors, the hidden scene can be reconstructed.

Buttafava et al. were the first to utilize a single pixel SPAD camera [But+15]. It is focused on the middle of the reflector wall while the laser scans an array of positions around it. Back projection is used as the reconstruction algorithm and an evaluation of the influence of the hidden object's albedo is performed, which confirms the intuitive assumption that lighter objects reflect more signal and are beneficial for the reconstruction.

The first method for NLoS reconstruction that does not rely on temporal resolution was presented by Katz et al. (see Section 3.4), however, it uses a different scene setup [Kat+14]. On the 3-bounce setup, a method presented in our own work (see Chapter 4) was the first to rely solely on intensity data. It runs in real time and on cheap hardware (as it does not use transient imaging) but performs only position and orientation tracking rather than full three-dimensional reconstruction [Kle+16]. It also introduced inverse rendering as novel reconstruction method.

Confocal setups

Although in previous work a variety of patterns for laser points and camera points are used, they share the characteristic of illuminating and measuring different points. O'Toole et al.

were the first to introduce the so called *confocal* setup, in which for each measurement the laser and camera points are always the same [OLW18]. The camera and laser view directions are combined using a beam splitter and with a galvanometer various positions at the reflector wall are scanned. This justifies also the term *coaxial* setup which focuses on the fact that both paths are aligned, however *confocal* remains the more popular description in later literature.

Having a laser and camera point coincide simplifies the image formation model, as the ellipses in Figure 2.7a become circles which turns the forward operator into a convolution. This allows to use a deconvolution-based reconstruction which is both much faster and more robust compared to traditional back projection as it offers a closed-form solution. It imposes however some new constraints such as requiring planar reflector wall and the inability to account non-retroreflective components of the light.

Heide et al. extends the image formation model of the confocal setup by adding occlusion (modeled by a visibility term) and surface normals [Hei+19]. The reconstruction is then done by solving a multi-convex optimization problem which is computationally more expensive than other approaches but allows to reconstruct more complex scenes.

Confocal setups are difficult to measure, since they contain a strong primary reflection (where the laser spot on the wall is in the field of view of the camera pixel). This leads to overexposure and the main signal (which is orders of magnitude weaker) is easily lost. A common approach is to defocus both points slightly to avoid capturing the primary reflection. This violates model assumptions, but in practice the impact of it is neglectible.

3.2 Reconstruction methods

Back projection

Back projection algorithms have been widely investigated in the context of computer tomography [KS88, Chap. 8] and were first introduced to NLoS imaging by Pandharkar et al. [Pan+11]. Even though better algorithms for geometry reconstruction are available today, it remains an easy to implement and robust algorithm for object tracking under the assumption that the scene mostly consists of a single, point-shaped object (see Section 2.3.2). As such, it is used in a number of publications that focus on extending the basic idea of NLoS imaging in various ways: Laurenzis et al. introduced the first setup using time-gated hardware [LV14], Gariepy et al. demonstrate person tracking from reflections off the floor rather than the wall [Gar+16], and Chan et al. demonstrates long distance tracking using a telescopic lens [Cha+17b].

Back projection for full geometry reconstruction was first demonstrated by Velten et al. [VRB11]. O. Gupta et al. experimented with a CoSaMP-based reconstruction [NT09] although notable improvements are only reported on synthetic data [Gup+12]. Buttafava et al. demonstrate that SPAD measurements are suitable input for back projection, and Arellano et al. developed an efficient, GPU-based back projection implementation [AGJ17].

La Manna et al. embed back projection into an iterative scheme [La +18]. In each iteration, a forward rendering operator is used to compare the current back projection result to the measurements and the difference is propagated into the reconstruction using either an additive or multiplicative mode. Upon convergence, the result is consistent with the forward

operator. For linear light transport operators (which is only true under strong assumptions, e.g. no occlusion), this method is an instance of *algebraic reconstruction techniques* [GBH70] and statements about convergence exist. The authors report improved results on synthetic data, on measured data the results are however less convincing, which the authors note might be due to shortcomings in the forward model.

Ahn et al. derive formal justification for the filtered back projection approach by showing that under certain assumptions on the imaging geometry, the Gram of the NLoS measurement operator is a convolution operator [Ahn+19]. With their mathematical framework, the authors then derive an optimized deconvolution kernel that offers improved reconstruction quality.

Conceptually similar to back projection is the space carving algorithm described by Tsai et al. [Tsa+17]. Here, only the first returning photons are used to identify areas in the reconstruction volume that are guaranteed to be free of geometry in contrast to marking potentially occupied areas. This approach is potentially more robust, since it is independent from any amplitudes, however, recovering details of regions with negative curvature or partially occluded parts of the scene is not possible.

Inverse rendering

In *inverse rendering*, a traditional (forward) rendering algorithm is used inside a numerical optimization loop to solve the inverse problem: retrieving scene parameters from rendered images (see Figure 4.1). While solving a single inverse problem is computationally expensive as it requires many evaluations of the forward model, it is still a widely used method since it can solve problems for which there is no analytical solution to the inverse. In NLoS imaging, an initial guess of the scene parameters can be refined by evaluating the forward model (which is usually derived from classic computer graphic research) and comparing it to the real scenes measurement. The difference between the two is measured by an objective function which then is minimized by algorithms like gradient descent. As appropriate forward models already exist or can be derived from existing ones, inverse rendering is a versatile tool and readily adopted to many reconstruction goals.

Naik et al. were the first to employ inverse rendering to NLoS imaging by fitting low-dimensional BRDF models to hidden surfaces [Nai+11].

In our own work (see Chapter 4), we develop a specialized, patch-based forward renderer fast enough to perform real-time tracking of hidden objects [Kle+16]. The algorithm can also reconstruct the orientation and even be used to classify objects of various shapes, but some ground-truth knowledge of the hidden objects is required.

Iseringhausen et al. demonstrate high-dimensional geometry reconstruction using inverse rendering [IH20]. They use a similar but extended renderer that takes occlusion into account in combination with a geometry model based on isosurfaces of three-dimensional Gaussian kernels. While a single reconstruction can take over a day, the method is shown to be significantly more robust to noise than many others.

Tsai et al. reconstruct even more surface detail, but require an initialization that is already close to the real geometry, due to the high non-convexity of the problem [TSG19]. They suggest to retrieve this initialization using a different method (and demonstrate it using the space carving algorithm of Tsai et al. [Tsa+17]) and proceed to optimize the position of

individual triangle vertices on the mesh. Although slow as well, this method generates by far the highest amount of surface details demonstrated so far.

In our own work (see Chapter 6) we tackle the problem of calibrating the visible part of the setup (which is a prerequisite for many methods) by presenting a calibration scheme based on specular reflections from multiple mirrors in the scene [Kle+20]. The specular reflections drastically simplify the image formation model, which enables automatic differentiation and an calibration result ready in a matter of minutes even with unoptimized code.

Machine learning

Some publications explore the feasibility of machine learning-based reconstructions, mostly in the form of neural networks. Given enough examples, a network of suitable architecture can be trained to automatically find a mapping between NLoS measurements and various kinds of output parameters, from low-dimensional position tracking to high-dimensional geometry reconstruction. There is no need to explicitly model light transport or camera noise characteristics or to calibrate the visible part of the setup, as all of these can be learned by the network. However, an explicit forward model is still necessary for the generation of synthetic training data since insufficient training data leads to over fitting and a lack of generalization to new environments.

Caramazza et al. were the first to use a neural network to distinguish different persons and their position (out of a small set of discrete possibilities) in a hidden volume [Car+18]. Although the output is low-dimensional and the network is only trained on a single setup geometry, this proved that machine learning is suitable for NLoS imaging problems.

Chen et al. reconstruct a two-dimensional view of the hidden scene from non-transient measurements with a neural network [Che+19]. Using only intensity data is a good example for the power of machine learning, since reconstructing this high amount of information without the temporal dimension has not been demonstrated before.

The work of Chopite et al. goes into a similar direction, although they use transient signals and reconstruct a depth map of the scene, rather than an albedo map [Cho+20].

So far, machine learning-based approaches have not been demonstrated to work on arbitrary new setups. Explicit reconstruction methods on the contrary can be quite easily calibrated to new setups and their reconstruction quality is in some sense depending on the calibration quality. However, in machine learning-based approaches, small derivations from the original setup can have unforeseeable effects which necessitates very thorough evaluation of such systems.

Wave-based

While back projection can be thought of as utilizing ray optics, similar algorithms can be formulated using wave-equations. This is a well-established approach in the field of seismology where acoustic waves are used to measure the reflectivity of the earth surface [Sto78]. The underlying mathematical formulation is similar to that of NLoS imaging.

Reza et al. were the first to adopt this approach [Rez+19]. Each point on the reflector plane is modeled as a complex phasor, computed by a Huygens's-like integral over the incoming light. These phasors are virtual, meaning that their frequency is of the same magnitude

as the hidden scene (cm or m) and not on the scale of the actual wavelength of the light (nm or µm).

Liu et al. present a geometry reconstruction algorithm based on this theoretical framework [Liu+19]. Compared to classical back projection, it reveals significantly more details and is more robust to noise while having the same computational complexity.

Lindell et al. adapt a wave-based reconstruction method known as f-k migration from the field of seismology [LWO19]. If offers a closed-form solution and is much more efficient than filtered back projection. In the original formulation it requires a confocal scanning setup, however the authors also present an extension that converts non-confocal measurements to confocal measurements based on *normal moveout correction* [Yil01], which originates also in the seismology community. Although the pre-processing is approximate, results are convincing in practice.

Overall, wave-based NLoS models are amongst the most promising reconstruction methods presented so far and offer both, high quality and reconstruction speed.

Other

Other reconstruction methods that do not quite fit in any of the above categories have been proposed as well.

Heide et al. models the hidden scene as a diffuse height field without occlusion [Hei+14]. With these assumptions, the light transport can be modeled as a linear equation system that can be solved using the Alternate Direction Method of Multipliers (ADMM) and sparsity priors.

Kadambi et al. draws inspiration from established antenna signal processing methods [Kad+16]. The reflector wall is viewed as a virtual antenna array (where each observed point is an antenna location that performs an omnidirectional measurement of the incoming light) and radio source localization algorithms are used for the reconstruction. Based on these, an analysis of recoverability is performed as well.

Pediredla et al. models the hidden room as a set of planes [Ped+17]. Since there is no analytical solution for the signal reflected by a plane, a dictionary of possible plane candidates is precomputed using Monte Carlo sampling. With this, the reconstruction becomes a combinatorical problem of selecting the best fitting planes that make up the room. The approach might be extended to a know set of hidden objects (such as for classification task), but the discrete nature of the dictionary makes it unsuitable for general, continuous scenes.

As an extension of the space carving algorithm of Tsai et al.[TSG19] (see discussion of back projection), Xin et al. [Xin+19] present a reconstruction technique called Fermat flow that is based on the Fermat principle [Sta72]. Their key observation is that points in the hidden scene that are either on the boundary of an object or result in a specular reflection for a certain camera and laser point result in discontinuities in the transient image and can be reconstructed using purely geometric reasoning. The first returning photons of the work of Tsai et al. are a subset of these so-called Fermat points. Since Fermat flow does not make use of intensity information, it can reconstruct objects with arbitrary BRDFs.

Scheiner et al. [Sch+20] use a Doppler radar with a frequency of 76 GHz - 81 GHz (corresponding to a wavelength of about 5 mm). Since diffuse reflection occurs from surface details in the range of the wavelength, this turns many everyday objects such as buildings, cars, and

guard rails into specular reflectors. By measuring the geometry of the reflector separately, the specular reflections can be explicitly computed and the sensing of the hidden object is effectively transformed into a line-of-sight imaging problem. Due to the Doppler radar, only moving objects can be captured. In their work they also perform a machine learning based refinement step one the object reconstructed by standard radar imaging techniques to perform a classification and predict trajectories.

3.3 Miscellaneous extensions

Although the main focus of most publications lies on the development of improved reconstruction methods, some work has also been done to transfer NLoS imaging from lab setups to more realistic scenarios.

As the wall of the hidden room might be cluttered or otherwise unsuitable for reflection, Gariepy et al. propose to reflect the signal off the floor instead [Gar+16]. La Manna et al. propose to use a second SPAD array to measure the movement of a curtain that acts as reflector in real time [La +20]. This approach allows to use almost arbitrary objects with complex and dynamic shape as reflectors.

Since NLoS is of potential interest for remote sensing applications, Chan et al. demonstrate a large distance sensing setup, where a hidden scene is observed from a stand-off distance of over 50 m using a telescope [Cha+17b].

Similarly, Metzler et al. focus a laser and a SPAD imager through a keyhole over a potentially long range which imposes the challenge of having just a single confocal observation point on the reflector wall available [MLW19]. For moving objects, the shape and trajectory can be jointly retrieved using an expectation maximization algorithm.

The hidden scene model commonly consist of some objects of interest plus a background signal from the room geometry. The later is usually regarded as a source of noise and filtered out or ignored. Pediredla et al. explicitly models this background signal as a set of mathematical planes and attempts to reconstruct them [Ped+17].

Usually, a single wavelength is used for the measurements, leading to monochrome reconstructions. By using multiple wavelengths, color information can be reconstructed as demonstrated by Musarra et al., and Chen et al. [Mus+19; Che+19].

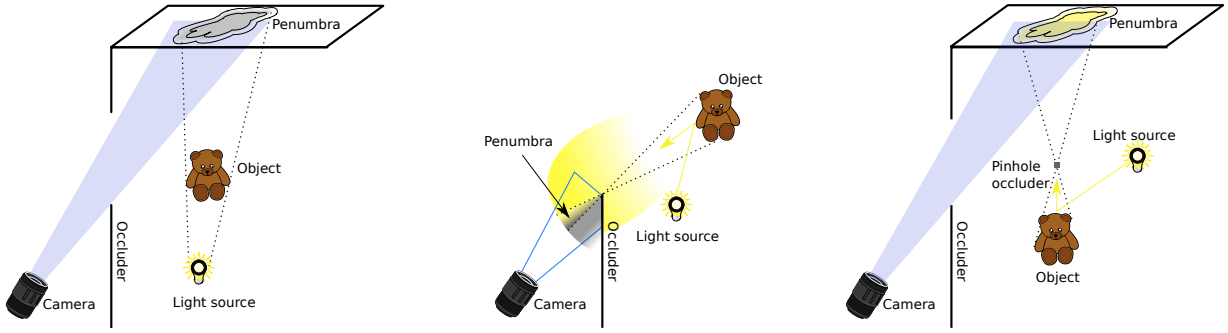
Lindell et al. [LWK19] use acoustic waves for NLoS reconstructions. Since the wave equation is the same for electromagnetic and acoustic waves, the confocal light cone transformation [OLW18] can be used for reconstruction after some domain specific adaptations.

3.4 Related problems

As outlined in the taxonomy in Figure 2.1, there are some problems closely related to transient NLoS imaging.

Partial occlusion

Instead of relying on the classic 3-bounce setup (shown in Figure 2.3a), some work attempts NLoS reconstruction from partial occlusion. These setups typically use *accidental illumina-*



(a) Object shadow at visible wall (b) Angle-dependent illumination of visible floor (c) Inverse pinhole

Figure 3.2: Different setups for occlusion based NLoS imaging. (a) A shadow of the hidden object is projected onto the visible wall. (b) Light reflected by the hidden object illuminates the visible floor. The intensity varies with the degree of occlusion by the wall. (c) Light from the hidden object reaches the wall unhindered, except when it comes from a single direction. In a traditional pinhole, only light from a certain direction would reach the wall.

tion rather than *active illumination*, meaning that a light source already present in the scene can be utilized rather than requiring a precisely controlled light source that is part of the measurement system.

Figure 3.2 shows an overview of occlusion-based setups. Baradad et al. observe an object and its shadow to reconstruct a four-dimensional light field of the hidden scene [Bar+18]. Later, Yedidia et al. extend this approach and do not rely on the knowledge of the shape anymore [Yed+19] (Figure 3.2a).

Bouman et al. analyze light reflected from an object onto the floor close to a corner (Figure 3.2b) [Bou+17]. The floor then encodes for each angle a one-dimensional projection of the object. Contributions from other parts of the scene can be filtered out if the hidden objects are moving and the static part of the illumination is gathered and subtracted from reconstruction measurements. This approach was extended by Seidel et al. who use a floor with non-homogeneous patterns (like tiles or stripes) which allows to also reconstruct static objects [Sei+19].

Saunders et al. use a small occluder within the hidden scene that projects partial shadows onto the reflector wall (Figure 3.2c) [SMG19]. The occluder realizes the function of an inverse pinhole, where instead of filtering light from a single direction all light except from a single direction passes. Similarly to pinhole cameras, smaller occluders lead to a higher resolution but a reduced contrast on the wall.

Since these setups are conceptually different from the classic 3-bounce setup, it is hard to compare transient imaging based approaches to occlusion based ones. However Thrampoulidis et al. use 3-bounce setup that also includes occluders and perform a study of how much can be reconstructed from either occlusion or temporal information [Thr+18].

Speckle pattern

When coherent light is reflected from a diffuse object, high-frequent speckle patterns are formed. The rough surface causes phase shifts in the scale of the wavelength which mod-

ulate the intensity by constructive and destructive interference. The speckle pattern is quasi-random, since the multitude of small surface variations cannot be modeled in practice. However, these patterns still encode information about the incoming light before the reflection, which can be decoded in various ways to perform NLoS reconstruction:

Feng et al. and Freund et al. first described and verified the so called *memory effect*, which describes that information about large-scale spatial variation is preserved while finer variations are lost when a wave travels through a scattering medium [Fen+88; FR88]. Katz et al. build on this principle to reconstruct a two-dimensional image around a corner [Kat+14].

Even if the hidden scene is illuminated by incoherent light, speckle-based approaches can be used by utilizing a spatial light modulator for wavefront shaping [KSS12].

Due to their high frequencies, speckle patterns are very sensitive to small motions which can be used for precise tracking of small objects, such as finger gestures [Smi+17]. If these patterns are reflected onto a wall in a NLoS setup, the same technique can be used for tracking around a corner [SOG18].

Speckle patterns are also used in lens-less imaging. Since no electronic circuit is fast enough to measure the phase of incoming visible light directly (only the intensity of the wave can be captured), a large part of the information is lost. *Phase retrieval* algorithms can be used to first reconstruct the phase and subsequently the full image [She+15; JEH15]. In a NLoS setup, the reflector wall then acts as a lens-less imager whose signal is recorded by the actual camera.

Lately, machine learning-based techniques have been applied to speckle patterns. Finding a matching decoder function to the quasi-random encoded speckle image is well suited for data-driven algorithms and can be used to classify digits or human poses [Lei+19]. Metzler et al. derive a noise model from spectral estimation theory and use it in combination with a neural network for robust phase retrieval from noisy data [Met+20]. This allows robust retrieval of two-dimensional images from hidden scenes.

Speckle imaging can also be combined with transient reconstruction techniques. Boger-Lombard and Katz reconstruct travel times from speckle patterns and proceed with a NLoS reconstruction using standard back projection [BK19].

Looking through occluders

Looking through objects is not only a problem closely related in application to looking around objects, but also shares some of the technologies and algorithmic methods with it. As outlined in Figure 2.1, the various methods can be categorized by the type of occluder they can look through.

For turbid media such as fog, muddy water, or rain and snow, small particles reduce the transparency in the medium by scattering or absorption, which imposes a severe problem for applications such as driving in bad weather or underwater imaging. Usually the particles are regarded as random, and unknown and are modeled by distribution functions. (A rare counter example is found in Iseringhausen et al., where rain drops on a transparent surface are individually estimated and their geometry is used to reconstruct a light field of the scene behind it by ray tracing.)

Since some photons still arrive at the camera without any (or very little) particle interaction, reconstructing the scene becomes a problem of filtering out the direct signal from

the noise [Wan+91; Bus05; Lau+12]. For temporally resolved measurements (such as in range-gated imaging), the direct signal will always have a shorter travel time than light with the same origin that was reflected by particles. However, light reflected by particles in front of the object might have the same path length as light coming straight from the object; thus the amount of noise that can be filtered out by a time gate has an upper limit.

Bijelic et al. solve the problem of imaging through bad weather by multisensor fusion [Bij+20]. Measurements from a stereo camera, a gated camera, a radar, a lidar, and a FIR camera are fed into a neural network which is trained to use the most reliable data in each situation to extract scene features.

For higher particle density, the medium turns into a diffusor and essentially no photon can travel through it without particle interaction. This problem is closely related to imaging around occluders using diffuse reflectors and various proposed solutions are demonstrated on both [KSS12; Kat+14; Kad+16; Xin+19]. Imaging through diffusors mostly relies on transient imaging [Sat+17; LW20] or speckle imaging [Ber+12].

Han et al. evaluate how the frequency of the electromagnetic waves influence the contrast that is achieved in sensing through diffusors by comparing near-infrared and terahertz waves [HCZ00]. As an extension of this idea it is possible to detect objects even through completely opaque (in the visible range) objects by using a wavelength that interacts less with the occluder. WiFi radiation (2.4 GHz - 5 GHz) passes through many types of walls and other obstacles relatively unhindered. However, since many room-sized objects are smooth with respect to the wavelength (where 5 GHz corresponds to a wavelength of about 6 cm), reflections are predominantly specular and receiving the reflected signal becomes a challenge (an effect popularly known from stealth bomber cloaking in the radar domain [McC08]). Karanam et al. use two drones to measure the signal of a scene hidden in a ring of cinder blocks [KM17]. By flying around the scene, the specular reflections can be captured. Adib et al. track humans through walls from a single location [AK13; Adi+15]. At different times during the movement, specular reflections from different body parts show up and can be merged in a post-processing step.

CHAPTER 4

Tracking objects outside the line of sight using 2D intensity images

This chapter was published as a peer-reviewed paper in the *Scientific Reports* journal by the *nature publishing group* in 2016 [Kle+16].
The authors are Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B. Hullin.

The observation of objects located in inaccessible regions is a recurring challenge in a wide variety of important applications. Recent work has shown that using rare and expensive optical setups, indirect diffuse light reflections can be used to reconstruct objects and two-dimensional (2D) patterns around a corner. Here we show that occluded objects can be tracked in real time using much simpler means, namely a standard 2D camera and a laser pointer. Our method fundamentally differs from previous solutions by approaching the problem in an analysis-by-synthesis sense. By repeatedly simulating light transport through the scene, we determine the set of object parameters that most closely fits the measured intensity distribution. We experimentally demonstrate that this approach is capable of following the translation of unknown objects, and translation and orientation of a known object, in real time.

4.1 Introduction

The widespread availability of digital image sensors, along with advanced computational methods, has spawned new imaging techniques that enable seemingly impossible tasks. A particularly fascinating result is the use of ultrafast time-of-flight measurements [Abr78; VRB11] to image objects outside the direct line of sight [Vel+12; Hei+14; LV14; Gar+16]. Being able to use arbitrary walls as though they were mirrors can provide a critical advantage in many sensing scenarios with limited visibility, like endoscopic imaging, automotive safety, industrial inspection and search-and-rescue operations.

Out of the proposed techniques for imaging occluded objects, some require the object to be directly visible to a structured [Sen+05] or narrow-band [SEL11; KSS12; Kat+14]

light source. Others resort to alternative regions in the electromagnetic spectrum where the occluder is transparent [Sum+11; AK13; Adi+15]. We adopt the significantly more challenging assumption that the object is in the direct line of sight of neither light source nor camera (Figure 4.1), and that it can only be illuminated or observed indirectly via a diffuse wall [Vel+12; Hei+14; LV14; But+15; Gar+16]. All the observed light has undergone at least three diffuse reflections (wall, object, wall), and reconstructing the unknown object is an ill-posed inverse problem. Most solution approaches reported so far use a back projection scheme as in computed tomography [PSV09], where each intensity measurement taken by the imager votes for a manifold of possible scattering locations. This *explicit* reconstruction scheme is computationally efficient, in principle real-time capable [Gar+16], and can be extended with problem-specific filters [Vel+12; Kad+16]. However, it assumes the availability of ultrafast time-resolved optical impulse responses, whose capture still constitutes a significant technical challenge. Techniques proposed in literature include direct temporal sampling based on holography [Abr78; Abr83; QM85], streak imagers [VRB11], gated image intensifiers [LV14], serial time-encoded amplified microscopy [GTJ09], single-photon avalanche diodes [Gar+15], and indirect computational approaches using multi-frequency lock-in measurements [Hei+13; Kad+13; Pet+15]. In contrast, *implicit* methods state the reconstruction task in terms of a problem-specific cost function that measures the agreement of a scene hypothesis with the observed data and additional model priors. The solution to the problem is defined as the function argument that minimizes the cost. In the only such method reported so far [Hei+14], the authors regularize a least-squares data term with a computationally expensive sparsity prior, which enables the reconstruction of unknown objects around a corner without the need for ultrafast light sources and detectors.

Here we introduce an implicit technique for detecting and tracking objects outside the line of sight in real time. Imaged using routinely available hardware (2D camera, laser pointer), the distribution of indirect light falling back onto the wall serves as our main source of information. This light has undergone multiple reflections; therefore, the observed intensity distribution is low in spatial detail. Our method combines a simulator for three-bounce indirect light transport with a reduced formulation of the reconstruction task [Gar+16; Kad+16]. Rather than aiming to reconstruct the geometry of an unknown object, we assume that the target object is rigid, and that its shape and material are either known and/or irrelevant. Translation and rotation, the only remaining degrees of freedom, can now be found by minimizing a least-squares energy functional, forcing the scene hypothesis into agreement with the captured intensity image.

Our main contributions are threefold. We propose to use light transport simulation to tackle an indirect vision task in an analysis-by-synthesis sense. Using synthetic measurements, we quantify the effect of object movement on the observed intensity distribution, and predict under which conditions the effect is significant enough to be detected. Finally, we demonstrate and evaluate a hardware implementation of a tracking system. Our insights are not limited to intensity-only imaging, and we believe that they will bring non-line-of-sight sensing closer to practical applications.

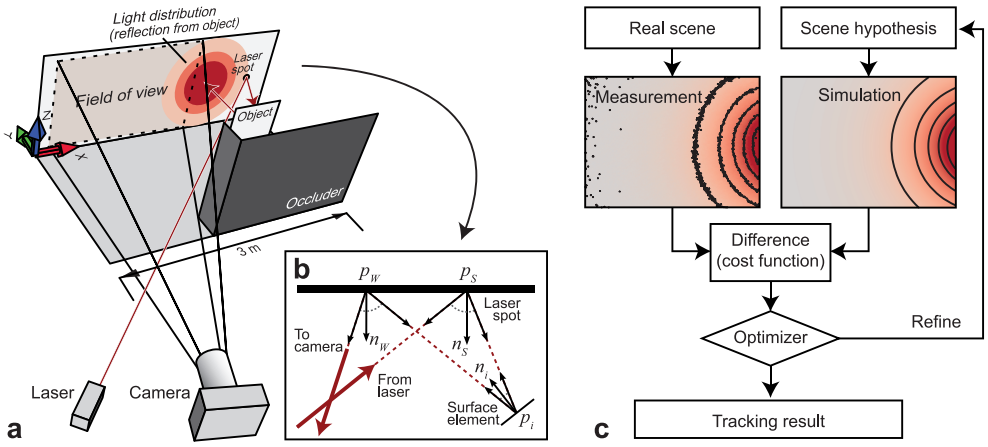


Figure 4.1: Tracking objects around a corner. **a**, Our experimental setup follows the most common arrangement reported in prior work, except that it does not use time-of-flight technology. A camera observes a portion of a white wall. To the right of the camera's field of view, a collimated laser illuminates a spot that reflects light toward the unknown object. The light distribution observed by the camera is the result of three diffuse light bounces (wall–object–wall) plus ambient contributions. **b**, Geometry of three-bounce reflection for a single surface element. **c**, Flow diagram of our tracking algorithm. Given shape, position and orientation of an object (the "scene hypothesis"), we simulate light transport to predict the distribution that this object would produce on the wall. By comparing this distribution to the one actually observed by the camera, and refining the parameters to minimize the difference, the object's motion is estimated.

4.2 Results

Light transport simulation (synthesis). At the center of this work is an efficient renderer for three-bounce light transport. Being able to simulate indirect illumination at an extremely fast rate is crucial to the overall system performance, since each object tracking step requires multiple simulation runs. Like all prior work, we assume that the wall is planar and known, and so is the position of the laser spot. The object is represented as a collection of Lambertian surface elements (*Surfels*), each characterized by its position, normal direction and area. As the object is moved or rotated, all its surfels undergo the same rigid transformation. We represent this transformation by the *scene parameter* p, which is a three-dimensional vector for pure translation, or a six-dimensional vector for translation and rotation. The irradiance received by a given camera pixel is computed by summing the light that reflects off the surfels. The individual contributions, in turn, are obtained independently of each other as detailed in the Methods section, by calculating the radiative transfer from the laser spot via a surfel to the location on the wall observed by a pixel. Note that by following this procedure, like all prior work, we neglect self-occlusion, occlusion of ambient light, and interreflections. To efficiently obtain a full-frame image, represented by the vector of pixel values $\mathbf{S}(p)$, we parallelized the simulation to compute each pixel in a separate thread on the graphics card. The rendering time is approximately linear in the number of pixels and the number of surfels. On an NVIDIA GeForce GTX 780 graphics card, the response from a moderately complex object (500 surfels) at a resolution of 160×128 pixels is rendered in 3.57 milliseconds.

To estimate the magnitude of changes in the intensity distribution that are caused by motion or a change in shape, we performed a numerical experiment using this simulation. In this experiment, we used a fronto-parallel view on a 2 m×2 m wall, with a small planar object (a 10 cm×10 cm white square) located at 50 cm from the wall. Object and laser spot were centered on the wall, but not rendered into the image. Figure 4.2 shows the simulated response thus obtained. By varying position and location of the object, we obtained difference images that can be interpreted as partial derivatives with respect to the components of the scene parameter p. Since the overall light throughput drops with the fourth power of the object-wall distance, translation in Y direction caused the strongest change. Translation in all directions and rotation about the X and Z axes affected the signal more strongly than the other variations. With differences amounting to several percent of the overall intensity, these changes were significant enough to be detected using a standard digital camera with 8- to 12-bit A/D converter.

Experimental setup. Our experiment draws inspiration from prior work [Vel+12; Hei+14; But+15; Gar+16; Kad+16]; the setup is sketched in Figure 4.1a. Here, due to practical constraints, some of the idealizing assumptions made during the synthetic experiment had to be relaxed. In particular, only an off-peak portion of the intensity pattern could be observed. To shield the camera from the laser spot and to avoid saturation and lens flare, we had to position the laser spot outside the field of view. The actual reflectance distribution of the wall and object surfaces was not perfectly Lambertian, and additional light emitters and reflectors, not accounted for by the simulation, were present in the scene. To obtain a

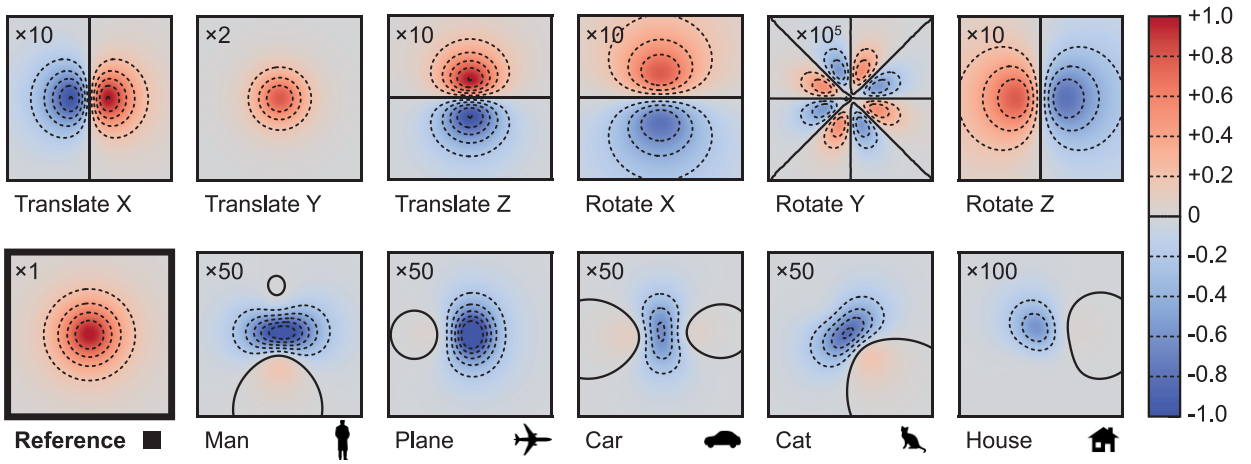


Figure 4.2: Intensity difference images. To investigate the effect of changes in object position and orientation on the intensity distribution observed on the wall, we performed a simplified synthetic experiment with an orthographic view of a 2 m×2 m wall, and laser spot and object centered with respect to the wall. The reference distribution (bottom left) was produced by a 10 cm×10 cm square-shaped object, located at 50 cm from the wall. Six difference images (top row), obtained by translating (±2.5 cm) and rotating (±7.5°) the object about the X, Y and Z axes, illustrate the distribution and magnitude of the respective change in the signal. The images shown in the bottom row visualize the difference caused by a change in shape. For display, each difference image has been amplified by the indicated factor (2 to 100,000) that also reflects the relative significance of the effect: Translations and rotations (except around the Y axis) caused the signal to change by roughly 1% per centimeter or per angular degree. A change in the object shape led to a peak difference around 1–2%, and rotation around the Y axis had a much smaller effect.

measured image $\mathbf{M}$ containing only light from the laser, we took the difference of images captured with and without laser illumination. Additionally, we subtracted a calibration measurement $\widehat{\mathbf{B}}$ containing light reflected by the background. A specification of the devices used, and a more detailed introduction of the data pre-processing steps, can be found in the Methods section.

Tracking algorithm (analysis). With the light transport simulation at hand, and given a measurement of light scattered from the object to the wall, we formulate the tracking task as a non-linear minimization problem. Suppose $\mathbf{M}$ and $\mathbf{S}(p)$ are vectors encoding the pixel values of the *measured* object term and the one predicted by the *simulation* under the transformation parameter or scene hypothesis p, respectively. We search for the parameter p that brings $\mathbf{M}$ and $\mathbf{S}(p)$ into the best possible agreement by minimizing the cost function

$$f(p) = \|\mathbf{M} - \gamma(\mathbf{M}, \mathbf{S}(p)) \cdot \mathbf{S}(p)\|_2^2, \qquad \text{where} \qquad \gamma(a,b) = \frac{a^T \cdot b}{\|b\|_2^2}. \tag{4.1}$$

The factor $\gamma(a,b)$ projects b to a, minimizing the distance $\|a - \gamma(a,b) \cdot b\|_2^2$. By including this factor into our objective, we decouple the recovery of the scene parameter p from any unknown global scaling between measurement and simulation, caused by parameters such as surface albedos, camera sensitivity and laser power. To solve this non-linear, non-convex, heavily over-determined problem, we use the Levenberg-Marquardt algorithm [Mar63] as implemented in the Ceres library [AMO15]. Derivatives are computed by numerical differentiation. When tracking six degrees of freedom (translation and rotation), evaluating the value and gradient of f requires a total of seven simulation runs, or on the order of 25 milliseconds of compute time on our system.

Tracking result. To evaluate the method, we performed a series of experiments that are analysed in Figure 4.4 and 4.5. The physical object used in all experiments was a car silhouette cut from plywood and coated with white wall paint, shown in Figure 4.3a. While our setup is able to handle arbitrary three-dimensional objects (as long as the convexity assumption is reasonable), this shape was two-dimensional for manufacturing and handling reasons.

For a given input image $\mathbf{M}$ and object shape, the cost function $f(p)$ in Equation 4.1 depends on three to six degrees of freedom that are being tracked. Figure 4.3b shows a slice of the function for translation in the XY-plane, with all other parameters fixed. Although the global minimum is located in an elongated, curved trough, only four to five iterations of the Levenberg-Marquardt algorithm are required for convergence from a random location in the tracking volume. In real-time applications, since position and rotation can be expected to change slowly over time, the optimization effort can be reduced to two to three iterations per frame by using the latest tracking result to initialize the solution for the next frame.

In **Experiment 1**, we kept the object's orientation constant. We manually placed the object at various known locations in an $60\,\text{cm} \times 50\,\text{cm} \times 60\,\text{cm}$ working volume, and recorded 100 camera frames at each location. These frames differ in the amount of ambient light (mains flicker) and in the photon noise. For each frame, we initialized the estimated position to a random starting point in a cube of dimensions $(30\,\text{cm})^3$ centered in the tracking volume, and refined the position estimate by minimizing the cost function, Equation 4.1. The

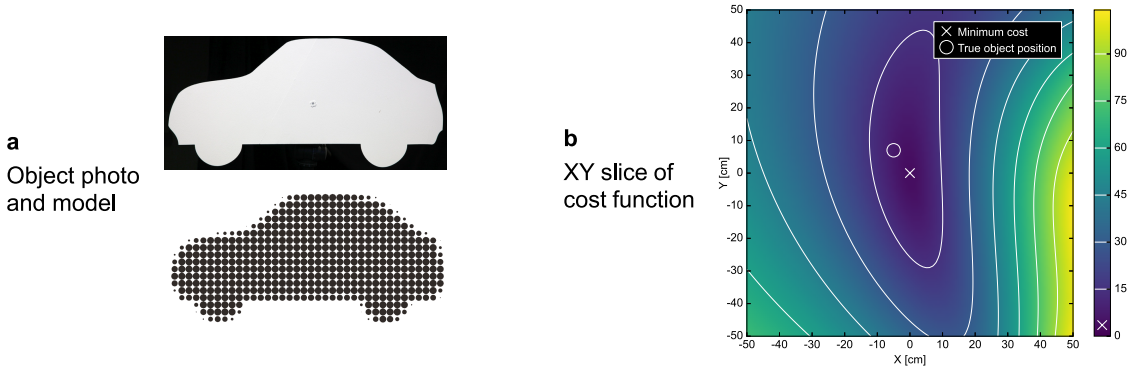


Figure 4.3: Object model and cost function used for tracking. **a,** photo of an object cut from white plywood, and its representation as surface elements (surfels). Note that although we use a flat object for demonstration, our method is also capable of handling three-dimensional objects. **b,** XY slice of the cost function for positional tracking, centered around the global minimum. With a perfect image formation model and in the absence of noise, the minimum (marked by cross) and the measured position of the object (marked by circle) should coincide at a function value of exactly $f(p) = 0$. Under real conditions, the reconstructed position deviated from the true one by a few centimeters, and the minimum was a small positive value.

results are shown in Figure 4.4a. From this experiment, we found positional tracking to be repeatable and robust to noise, with a sub-cm standard deviation for each position estimate. The root-mean-square distance to ground truth was measured at 4.8 cm, 2.9 cm and 2.4 cm for movement along the X, Y and Z axis, respectively. This small systematic bias was likely caused by a known shortcoming of the image formation model, which does not account for occlusion of ambient light by the object.

In **Experiment 2**, we kept the object at a (roughly) fixed location and rotated it by a range of $\pm 30°$ around the three coordinate axes using a pan-tilt-roll tripod with goniometers on all joints. Again, per setting we recorded 100 frames that mainly differ in the noise pattern. We followed the same procedure as in the first experiment, except that this time we jointly optimized for all six degrees of freedom (position and orientation). The results are shown in Figure 4.4(b). As expected, the rotation angles were tracked with higher uncertainty than the translational parameters, although the average reconstructions for each angle remain stable. We identify two main sources for the added uncertainty: the increased number of degrees of freedom and the pairwise ambiguity between X translation and Z rotation, and between Z translation and X rotation (Figure 4.2). We recall that in the synthetic experiment, the effect of Y rotation was vanishingly small; here, the system tracked rotation around the Y axis about as robustly as the other axes. This unexpectedly positive result was probably owed to the strongly asymmetric shape of the car object.

So far, we assumed that the object's shape was known. Since this requirement cannot always be met, we dropped it in **Experiment 3**. Using the data already captured using the car object for the first experiment, we performed the light transport simulation using a single oriented surface element instead of the detailed object model. Except for this simplification, we followed the exact same procedure as in Experiment 1 to track the now unknown object's position. The results are shown in Figure 4.5(a). Despite a systematic shift introduced by

a
Experiment 1:
Position tracking

b
Experiment 2:
Rotation tracking

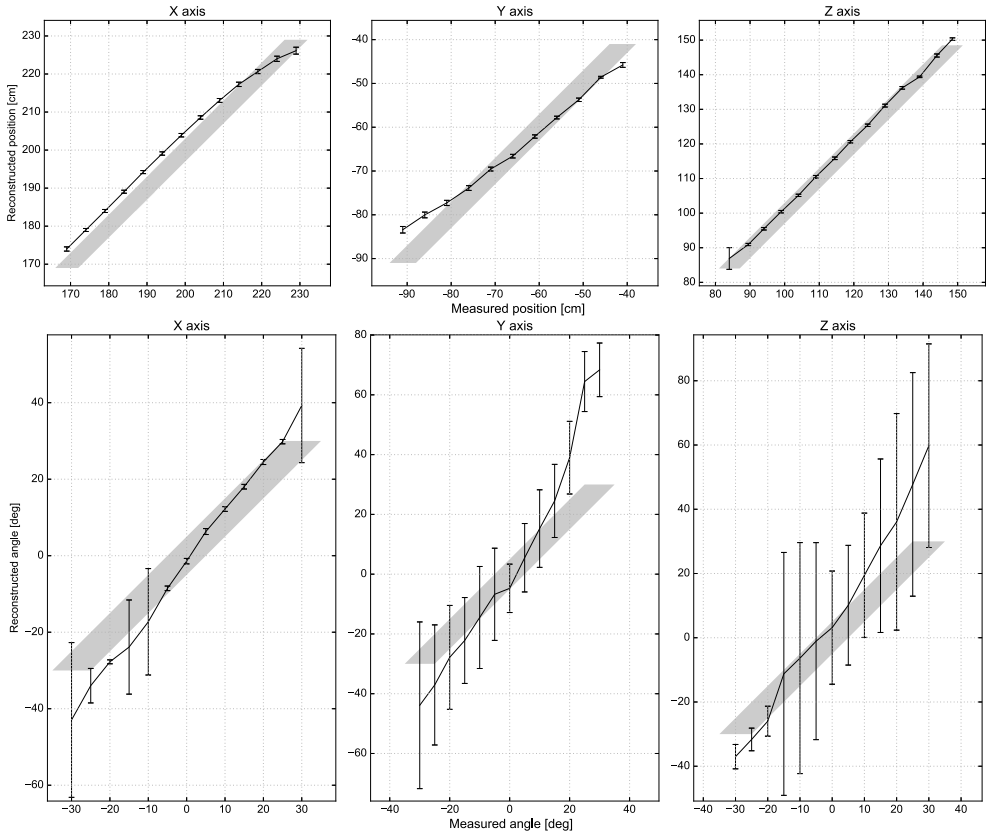


Figure 4.4: Tracking a known object. **a,** Result of three tracking sessions where the object was translated along the X, Y and Z axes (Experiment 1). We recorded 100 input images at each position and reconstructed the object position for each input image independently. Plots and error bars visualize the mean and standard deviation of the recovered positions. The area shaded in gray is the confidence range for the true position which was determined using a tape measure. **b,** Result of three tracking sessions where the object was rotated around the X, Y and Z axes (Experiment 2). From 100 input images, we jointly reconstructed translation and rotation. Shown are mean and standard deviation of the recovered rotation angle. The higher uncertainty reflects the fact that rotation in general has a smaller effect on the signal, and the ambiguity between translational and rotational motion (also see Figure 4.2).

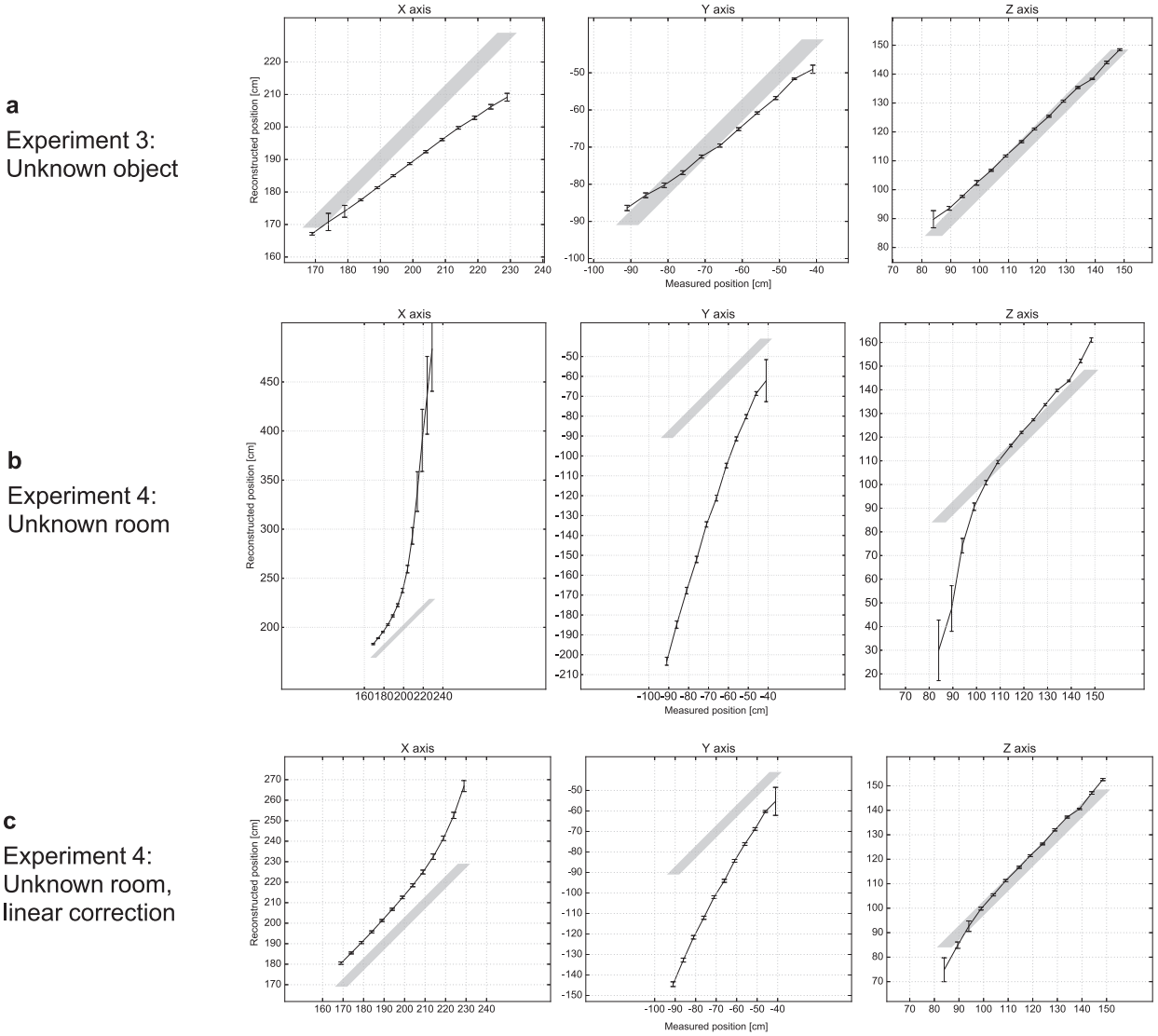


Figure 4.5: Tracking of an unknown object, or in an unknown room. **a,** Result of Experiment 3: Positional tracking as in Experiment 1, but with no knowledge about the object shape. We used a single oriented surface element for the light transport simulation. **b,** Result of Experiment 4: Positional tracking as in Experiment 1, but without subtracting the pre-calibrated room response. The estimated absolute position greatly deviated from the ground-truth position (shaded areas). **c,** Subtraction of a linear fit significantly reduced the tracking error and made the tracking task feasible even in the absence of a background measurement. In all cases, the standard deviation (error bars) remained small, indicating that changes in position could still be robustly detected.

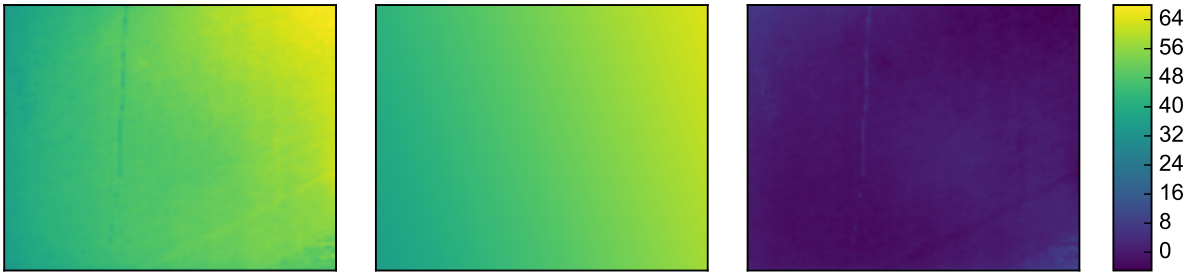


Figure 4.6: Approximating the background term by a linear model. From left to right, in arbitrary units: background term $\widehat{\mathbf{B}}$ obtained through calibration, linear approximation of $\widehat{\mathbf{B}}$, residual background term after subtraction of linear component.

the use of the simplified object model, the position recovery remained robust to noise and relative movement was still detected reliably.

The need for a measured background term $\widehat{\mathbf{B}}$ can hinder the practical applicability of our approach as pursued so far. In **Experiment 4**, we lifted this requirement. When omitting the term without any compensation, the tracking performance degraded significantly (Figure 4.5(b)). However, we observed that the background image, caused by distant scattering, was typically smooth and well approximated by a linear function $g(u, v) = au + bv + c$ in the image coordinates u and v (Figure 4.6). We extended the tracking algorithm to fit such linear models to both input images $\mathbf{M}$ and $\mathbf{S}(p)$, and subtract the linear portions prior to evaluating the cost function (Equation 4.1). This simple pre-processing step greatly reduced the bias in the tracking outcome and enabled robust tracking of object motion (Figure 4.5(c)) even in unknown rooms.

The supplementary video to this paper shows two real-time tracking sessions (Session 1: translation only; Session 2: translation and rotation) using the described setup. A live view of the hidden scene is shown next to the output from the tracking software. The average reconstruction rate during these tracking sessions was 10.2 frames per second (limited by the maximum capture rate of our camera-laser setup) for Session 1, and 3.7 frames per second (limited by computation) for Session 2. The two-dimensional car model was represented by 502 surfels; the total compute time required for a single tracking step was 72.9 ms for translation only, and 226.1 ms for translation and rotation.

4.3 Discussion

The central finding of this work is that the popular challenge of tracking an object around a corner can be tackled without the use of time-of-flight technology. By formulating an optimization problem based on a simplistic image formation model, we demonstrated parametric object tracking only using two-dimensional images with a laser pointer as the light source. In a room-sized scene, our technique achieves sub-cm repeatability, which puts it on par with the latest time-of-flight-based techniques [Kad+16; Gar+16]. However, as our technique does not rely on temporally resolved measurements of any kind, it has the unique property of being scalable to very small scenes (down to the diffraction limit) as well as large scenes (sufficient laser power provided). We note that the analysis-by-synthesis approach *per*

se is not limited to pure intensity imaging, but may form a valuable complement to other sensing modalities as well. For instance, a simple extension to the light transport model would enable it to accommodate time-of-flight or phase imaging.

A key feature of the analysis-by-synthesis paradigm is its transparency. Putting a virtual experiment (simulation) alongside the real experiment enables a rigorous quantitative analysis of the sensing problem. Using difference images, for instance, we investigated the influence that parameter changes have on the signal, and predicted the detectability of centimetre-scale motion. The same mechanism could also be used to obtain robustness estimates regarding additional unknowns in the scene model, such as non-diffuse object reflectance or the presence of additional objects. With these options, our approach offers a significant advantage over existing non-line-of-sight sensing techniques.

The real-time performance of our technique is determined by four main factors: the capture rate of the camera (constrained by exposure time and read-out bandwidth), performance of the compute system, the discretization of the model into surfels and the number of translational and rotational degrees of freedom afforded to the model. Other factors, in particular the question whether object and room are known, are irrelevant with this regard.

We identify four main limiting factors to the resolution and repeatability of our technique. Firstly, shortcomings in the models for scene and light transport can introduce a systematic bias. We exemplarily demonstrated how additional heuristic pre-processing steps can mitigate this bias. In usage scenarios where systematic errors preclude quantitative tracking, simpler sensing tasks, like the detection of object motion, will still remain possible. Secondly, the tracking of additional parameters like rotation, non-rigid objects or multiple object positions, is sensitive to image noise. The adoption of advanced filtering techniques or multi-frame averaging will further improve the tracking quality. Furthermore, certain applications will require a careful selection of the degrees of freedom afforded to the model. Thirdly, like in all prior work, we assumed knowledge about the geometry and angular reflectance distribution of a wall that receives light scattered by the unknown object. Thanks to recent progress in mobile mapping [Pue+13], highly detailed geometry and albedo texture data is already widely available for many application scenarios; if not, it can be recovered using existing line-of-sight sensing methods. Lastly, our tracking result is the outcome of a local parameter search (Levenberg-Marquardt) and hence not guaranteed to be the global optimum of the cost function, Equation 4.1. Although we never experienced convergence problems in practice, some situations may necessitate a combination of global and local optimization strategies.

The prospective of being able to sense beyond the direct line of sight can benefit many application fields. So far, the deployment of existing approaches has been hindered by practical limitations such as long capture times and device costs. As we were able to show here, these limitations can in principle be overcome if the problem can be reduced to a small number of degrees of freedom. One of the first applications of such reduced models could be in urban traffic safety, where the motion of vehicles and pedestrians is constrained to the ground plane. Extrapolating from our results, we believe that more detailed forward models and efficient simulation techniques can become a source of profound insight about non-line-of-sight sensing problems—and, eventually, enable the first truly practical solutions for looking around corners.

4.4 Methods

Light transport simulation. Accurate simulation of indirect illumination is computationally expensive and can take hours to complete. By assuming that all light has undergone exactly three reflections, we achieved a reduced overall computational complexity that is linear in the number of pixels and the number of surfels n. The geometry of this simulation is provided in Figure 4.1b. Each camera pixel observes a radiance value, L, leaving from a point on the wall, p_W, that, in turn, receives light reflected by the object's surfels. The portion contributed by the surfel of index $i \in \{1 \dots n\}$ is the product of three reflectance terms, one per reflection event; and the geometric view factors known from radiative transfer [ÇG14; Gor+84]:

$$\begin{aligned} L_i := \quad & \rho_0 \cdot f_s(p_L - p_S, p_i - p_S) && \text{(laser spot)} \\ & \cdot \tfrac{(n_S \circ (p_i - p_S)) \cdot (n_i \circ (p_S - p_i))}{\|p_S - p_i\|_2^2} \cdot f_i(p_S - p_i, p_W - p_i) \cdot A_i && (i^{\text{th}} \text{ surfel}) \\ & \cdot \tfrac{(n_i \circ (p_W - p_i)) \cdot (n_W \circ (p_i - p_W))}{\|p_i - p_W\|_2^2} \cdot f_W(p_i - p_W, p_C - p_W) && \text{(wall)}, \end{aligned} \tag{4.2}$$

where the operator

$$v \circ w := \begin{cases} \frac{v^T \cdot w}{\|v\|_2 \cdot \|w\|_2} & \text{if } v^T \cdot w > 0 \\ 0 & \text{otherwise} \end{cases}$$

denotes a normalized and clamped dot product as used in Lambert's cosine law. Each line in Equation 4.2 models one of the three surface interactions. n_S, n_i and n_W are the normal vectors of laser spot, surfel and observed point on the wall, and $f_{\{S,i,W\}}(\omega_{\text{in}}, \omega_{\text{out}})$ are the values of the corresponding bidirectional reflectance distribution functions (BRDF). The incident and outgoing direction vectors ω_{in} and ω_{out} that form the arguments to the BRDF are given by the scene geometry. In particular, the vectors p_L, p_S, p_i, p_W and p_C represent the positions of, in this order: the laser source, the laser spot on the wall, the i^{th} surfel, the observed point on the wall, and the camera (center of projection). A_i is the area of the i^{th} surfel, and ρ_0 a constant factor that subsumes laser power and the light efficiencies of lens and sensor. This factor is cancelled out by the projection performed in the cost function Equation 4.1, so we set it to $\rho_0 = 1$ in simulation. The total pixel value is simply computed by summing Equation 4.2 over all surfels:

$$L_{\text{total}} := \sum_{i=1}^{n} L_i \tag{4.3}$$

This summation neglects mutual shadowing or interreflection between surfels, an approximation that is justifiable for flat or mostly convex objects. For lack of measured material BRDFs, we further assume all surfaces to be of diffuse (Lambertian) reflectance such that $f_{\{S,i,W\}} := \text{const} = 1$, again making use of the fact that the cost function (Equation 4.1) is invariant under such global scaling factors. If available, more accurate BRDF models as well as object and wall textures can be included at a negligible computational cost.

Capture devices. Our image source was a Xenics Xeva-1.7-320 camera, sensitive in the near-infrared range (900 nm–1,700 nm), with a resolution of 320×256 pixels at 14 bits

per pixel. We used an exposure time of 20 ms. The laser source (1 W at 1.550 nm) was a fiber-coupled laser diode of type SemiNex 4PN-108 driven by an Analog Technologies ATLS4A201D laser diode driver and equipped with a USB interface trigger input. On the output side of the fiber, we fed the collimated beam through a narrow tube with absorbing walls to reduce stray light.

A desktop PC with an NVIDIA GeForce GTX 780 GPU, 32GB of RAM and an Intel Core i7-4930K CPU controlled the devices and performed the reconstruction.

Measurement routine and image pre-processing. After calibrating the camera's gain factors and fixed pattern noise using vendor tools, we assumed that all pixels had the same linear response. All images were downsampled to half the resolution (160×128 pixels) prior to further processing. Due to the diffuse reflections, apart from noise, the measurements do not contain any information of high spatial frequency. Thus, moderate downsampling is a safe way to improve the performance of the later reconstruction.

The images measured by the camera are composed of several contributions, each represented by a vector of pixel-wise contributions: *ambient* light not originating from the laser, $\mathbf{A}$; laser light scattered by static *background* objects present in the scene, $\mathbf{B}$; and laser light scattered by the dynamic *object*, $\mathbf{O}$. All measured images are further affected by noise, the main sources being photon counting noise and signal-independent read noise. We assume the scene to remain stationary at least during short time intervals between successive captures. Further assuming the spatial extent of the object to be small, shadowing of $\mathbf{A}$ and $\mathbf{B}$ by the object, as well as ambient light reflected by the object, can be neglected. By turning the laser on and off, and inserting and removing the object, the described kind of setup can capture the following combinations of these light contributions:

Laser off (0), object absent (0):	$\mathbf{I}_{00} = \mathbf{A} + \text{noise}$
Laser on (1), object absent (0):	$\mathbf{I}_{10} = \mathbf{A} + \mathbf{B} + \text{noise}$
Laser off (0), object present (1):	$\mathbf{I}_{01} = \mathbf{A} + \text{noise}$
Laser on (1), object present (1):	$\mathbf{I}_{11} = \mathbf{A} + \mathbf{B} + \mathbf{O} + \text{noise}$

The input image to the reconstruction algorithm, $\mathbf{M}$, was obtained as the difference of images captured in quick succession with and without laser illumination. Additionally, we subtracted a calibration measurement containing light reflected by the background:

$$\mathbf{M} := \mathbf{I}_{11} - \mathbf{I}_{01} - \widehat{\mathbf{B}} \approx \mathbf{O} + \text{noise}, \tag{4.4}$$

The addition or subtraction of two input images increases the noise magnitude by a factor of about $\sqrt{2}$. The background estimate $\widehat{\mathbf{B}}$ was captured with the object removed by recording difference images with and without laser illumination. We averaged $n = 300$ such difference images to reduce noise in the background estimate:

$$\widehat{\mathbf{B}} := \frac{1}{n} \sum_{i=0}^{n} \left(\mathbf{I}_{10}^{(i)} - \mathbf{I}_{00}^{(i)} \right) \overset{n \gg 1}{\approx} \mathbf{B} \tag{4.5}$$

CHAPTER 5

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

This chapter was published as a peer-reviewed paper at the *British Machine Vision Conference* in 2018 [Kle+18].
The authors are Jonathan Klein, Martin Laurenzis, Dominik L. Michels, and Matthias B. Hullin.

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.

5.1 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 [Kir+09; Vel+12].

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.

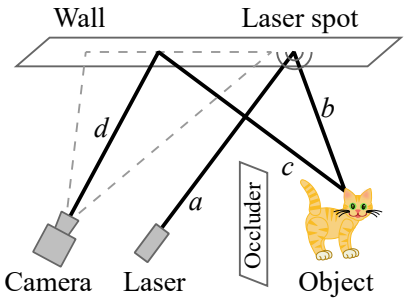


Figure 5.1: In the most common scenario of NLoS reconstruction, the path traveled by the light (laser-wall-object-wall-camera) consists of four segments a–d connected by three diffuse reflections.

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.

5.2 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 [Adi+15; AK13; FC99], or that exploit coherence properties of light, reconstructing objects using interferometry or speckle correlation [Kat+14] 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.

5.2.1 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 [Sen+05]. 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 [Kad+16] 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.

This leads to a canonical scene arrangement that has been prominently featured in most prior works [AGJ17; But+15; Car+17; Cha+17a; Gar+16; Hei+14; Kad+16; Kir+09; Kle+16; LV14; Ped+17; Shr+16; Vel+12; War+16; Tsa+17; Hei+18] and that is illustrated in Figure 5.1. 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 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 [Kad+16]. 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. [Vel+13].

5.2.2 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* $\boldsymbol{I}(u, v, \tau)$, where u and v are the usual image coordinates and τ is the travel time of light (Figure 5.2). We refer the reader to a recent survey on this general topic [Jar+17]. Among the devices used for the purpose are fast photodiodes [Kir+09], streak tubes [Vel+12], gated image intensifiers [LV14] and avalanche photodetectors [But+15; Cha+17a; Gar+16; Ped+17; Hei+18]. 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 [Hei+14; Kad+16; Shr+16] or even entirely unmodulated intensity images [Kle+16].

5.2.3 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 [Cha+17a; Kad+16; Kir+09; Kle+16; Shr+16] via characterization of room shapes [Ped+17] to the recovery of full three-dimensional shapes [But+15; Hei+14; Vel+12; Hei+18]. 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

[1] This also excludes settings like the one described by Jin et al. [Jin+14] that employ a pinhole to indirectly image the hidden scene.

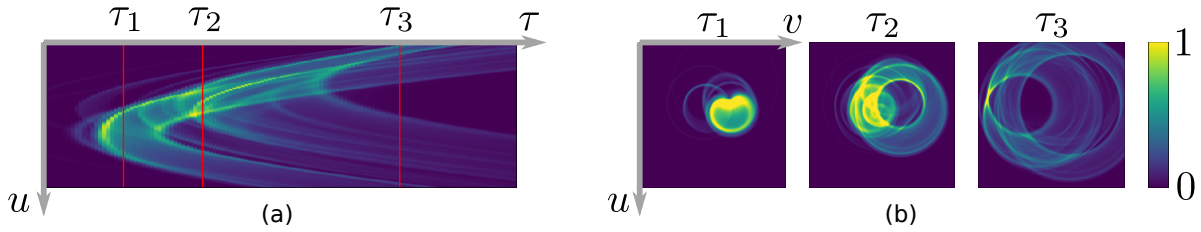


Figure 5.2: Slices of an unwarped transient image $\boldsymbol{I}(u, v, \tau)$ of light reflected by an object onto a wall as illustrated in Figure 5.1. u–τ slices (a) resemble streak images, whereas u–v slices (b) can be interpreted as frames of a video of light in flight. (Range normalized for display.)

forward model. For instance, researchers have proposed radiative transfer simulations based on oriented surface patches [Kir+09; Kle+16; Ped+17], derived a linearized light transport tensor [Hei+14; Hei+18], and exploited additional geometric constraints to express light transport as a convolution of light cones [OLW18]. The problem of reconstructing the scene $\boldsymbol{s}$ 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 back projection 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 [Vel+13].

We are not aware of any systematic investigation as to which of these approaches is best suited for a given reconstruction problem. For the back projection technique, La Manna et al. compared different flavors (additive/multiplicative back projection) as well as different iteration and filtering strategies [La +17].

5.3 Challenge design

On the highest level, the non-line-of-sight reconstruction problem addressed in our challenge is: given the transient image $\boldsymbol{I}(u, v, \tau)$, what is the scene $\boldsymbol{s}$? Here, $\boldsymbol{s}$ 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.

5.3.1 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 5.3. 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 5.5a.

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

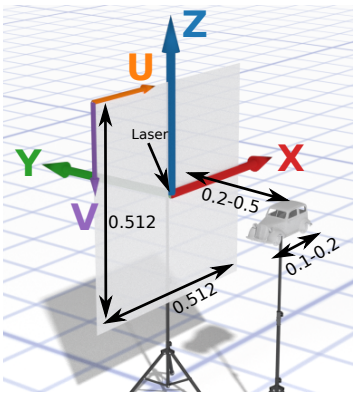


Figure 5.3: Our unified scene geometry.

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 [OLW18]) 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 [Wal+07]. Since material reconstruction is not part of the benchmark in this first iteration, all material parameters are provided.

5.3.2 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 5.3 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 [Kle+16] 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 [Kle+16] by integrating over the time dimension). An example of a transient image can be seen in Figure 5.2.

All our data are time unwarped (path segments a and d in Figure 5.1 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.

5.3.3 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 5.3. Thus synthetic renderings provide both high-quality transient images and high-quality ground truth data.

As rendering tool, we extended `pbrt-v3` [PJH16], 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.

5.3.4 Submission

The data sets 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 data sets, users can download an SDK with functions for data handling, error metrics and a base-line reconstruction algorithm, namely ellipsoidal back projection.

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.

5.4 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.

5.4.1 Materials

Our data set 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 [Wal+07]. 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.

5.4.2 Geometry reconstruction

The goal of this challenge is to reconstruct the object's geometry from a single transient image as illustrated in Figure 5.4a. 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 [RR96; SZL92] or volume [LT98]. A global comparison between two meshes can be achieved using an error metric based on the Hausdorff distance [KLS96]. 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 $t := (\boldsymbol{v}_0^t, \boldsymbol{v}_1^t, \boldsymbol{v}_2^t) \in \mathcal{T}$ with its corner vertices $\boldsymbol{v}_i^t$. Its center is given by $\boldsymbol{c}^t = (\boldsymbol{v}_0^t + \boldsymbol{v}_1^t + \boldsymbol{v}_2^t)/3$ and its area by

$A^t = \|(\boldsymbol{v}_1^t - \boldsymbol{v}_0^t) \times (v_2^t - \boldsymbol{v}_0^t)\|/2$ 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

$$d\left(\mathcal{M}_0, \mathcal{M}_1\right) = \sum_{t_0 \in \mathcal{M}_0} \frac{A^{t_0}}{A^{\mathcal{M}_0}} \min_{t_1 \in \mathcal{M}_1} \|\boldsymbol{c}^{t_0} - \boldsymbol{c}^{t_1}\| \tag{5.1}$$

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

$$D\left(\mathcal{M}_0, \mathcal{M}_1\right) = \max\left(d\left(\mathcal{M}_0, \mathcal{M}_1\right), d\left(\mathcal{M}_1, \mathcal{M}_0\right)\right). \tag{5.2}$$

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 $d\left(\mathcal{R}, \mathcal{G}\right)$ and $d\left(\mathcal{G}, \mathcal{R}\right)$ as they represent different quality indicators. For example, $d\left(\mathcal{R}, \mathcal{G}\right) = 0$ is reached if only a single point is reconstructed correctly while $d\left(\mathcal{G}, \mathcal{R}\right) = 0$ 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 $d\left(\mathcal{M}_0, \mathcal{M}_1\right)$. Likewise $\mathcal{M}_0 \subset \mathcal{M}_1$ is handled by $D\left(\mathcal{M}_0, \mathcal{M}_1\right)$. 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 [LC87] that is provided as part of the SDK. Contestants concerned about the triangulation quality are encouraged to use a different implementation.

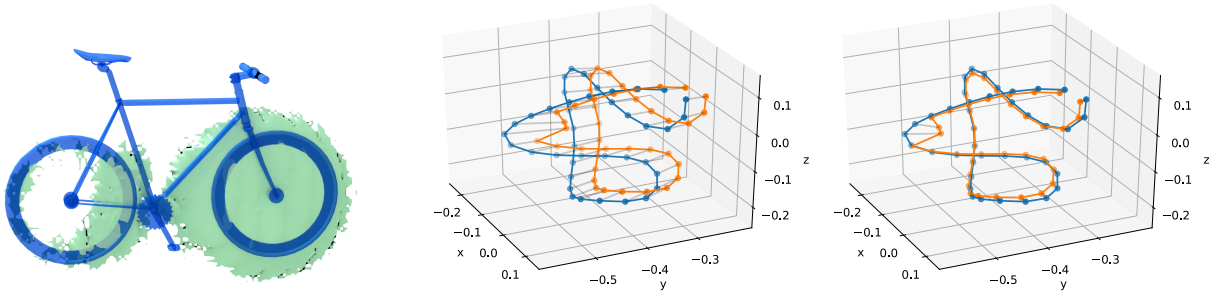
5.4.3 Position and orientation tracking

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 5.4c. 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 and P' with $\boldsymbol{P} = (\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n-1})^\mathsf{T}$ with $\boldsymbol{p}_i = (p_i^x, p_i^y, p_i^z) \in \mathbb{R}^3$ are compared by computing the root-mean-square (RMS) error:

$$\|\boldsymbol{P} - \boldsymbol{P}'\|_{\text{pos}} = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} \|\boldsymbol{p}_i - \boldsymbol{p}_i'\|^2}. \tag{5.3}$$



(a) (b) (c)

Figure 5.4: (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.

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 $S = \|(\boldsymbol{P} - \boldsymbol{x}) - \boldsymbol{P}'\|_{\text{pos}}$, the length of the offset $\|\boldsymbol{x}\|$ 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 $\boldsymbol{q}$ and $\boldsymbol{q}'$ in quaternion representation, the difference is computed by the unit quaternion dot product metric

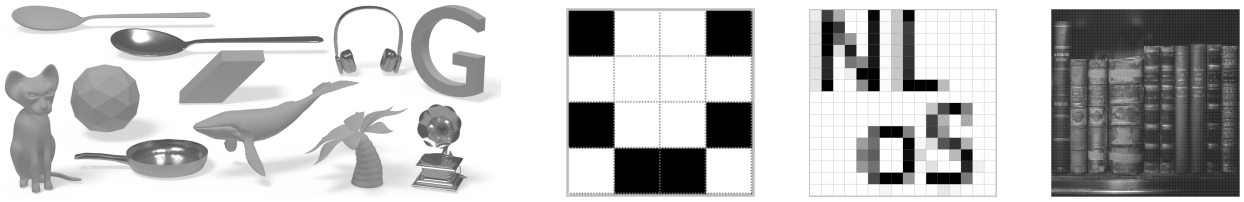
$$\|\boldsymbol{q} - \boldsymbol{q}'\|_{\text{quat}} = 1 - |\langle \boldsymbol{q}, \boldsymbol{q}' \rangle| \in [0, 1]\,,$$

where $|\langle \cdot, \cdot \rangle|$ denotes the absolute value of the dot product between the four components of the quaternions [Huy09]. 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.

5.4.4 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 5.5a. 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 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.



(a) (b) 4×4 (c) 16×16 (d) 128×128

Figure 5.5: a) Classification data set. Overview of the eleven different models used for the classification challenge. b-d) Example textures from the texture reconstruction challenge. The textures have different resolutions and different color depths.

5.4.5 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 5.5.

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 $\boldsymbol{T} \in [0,1]^{n \times n}$ in which n is a power of 2, we compute a Laplacian pyramid by iteratively blurring and downsampling $\boldsymbol{T}$, and storing the differences between the steps in the individual pyramid layers [BA87]. This essentially decomposes the image into its different frequency components.

Let $\boldsymbol{T}_n, \boldsymbol{T}_{n/2}, \ldots, \boldsymbol{T}_1$ respectively $\boldsymbol{T}'_n, \boldsymbol{T}'_{n/2}, \ldots, \boldsymbol{T}'_1$ denote the individual pyramid layers. The differences between each layer are computed by applying the Frobenius norm $\|\cdot\|_\mathrm{F}$ onto the texture difference $\boldsymbol{T}_n - \boldsymbol{T}'_n$ 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

$$\sqrt{\frac{\|\boldsymbol{T}_n - \boldsymbol{T}'_n\|_\mathrm{F}^2}{n^2}}, \sqrt{\frac{\|\boldsymbol{T}_{n/2} - \boldsymbol{T}'_{n/2}\|_\mathrm{F}^2}{(n/2)^2}}, \ldots, |\boldsymbol{T}_1 - \boldsymbol{T}'_1|. \tag{5.4}$$

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

$$\|\boldsymbol{T} - \boldsymbol{T}'\|_\mathrm{img} = \sqrt{\frac{\sum_{i=0}^{\log_2(n)} \|\boldsymbol{T}_{n/2^i} - \boldsymbol{T}'_{n/2^i}\|_\mathrm{F}^2 / (n/2^i)^2}{\log_2(n) + 1}} \tag{5.5}$$

is used as a quality metric.

5.5 Reconstruction results

With the exception of Arellano et al. [AGJ17], reconstruction code is not available, making the comparison challenging. We therefore seed our reconstruction challenge with the de-facto standard reconstruction method, ellipsoidal back projection, using the implementation of [AGJ17]. Its generality and ubiquity (as well as the lack of general-purpose alternatives) makes back projection 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 5.4. Exact numbers for each scene are found in the supplementary material.

5.6 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/`.

Supplemental material

This document contains supplementary information for *A Quantitative Platform for Non-Line-of-Sight Imaging Problems*. Sections are self-contained and they are not necessarily meant to be read in order.

5.A Challenge design

5.A.1 Setup size and geometry

Our setup uses arbitrary units, but nevertheless we have to decide on the proportions of the individual parts. As the setups in the previously published work vary greatly (see Table 5.1),

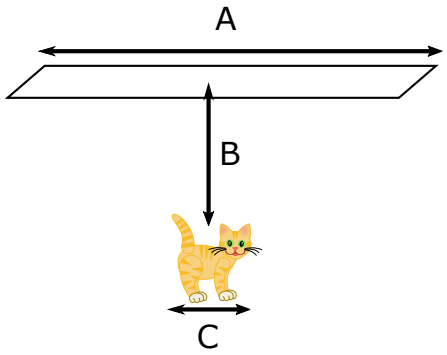


Figure 5.6: We categorize existing setups by the ratios of the reflector (A), distance to the object (B) and the size of the object(C).

we focused on three main quantities: The size of the reflector, the distance between object and reflector and the size of the object itself (see Figure 5.6).

We chose the size of our setup (shown in Figure 5.3) by taking the geometric mean of the individual values and applied some manual adjustment (e.g., particularly difficult proportions were weighted less, if the reported results on them are inferior to the average).

Other interesting quantities of setups are its temporal resolution T, as well as the number and distribution of directly visible scene locations that are illuminated (N_i) and observed (N_o) over the course of the measurement routine. They are also shown in Table 5.1. We oriented our temporal resolution towards the resolution of streak cameras [Vel+12], as they offer the highest resolution and use a single illumination point with a regular grid of observation points.

Table 5.1: Key specifications for various setups reported in literature: temporal resolution T (histogram bin size or point spread function, whichever is greater); numbers of observed (N_o) and illuminated (N_i) locations; scene dimensions A, B and C as illustrated in Figure 5.6; and ratios of these dimensions. All values are approximate; those in parentheses have been estimated by authors of this paper from information provided in the respective works. Entries marked AMCW or ∞ denote amplitude modulated correlation sensors and steady-state intensity imagers, respectively.

Ref.	T [ps]	N_o	N_i	A [cm]	B [cm]	C [cm]	A/B	B/C
[Kir+09]	250	1	1	(2.5)	(4.3)	(3)	0.6	1.4
[Nai+11]	1.6	(672×512)	> 1	25	(15)	(1–1.5)	1.6	12
[Vel+12]	15	672	> 1	40×25	25	1.5×8.2	1.4	5
[Hei+14]	AMCW	160×120	1	200	150	(80)	1.3	1.8
[But+15]	30	1	185	100×80	150	40	0.6	3.8
[Kad+16]	AMCW	176×144	1	200×100	100	4	1.5	25
[Kle+16]	∞	320×240	1	130	(60)	80×30	2.1	1.2
[Gar+16]	110	32×32	1	15×30	45	30×10	0.5	2.6
[Cha+17a]	64	3	1	30	50	15×15	0.6	3.3

5.B Data sets

All data sets are rendered using the scene arrangement illustrated in Figure 5.3. Data is provided as transient images $\boldsymbol{I}(u, v, \tau)$, where the image coordinates $(u, v) \in [0, 1, \ldots, 255]^2$ address square-shaped wall elements ("pixels") of size 0.002^2 in the X-Z plane, and the τ dimension is discretized in 1600 bins of size $d\tau = 0.001$ starting at $\tau = 0$ (τ is a measure of the path length and must be divided by the speed of light to retrieve the travel time). The exchange format for transient images is specified in Section 5.D. Extents and discretization of the temporal dimension are specified within each data set.

5.B.1 Geometry reconstruction

We consider geometry reconstruction the most important challenge, as it is not only the focus of most of the previous work, but also the most general problem with the highest number of degrees of freedom. In most scenarios, once a full geometric scene model has been reconstructed, derivative information such as object positions or classes can be obtained more easily from three-dimensional geometry than from from raw transient images.

We split our test scenes into several categories that test different capabilities of solvers. Each scene contains a single object which has to be reconstructed. We define five categories of objects:

Cat. 1 Two-dimensional shapes. This category contains two-dimensional objects of a certain thickness perpendicular to the wall. This category is closely related to the texture reconstruction challenge (Section 5.B.4).

Cat. 2 Simple geometric shapes. Objects with simple mathematical descriptions without additional surface details.

Cat. 3 Simple objects. Everyday objects from the real world, with limited geometric detail. Objects in this category are not easily approximated by the shapes of the previous category.

Cat. 4 Complex objects. Highly non-convex objects with complex shape but without fine surface details.

Cat. 5 Difficult objects. Objects with thin elements, fine structures and complex topology (e.g. many holes).

A full listing of data sets is given in Table 5.2. To facilitate the development and refinement of reconstruction techniques, ground truth geometry in `.obj` format is provided for some of the data sets; for all others, the true geometry remains unknown.

5.B.2 Position and orientation tracking

For tracking, three different rigid objects are used: a sphere, a golem figurine and an airplane (see Table 5.3). For each object, there are a total of four challenges:

Table 5.2: Objects and their categories for the geometry reconstruction challenge.

Category	Data set name	Material	Ground truth provided
1	`LetterK`	diffuse	yes
1	`LetterQ`	diffuse	no
2	`Box`	diffuse	yes
2	`Cone`	diffuse	no
3	`StanfordBunny`	diffuse	yes
3	`UtahTeapot`	diffuse	yes
3	`Ax`	diffuse	no
3	`Hammer`	diffuse	no
3	`Cup`	specular	no
4	`StanfordDragon`	specular	yes
4	`Dinosaur`	diffuse	no
4	`FlyingDragon`	diffuse	no
4	`IndoorPlant`	diffuse	no
5	`Chair`	diffuse	no
5	`Bike`	specular	no
5	`Greenhouse`	specular	no

1. movement along the three main axes without rotation,
2. rotation at a fixed position along the three main axes,
3. movement along a complex path with constant orientation and
4. movement along a complex path while changing orientation.

Due to its symmetry, the challenges involving rotations are not included for the sphere data set. The paths are different for each object and challenge, e.g. the golem moves along one path with constant orientation and along another path for the combined orientation and translation.

The object origins lie in the center of mass for each object. This way, the object position is uniquely defined for an ideal reconstruction of the scene, however we do expect a certain bias in the position in realistic scenarios. Therefore we consider the residual RMS to be the most important metric.

Each path forms a loop and contains 40 frames. The axes movement and static rotation consist of 30 frames, with 10 frames for each axis. The movements and rotations around the axes are designed to be easy to reconstruct and to make systematic errors and missed frames obvious.

The exact geometry of the golem is given as an `.obj` file and can be used to improve the reconstruction (e.g., by fitting it into a partial geometric reconstruction of each frame). The shape of the airplane is not revealed to test tracking of unknown objects.

5.B.3 Object classification

The goal of the object classification task is to assign the transient image to one of eleven object geometries, as listed in Table 5.4 and shown in Figure 4a in the main paper. All

Table 5.3: Overview of the object tracking data sets.

Data set name	Material	Shape known	Position	Rotation
`SphereAxesPos`	diffuse	yes	yes	no
`SpherePathPos`	diffuse	yes	yes	no
`GolemAxesPos`	diffuse	yes	yes	no
`GolemAxesRot`	diffuse	yes	no	yes
`GolemPathPos`	diffuse	yes	yes	no
`GolemPathRot`	diffuse	yes	yes	yes
`AirplaneAxesPos`	specular	no	yes	no
`AirplaneAxesRot`	specular	no	no	yes
`AirplanePathPos`	specular	no	yes	no
`AirplanePathRot`	specular	no	yes	yes

Table 5.4: Overview of the object classification data set.

Data set name	Material
`Cat`	diffuse
`Icosphere`	diffuse
`LetterG`	diffuse
`Parallelepiped`	diffuse
`Plant`	diffuse
`SpoonDiffuse`	diffuse
`Whale`	diffuse
`Gramophone`	specular
`Headphones`	specular
`Pan`	specular
`SpoonSpecular`	specular

object shapes are given as `.obj` files and should be used to perform the classification.

The objects were scaled to equal surface area to prevent classifying by the size of the object, i.e. amount of reflected light. However, this approach is not necessarily perfect as very compact or concave objects appear smaller when scaled by their surface area.

5.B.4 Planar textures

In this challenge, a textured planar target of extent $(x, z) \in [-0.1, 0.1]^2$ placed in front of the laser spot at $y = -0.3$ (see Figure 5.7) is the subject of the reconstruction. The texture, represented by a grayscale image of dimension $n \times n$, modulates the albedo (diffuse reflectance) of the surface. Each value $\rho_{s,t}$ within the texture covers a square-shape region of size $(0.2/n, 0.2/n)$ with a value in the range of $[0, 1]$. We provide a variety of data sets that feature black-and-white and grayscale textures of different resolution (Table 5.5).

For the evaluation, results with arbitrary resolutions are supported. If the reconstruction $\boldsymbol{T}'$ has the same resolution as the reference, identical decomposition steps can be applied. Otherwise it is scaled to the closest power of 2 before decomposing it. If the pyramid of

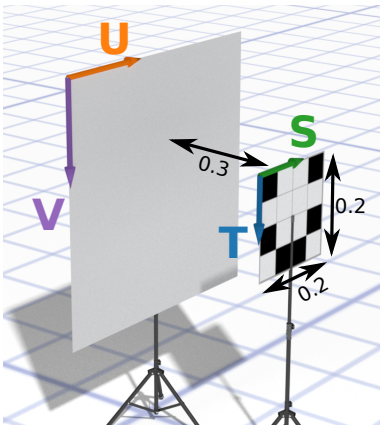


Figure 5.7: Texture reconstruction setup.

Table 5.5: Overview of the texture reconstruction data sets.

Data set name	Resolution	Pixel depth
`Character`	4×4	$\{0, 1\}$
`Digit`	4×4	$\{0, 1\}$
`Letter`	4×4	$\{0, 1\}$
`Smiley`	4×4	$\{0, 1\}$
`House`	16×16	$\{0, 0.25, 0.5, 0.75, 1\}$
`Number`	16×16	$\{0, 0.25, 0.5, 0.75, 1\}$
`Pattern`	16×16	$\{0, 0.25, 0.5, 0.75, 1\}$
`Text`	16×16	$\{0, 0.25, 0.5, 0.75, 1\}$
`Books`	128×128	$[0, 1]$
`Concert`	128×128	$[0, 1]$
`Fan`	128×128	$[0, 1]$
`Industrial`	128×128	$[0, 1]$

the reconstruction contains more layers (i.e. it had a higher input resolution), the highest layers are discarded; if it contains fewer layers, the missing ones are filled with zeros (as it did not contain any information about the higher frequencies). Therefore, reconstructions in the reference solution are preferred.

5.C Rendering

We used a modified version of `pbrt-v3` [PJH16] to render the transient images.

5.C.1 Importance sampling

The transient images of our scenes contain only light from indirect reflections, which can make the rendering very inefficient if no special care is taken. Sampling from the wall towards the light source is futile, as the laser spot illuminates only the object, not the wall. Sampling the hemisphere over the wall is inefficient, as most objects of interest only cover a small solid angle on the hemisphere and are thus unlikely to be hit.

To improve the performance for this light transport scenario, we implemented a custom importance sampling that is heavily inspired from the area light source sampling already implemented in `pbrt-v3`. The triangles of the object are stored in a special list and both the wall and the object triangle are tagged with specific flags. During path tracing, everytime a ray hits the wall, we can sample directly into the direction of the object, as the wall can only receive light that was reflected from the object. These samples need to be normalized by considering the area, angle and distance of the triangle of the object to ensure the correct expected value of the sampling.

Our tests show that this custom importance sampling is two to three orders of magnitude more efficient than naive sampling. Physical correctness is preserved by only eliminating zero-radiance paths, e.g. interreflections on the object surface remain untouched by this optimization.

5.D Transient image files

As of writing this paper, there is no standard format for storing transient images. It would seem like an canonical and attractive choice to resort to standard image formats for which suitable I/O libraries exist; however, such files would have to be accompanied by separate metadata specific to the data set, and most image formats rigidly adhere to a Cartesian pixel arrangement. Custom formats have also been proposed, for example Arellano et al.'s `.float` and `.lasers` files [AGJ17], a format that would require thousands of individual files to store a single data set from our database, and in its current form is not expressive enough to cover new capture geometries such as O'Toole et al.'s confocal setting [OLW18]. We therefore propose a new format that is compact (one file per data set), easy to read and write, and at the same time flexible enough to cater to the needs of emerging research directions. Here we give a high-level overview over the file format. A detailed implementation guide comes as part of the SDK.

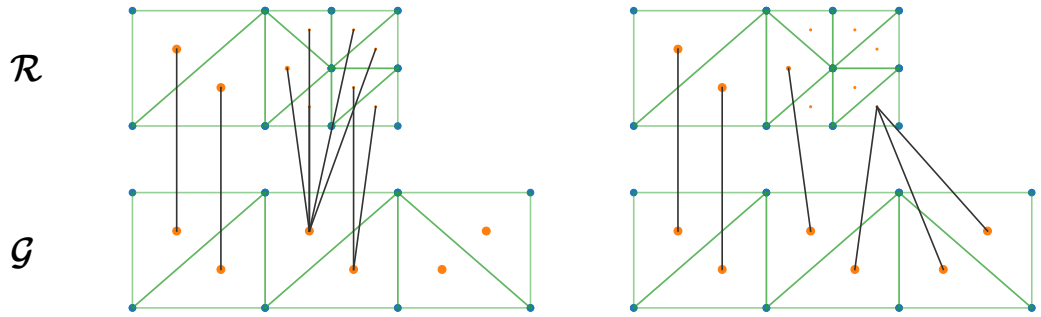
A transient image file consists of 4 blocks: The *file header* contains general information and the sizes of each remaining block. The *pixel data* block contains a linear array of transient pixels. The pixel interpretation block is an efficient representation of the illumination and observation points of each pixel. Finally, the *image properties* block contains arbitrary, JSON-encoded meta-data of the image.

For the *pixel interpretation* and *image properties* blocks we made some noteworthy design choices, which we will discuss in the following.

5.D.1 Pixel interpretation block

Traditionally, a single point on the reflector is illuminated while a regular grid of points on the reflector is observed. Due to the reciprocity of the light transport, this can be reversed, e.g. as done by Buttafava et al. [But+15]. In general, multiple illumination and observation points can be used (which may or may not be arranged in a regular grid), and in the extreme case of [OLW18], a unique illumination and observation point is used for every pixel.

To support all these cases and still have an efficient representation of our data, the observation and illumination points can be stored in different modes. Mode 0 is the most general



(a) Top to bottom: $d(\mathcal{R}, \mathcal{G})$ (b) Bottom to top: $d(\mathcal{G}, \mathcal{R})$

Figure 5.8: Illustration of the correspondence selection for our surface comparison metric. For each triangle of the source, the closest triangle of the target is selected. The lower object is the ground-truth geometry $\mathcal{G}$, while the upper object is the reconstruction $\mathcal{R}$.

one and requires no structure in the observation or illumination points. They are stored for every pixel individually, however this introduces a certain overhead (which becomes negligible, if each pixel consists of a large number of bins). The SDK provides an upconverter to Mode 0 from the other, more specialized ones.

Mode 1 assumes a single illumination point and a regular grid of observation points, thus the transient image has a meaningful x and y resolution. Observation point positions are implicitly stored by the grid properties and their position in the linear *pixel data* array (as it is the case in raster graphic formats). Mode 2 is the reciprocal case, where the roles of observation and illumination points are reversed.

5.D.2 Image properties block

Image meta-data is widely used to store additional information such as the camera settings used to capture the image. In Transient image files they are stored as an UTF-8 encoded JSON string at the very end of the file. A number of standard fields are specified, however users are free to add their own ones.

This approach has multiple advantages: Meta-data can be read and written using a binary-compatible text editor, all fields are optional, new properties can easily be added, and readers and writers are quick to implement due to the wide availability of JSON en-/decoders.

5.E Comparison metrics

Figure 5.8 shows how the closest points are selected during the evaluation of the asymmetric mesh-to-mesh distance (according to Equation 5.2). The reconstruction $\mathcal{R}$ (top) misses the right third of the surface that is known to exist in the ground-truth mesh $\mathcal{G}$ (bottom), and it has a finer tessellation in the middle segment. In the distance from reconstruction to ground truth, this results in more connections in the middle segment, compared to the left segment.

Table 5.6: Asymmetric reconstruction errors between a fast back projection reconstruction (M1), a ground truth mesh (M2) and the same mesh after one level of Catmull-Clark subdivision (M3). In boldface, the distance $d(\mathcal{G}, \mathcal{R})$ from ground truth to a test geometry measures the incompleteness of the reconstructed surface. As intended by design, this distance measure is significantly less sensitive to remeshing (third row) than it is to actual missing geometry (first and second rows).

Comparison	$d(\mathcal{R}, \mathcal{G})$	$d(\mathcal{G}, \mathcal{R})$
$\mathcal{R}$ = M1, $\mathcal{G}$ = M2	$5.255 \cdot 10^{-3}$	$\mathbf{1.819 \cdot 10^{-2}}$
$\mathcal{R}$ = M1, $\mathcal{G}$ = M3	$5.132 \cdot 10^{-3}$	$\mathbf{1.813 \cdot 10^{-2}}$
$\mathcal{R}$ = M3, $\mathcal{G}$ = M2	$4.729 \cdot 10^{-4}$	$\mathbf{6.438 \cdot 10^{-5}}$

Some of these connections are shorter and some are longer, but their average is roughly the same as in the case of equal tessellation. As they are weighted by the triangle area, the total cost for the middle part does not increase significantly by the additional connections. Overall, the cost is quite similar to the leftmost geometry segment.

The right part of $\mathcal{G}$ is missing in $\mathcal{R}$. In the distance from original to reconstruction ($d(\mathcal{G}, \mathcal{R})$), this results in longer connections and thus in increased cost.

Additionally, the misalignment of the meshes increases the length of all connections and thus the overall distance.

The tessellation of an object does have a certain influence on the comparison result; however, it is small. To avoid bias from tesselation, all used models are tessellated finer than the maximal expected reconstruction resolution. Furthermore, trivial subdivision of the triangles can further reduce this bias if needed and does not require to render or reconstruct the scenes again.

In Table 5.6, we compare the reconstruction error introduced by a change in tessellation to that of a state-of-the-art reconstruction.

5.E.1 Back face culling

Under normal conditions, it cannot be expected that the backside of an object can be well reconstructed, as usually little to no information will reach the wall. Therefore triangles facing away from the reflector are discarded before the mesh distance is evaluated.

Figure 5.9 shows an example of a triangle that is pointing away from the reflector but still visible from the laser spot. We use this visibility of the laser spot as a culling criterion, instead of only checking whether the face normal is pointing away from the reflector. This is still not always correct, as global illumination can also allow light from culled triangles to reach the reflector (and likewise, triangles facing towards the reflector can be completely occluded), but taking all of these effects into account is not possible in a simple and transparent manner.

5.F Tools

We provide a variety of tools for handling the transient images. At the core are loaders and writers for various programming languages including C++, Python and Matlab. Together

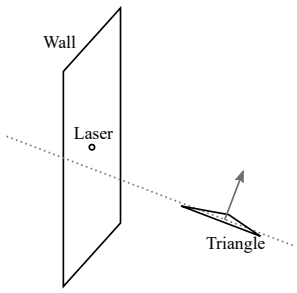


Figure 5.9: Culling of back faces. Triangles that are facing away from the laser spot $(x, y, z) = (0, 0, 0)$ are removed from the comparison. All others are kept, even those whose normal vector points in positive z direction (away from the wall).

Table 5.7: Fast back projection reconstruction results. a) Geometry reconstruction: The greater number of the two (marked in bold) is the symmetric distance of reference and reconstruction (see Equation 5.2). b) Position tracking: Columns from left to right: RMS distance (see Equation 5.3), offset length, RMS residual after subtraction of offset, completeness of trajectory (percentage of recovered frames).

(a)

Scene	$d(\mathcal{R}, \mathcal{G})$	$d(\mathcal{G}, \mathcal{R})$
Axe	0.00376	**0.00645**
Bike	0.00675	**0.00916**
Chair	0.00429	**0.0137**
Cone	**0.0129**	0.00867
Cube	**0.0743**	0.00686
Cup	0.00809	**0.0283**
Dino	0.00500	**0.0172**
Dragon	**0.0128**	0.00453
Greenhouse	0.0133	**0.0200**
Hammer	**0.0108**	0.00876
IndoorPlant	0.00438	**0.0162**
K-Letter	**0.0200**	0.00734
Q-Letter	0.00625	**0.00631**
StanfordBunny	0.00356	**0.0155**
StanfordDragon	0.00359	**0.0202**
UtahTeapot	0.00341	**0.0362**

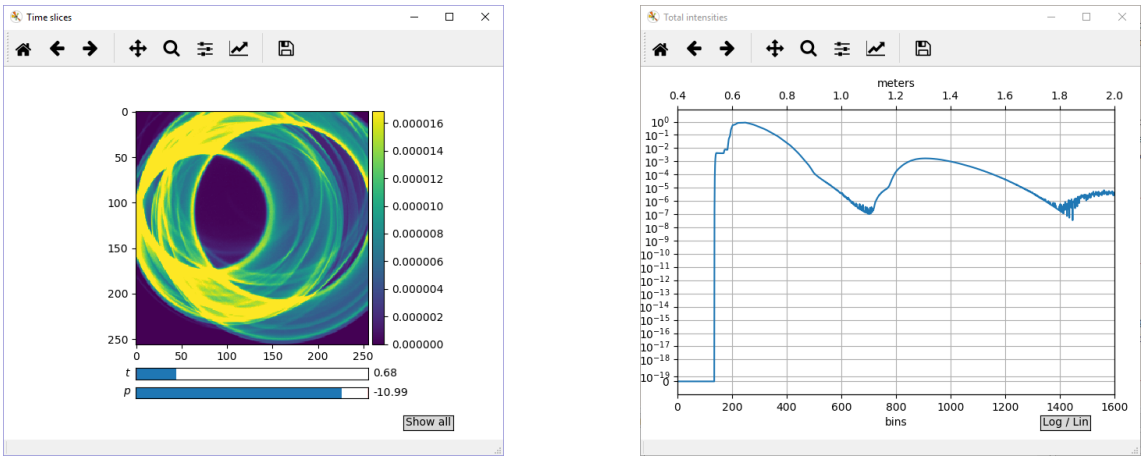
(b)

Scene	RMS dist.	‖Offset‖	RMS res.	Compl. [%]
Golem Axes	0.0238	0.0216	0.0101	100
Golem Path	0.0685	0.025	0.0638	100
Sphere Axes	0.0488	0.0485	0.0056	100
Sphere Path	0.0535	0.0504	0.018	100
Plane Axes	0.0269	0.0127	0.0237	100
Plane Path	0.05	0.0167	0.0472	100

with the file format description, they should allow a quick integration of our data sets in other frameworks.

5.F.1 Image viewer

We provide a simple viewer for transient images based on Python and Matplotlib. The user can scroll through time and adjust the intensity scale. A second view shows the transient image integrated over the spatial domain. The resulting histogram illustrates how much light arrived at what time, on either a linear and logarithmic scale. The viewer is shown in Figure 5.10.



(a) Time slices (b) Histogram

Figure 5.10: Our viewer shows time slices and histograms of transient images.

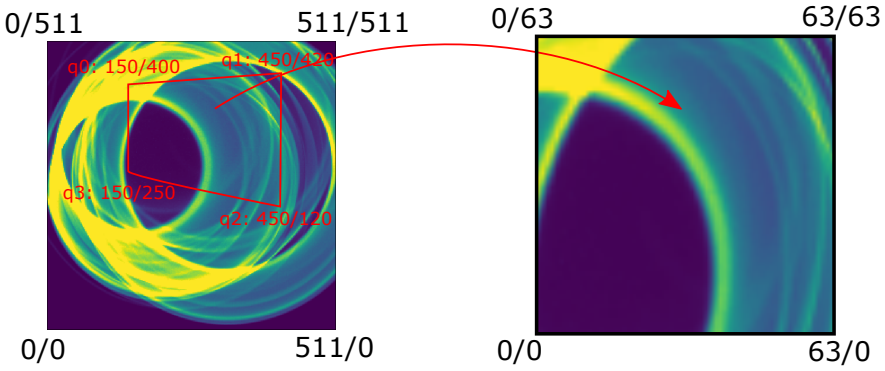


Figure 5.11: The camera converter uses a homography defined by four points pairs to resample a transient image.

5.F.2 Setup converter

Many setups seen in the real world have different illumination and viewing geometries. While a change in laser spot position would require re-rendering the scene, other camera placings and projections can be accommodated to make results more comparable. To this end, we offer a resampling tool to convert transient images to different camera positions.

At the heart of the resampling is a homography as depicted in Figure 5.11. The user defines the four point pairs in the old and new image from which the homography is computed. Applying it to the image jointly crops, transforms and rescales the image. The user can also specify a camera and laser position as three-dimensional coordinates which are used to compute a temporal offset for each output pixel. Additional, the temporal window of the output image can be changed. For the resampling, a Mitchell-Netravali filter with customizable size for both spatial and temporal filtering is used. If no size is specified, reasonable filter sizes are computed from the homography.

The tool is written in C++ with no external dependencies. It is implemented as command line tool and thus ready for integration in batch processing.

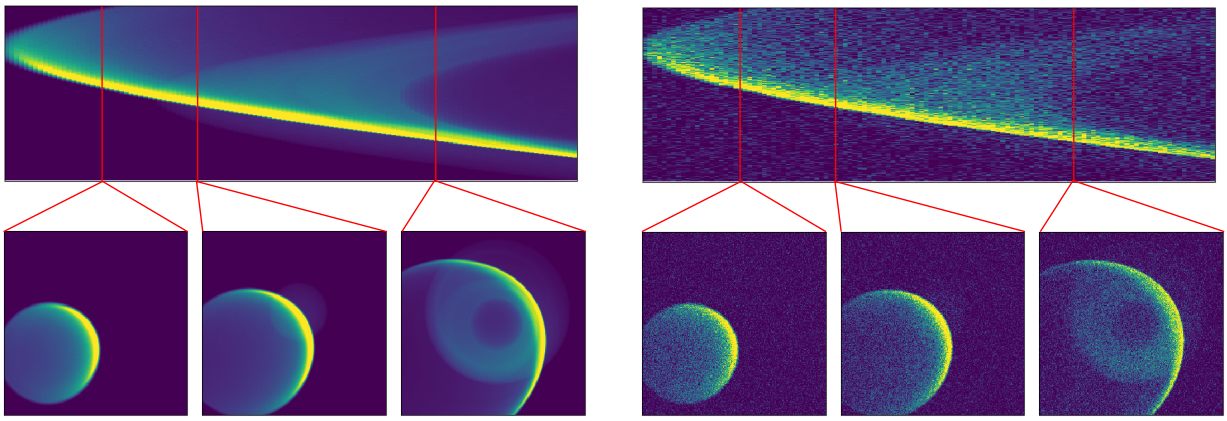


Figure 5.12: Transient image of the `Hammer` scene before (left) and after (right) applying the noise model `SPAD`.

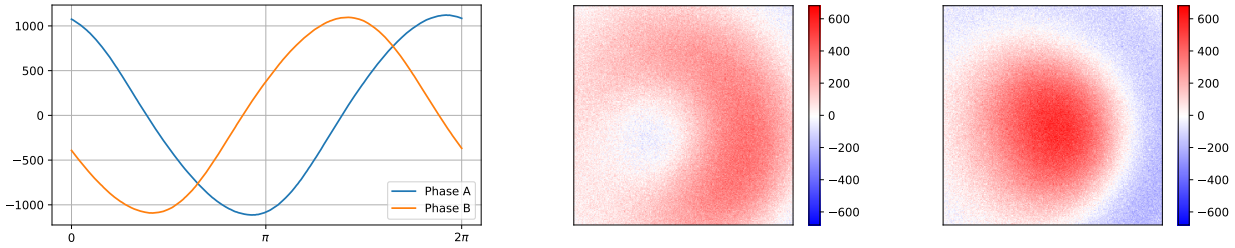


Figure 5.13: Synthetic AMCW measurements of the `Hammer` scene. Left: Measured demodulation functions from a *PMDTec CamBoard nano* AMCW camera. Middle/Right: The two phase images (0° and 90°) computed using the sensor model.

5.F.3 Fast back projection integration

All scripts that were used to export the transient images to the *fast back projection* solver by Arellano et al. [AGJ17] are available on our website. This allows the user to set up a complete reconstruction pipeline and re-evaluate all results in the paper.

5.F.4 Sensor models / noise

The suite of scripts and tools contains two noisy sensor models to reflect the characteristic behavior of two important types of device: `AMCW`, a simple model of a correlation ToF sensor (4-tap near-sinusoidally modulated correlation time-of-flight measurement with Skellam-distributed shot noise) and `SPAD`, a single-photon counter with Poisson-distributed shot noise and dark counts. Figure 5.12 shows an example data set before and after applying the `SPAD` model with standard settings. Figure 5.13 shows images for the same scene as seen through the `AMCW` sensor model.

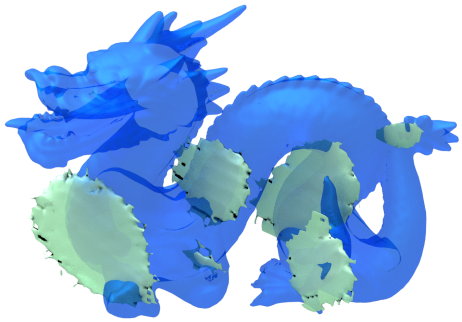


Figure 5.14: Reconstruction of the Stanford Dragon. The ground truth geometry is shown in blue, while the reconstructed geometry is green.

5.G Reconstruction results

Our evaluation metrics aim to make different reconstruction algorithms comparable by reducing their overall performance to a single number (that are shown in Table 5.7). However, these error terms arise from various characteristics of the algorithm which are interesting to study, as they increase the understanding of the behavior and point at possible improvements. Thus we now discuss some characteristics of the back projection example used in the paper. It should be noted that we did not tune parameters to achieve the highest possible accuracy. Instead, these examples serve to illustrate the typical behavior of a reconstruction algorithm.

5.G.1 Geometry reconstruction

Figure 5.14 shows the reconstruction of the Stanford Dragon model. The reconstruction mainly consists of planar patches that are parallel to the wall which is very typical and seen in almost all reconstructions. The Laplacian filter used after back projection to isolate surfaces that favors flat structures and thus struggles with surfaces that are curved or not aligned with the wall. The resulting reconstructions are often incomplete, low in detail, and they feature a distinctive “cloud-of-pancakes” look.

5.G.2 Position tracking

Our naive tracking position implementation reconstructs first the object geometry and then uses its center of mass as object position. Thus it is very vulnerable to incomplete geometry reconstructions.

Figure 5.15 shows two frames of the `AirplaneAxesPos` data set, where the plane moves along the X axis. Both reconstructions are incomplete and favor geometry close to the laser spot, which lies in between both positions. When the center of masses are computed, the movement of the object thus appears to be smaller than it actually is (see Figure 5.16). A more sophisticated algorithm could be aware of this shortcoming and try to fit the given object geometry into its reconstruction to determine which part of the plane was reconstructed.

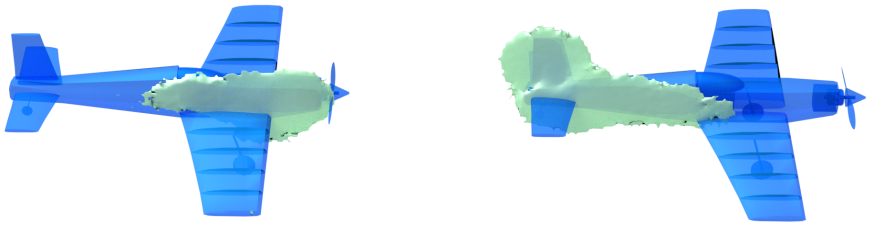


Figure 5.15: Reconstructed geometry as it is used during position tracking. The ground truth geometry is shown in blue, while the reconstructed geometry is green. In different frames, different parts of the plane were reconstructed, resulting in an error in the position reconstruction.

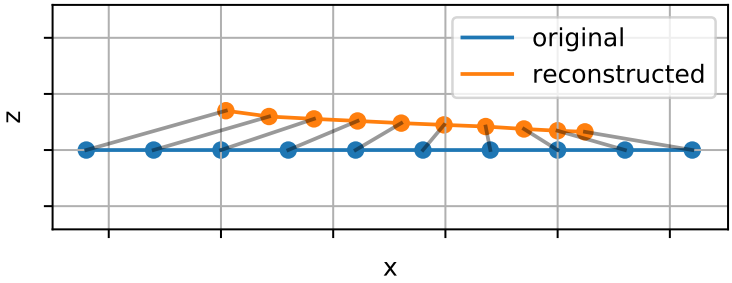


Figure 5.16: Reconstructed trajectory of the airplane for the movement along the X axis. Apart from the offset in Z direction, the reconstructed path is too short.

CHAPTER 6

A Calibration Scheme for Non-Line-of-Sight Imaging Setups

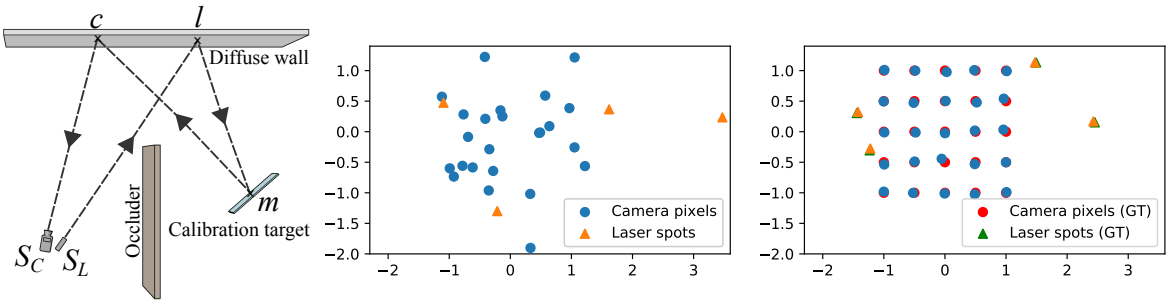
This chapter was published as a peer-reviewed paper in the *Optics Express* journal by the *Optical Society* in 2020 [Kle+20].
The authors are Jonathan Klein, Martin Laurenzis, Matthias B. Hullin, and Julian Iseringhausen.

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.

6.1 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 Figure 6.1a for a schematic illustration. The available operation modes [Vel+12; Kir+09; Hei+19; OLW18; Gar+15; IH20] support not only object detection



a) Scene setup b) Initialization c) Calibration result

Figure 6.1: (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 two-dimensional 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.

and tracking of components of the occluded scene but extend to the full reconstruction of three-dimensional 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 [Mus+19], randomly scattered measurement locations [But+15] as well as low-resolution imagers and even correlation time-of-flight sensors [Hei+14]. Additionally, task-specific constraints (e.g., pixel positions restricted to a scan line) are easily integrated in the method.

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.

6.2 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 [KSS12; Kle+16; Bou+17; Thr+18; Sei+19; Che+19], 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 [Jar+17]. Seminal works include the recovery of low-parameter geometry and reflectance models from transient measurements [Kir+09; Nai+11] as well as the first reconstruction of distinct shapes [Vel+12]. Since then, significant effort has been devoted to unlock novel sensor technologies and interferometric setups for transient imaging [Hei+13; Gar+15; Gki+15] while simultaneously improving the performance of the de-facto standard reconstruction technique, ellipsoidal error back projection [Vel+12; LV14; AGJ17]. Recent additions to the non-line-of-sight reconstruction problem include the introduction of the confocal capture setting [OLW18] as well as attempts to cast the problem into paradigms borrowed from wave optics and seismic tomography [Liu+19; LWO19]. While most of these works rely on volumetric representations for the hidden target, other researchers have explored alternative, surface-driven representations as well [Ped+17; IH20; TSG19]. 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 [Hei+14; Hei+19]. Lately, there has also been some work introducing machine learning algorithms to NLoS reconstructions [Cho+20; Met+20; Che+19; Car+18].

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. [Gar+15; Kle+16]), extending the setup by dedicated calibration hardware is a common approach. Buttafava et al. uses a web cam to estimate the three-dimensional position of the visible laser spot, however the webcam itself is manually calibrated using a dot pattern [But+15]. 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 [La +20]. 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 [OLW18; LWO19]. Speckle correlation based approaches (e.g. [KSS12; Met+20]) 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 [Car+18; Cho+20]. However, a such trained network cannot be transferred to new setups.

6.3 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 [Nai+11] or retro-reflective objects [OLW18]), 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 [IH20; Kle+16]. 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 [Kle+16] 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.

6.3.1 Image formation model

Figure 6.1a 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 $m \in M$.

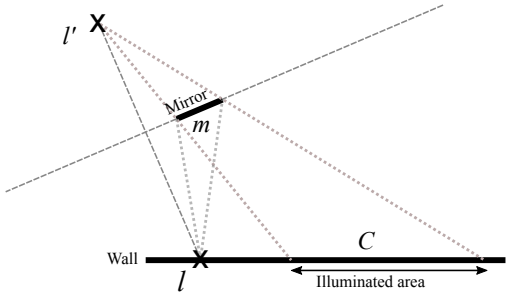


Figure 6.2: To assess the optical path $l \to m^r \to c$, we use a similarity relation: The laser spot l illuminates the wall as if it was reflected on the mirror plane, resulting in a virtual light spot l'.

Whether the pixels lie on a fixed grid (as for two-dimensional 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 6.3.3). 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 three-dimensional 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 $l \to m \to c$, 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 6.A.1).

A complete measurement consists of a series of paths $P_{i,j,k} = S_L \to l_i \to m_j \to c_k \to S_C$ (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 three-dimensional positions of laser spots and camera pixels).

For the time of flight computation we need to compute the length of a path $S_L \to l \to m \to c \to S_C$. 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 $l \to m^r \to c$ is equal to the path length $l' \to c$, where l' is the point l mirrored at m (see Figure 6.2).

A mirror plane is represented in the Hesse normal form as normal vector n and scalar

offset d. Then

$$l' = l - 2(n \cdot l + d)n \tag{6.1}$$

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

$$f\left(S, l, c, m\right) = \|l - S_L\| + \|c - l'\| + \|S_C - c\| . \tag{6.2}$$

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 Figure 6.2). In this case, the path can simply be removed from the optimization (see Section 6.3.2).

6.3.2 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

$$\underset{S,L,C,M}{\arg\min} \sum_{l \in L} \sum_{c \in C} \sum_{m \in M} \left\| f\left(S, l, c, m\right) - t_{l,c,m} \right\|^2 . \tag{6.3}$$

using a standard gradient descent algorithm (BFGS [TQ04]).

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 6.4. Due to the compact image formation model automatic differentiation can be used for gradient computation.

6.3.3 Parameterization

In Equation 6.3, 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 $g : p \to (S, L, C, M)$ 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.

Planar walls

Most current non-line-of-sight reconstruction approaches assume planar walls (with exceptions such as [LWO19; La +20]). 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.

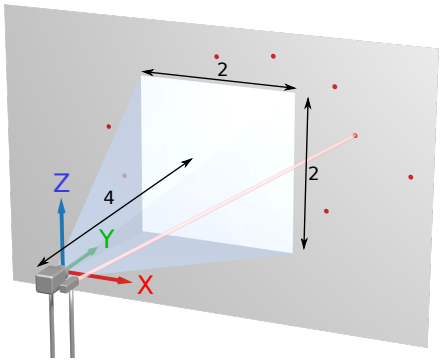


Figure 6.3: Setup used for the synthetic evaluation. The camera and laser are in the origin, the red dots mark the laser spot positions. A total of 40 wall-facing mirrors (not shown here) are placed between camera and wall.

Regular grids

On two-dimensional 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 two-dimensional coordinates of the image sensor to two-dimensional coordinates on the wall. Since two-dimensional 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 6.10 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.

6.4 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.

6.4.1 Evaluation setup

Our standard evaluation setup (shown in Figure 6.3) 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.

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 6.1 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

$$\mathrm{RMS}(P, Q) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \|P_i - Q_i\|_2^2}. \tag{6.4}$$

As mentioned before, the calibrated setup might be in a different coordinate system and naively applying Equation 6.4 can result in high errors even for actually good result. Therefore we use the Kabsch algorithm [Kab76] 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 Figure 6.1 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.

6.4.2 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:

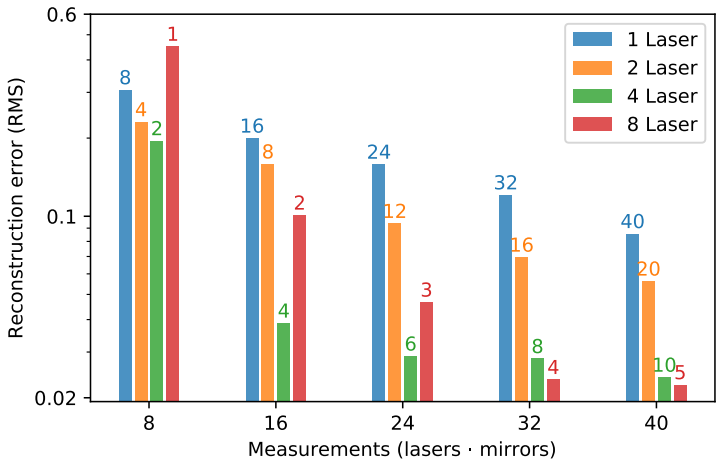


Figure 6.4: Calibration performance depending on the total number of measurements. Each data point shows the mean of 100 individual optimizations. The numbers above the bars show the number of mirrors used for that data point.

- 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 two-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 Figure 6.4 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 camera. We conclude that 32 measurements using at least 4 laser positions are a lower bound for a sufficiently accurate reconstruction.

Figure 6.5 shows the same data set as in Figure 6.4, but this time decoded in terms of its dependence on the initialization error. We find that within generous bounds the initialization

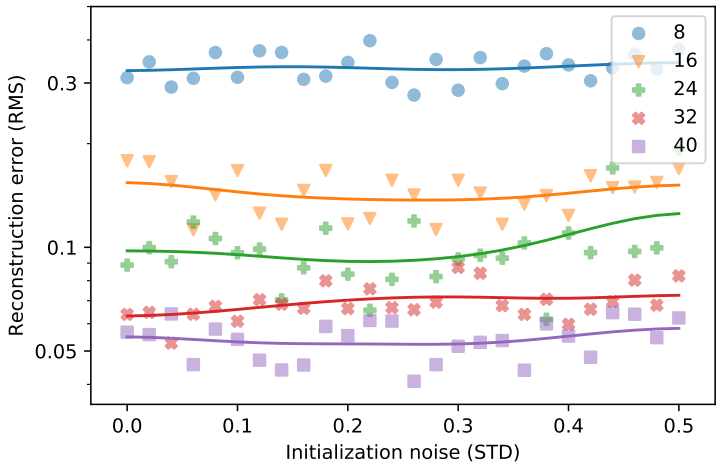


Figure 6.5: Reconstruction error for various levels of initialization noise. The data is the same as shown in Figure 6.4, averaged over all laser/mirror combinations for a specific number of measurements (shown in different colors / markers).

has no effect on the convergence of the optimization; the RMS error primarily depends on the number of measurements involved.

Figure 6.6 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 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 6.7 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.

The default parameterization supports arbitrarily shaped walls such as the curved wall shown in Figure 6.8.

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.

Additional evaluations of the impact of the setup geometry can be found in Appendix 6.C.

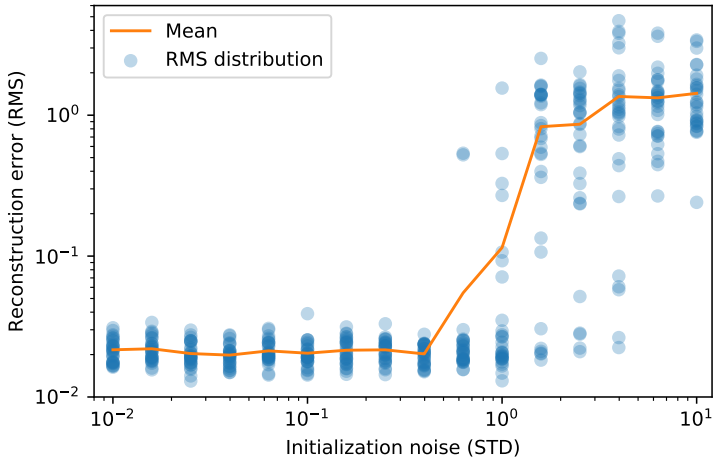


Figure 6.6: Reconstruction success depending on initialization noise. Time-of-flight noise is fixed at 0.02. The blue distribution of the individual optimization results gives an better intuition than the orange mean value - the results split in two distinct clusters for increased noise. Note that both axes are in log scale.

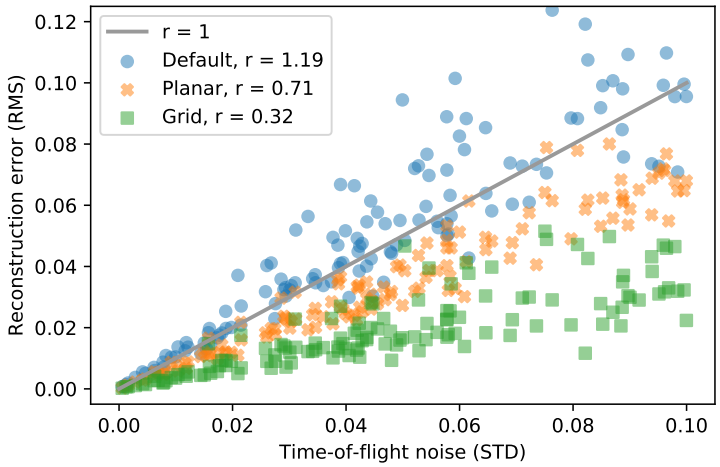


Figure 6.7: Calibration error depending on the time-of-flight noise and parameterization. The r-values show the slope of a linear fit for each parameterization.

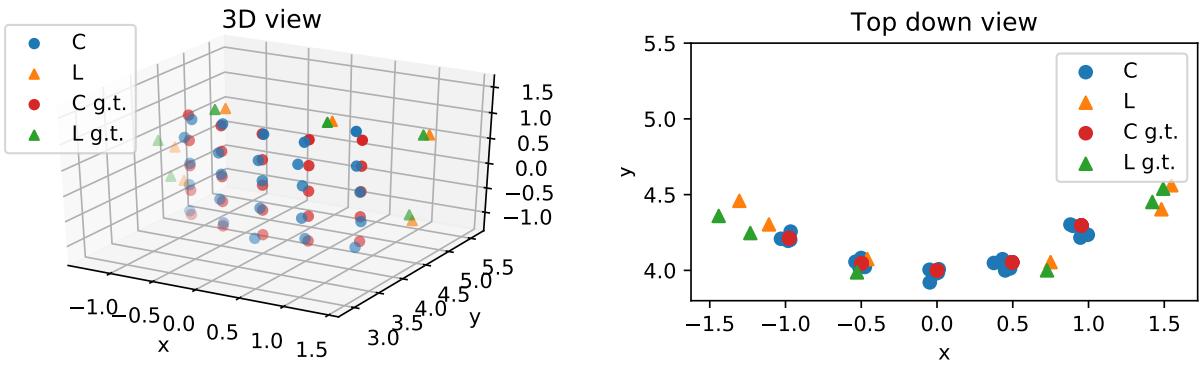


Figure 6.8: Example calibration of a curved wall. The setup consists of 6 lasers and 6 mirrors, the initialization noise is 0.5, the time-of-flight noise is 0.1. The RMS error of the calibration is 0.099.

6.4.3 Implementation and runtime

In our prototype Equation 6.3 is implemented purely in Python, the optimization of Equation 6.3 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.

6.5 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 Figure 6.9. It uses a total of 7 different laser spot and 7 mirror positions (as described in Section 6.4.2 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.

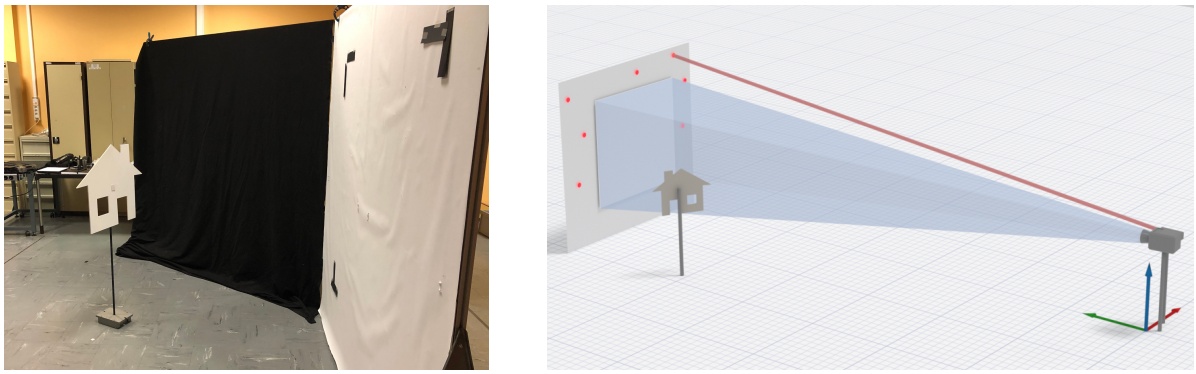


Figure 6.9: Photograph and schematic of our experimental setup. The reconstruction target is a house shape outside the field-of-view of the camera. The red spots on the wall show the 7 laser spot positions that are used.

6.5.1 Calibration results

Our hardware setup consists of a Princeton Lightwave 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 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 6.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 $\times$ 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 $l \to m \to c$ paths, we can automatically detect and remove invalid paths from the optimization (see Appendix 6.A.2 for details on the detection).

Figure 6.10 shows the calibration results. We evaluate a series of different initialization noise values (see Section 6.4.1), 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

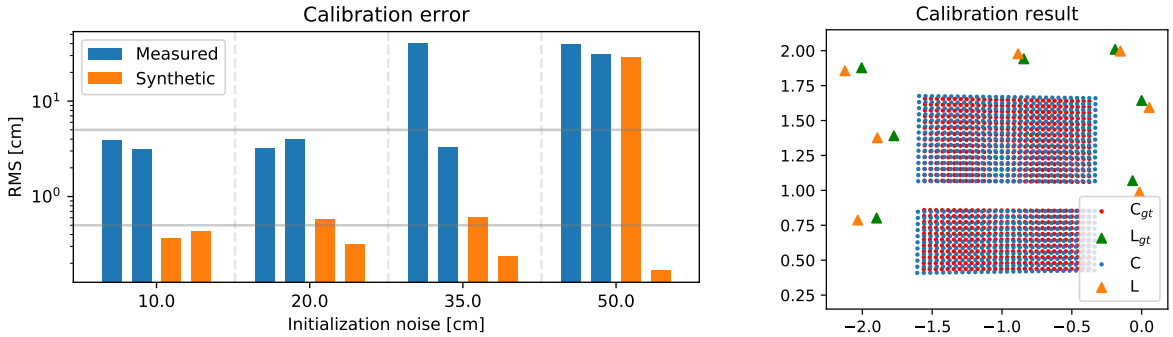


Figure 6.10: Calibration quality on the experimental setup. Left: The RMS is computed according to Equation 6.4. For each noise level, two initializations are created which are shared by the evaluations on both the measured and synthetic time-of-flight data. The gray lines show the 0.5 cm and 5 cm error. 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.

is given by the sensor pattern). In real applications, the worst case ($\sigma = 50$ 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 6.4.

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 6.4 with $\sigma = 0.5$ cm. The results are shown in Figure 6.10. As expected from the significantly lower noise level, the calibration results are about an order of magnitude better than for the measured data.

6.5.2 Reconstruction results

For object reconstruction, we use the phaser-field back projection algorithm described in Liu et al. [Liu+19]. 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 Figure 6.11 parameters are different for each row, but constant within a row). Details about the reconstruction parameters are found in Appendix 6.B.

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. [IH20] 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 tran-

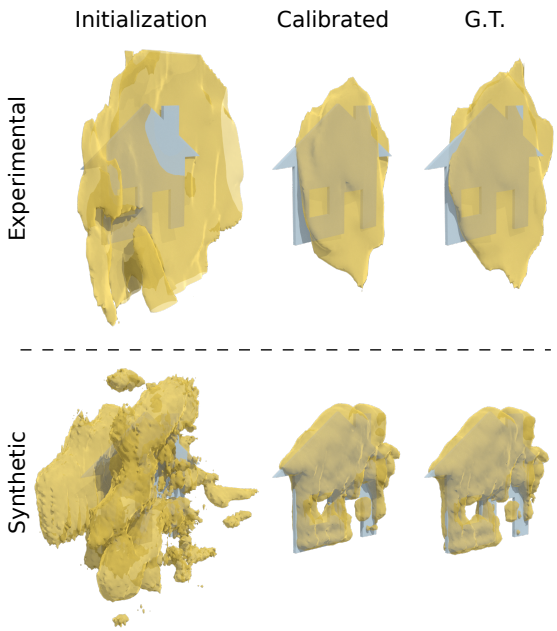


Figure 6.11: Object reconstructions obtained from different setups: *Initialization*: the ground truth setup perturbed with a noise level $\sigma = 10$ cm, *Calibrated*: the setup obtained by the presented calibration method, *G.T.*: the ground truth / measured setup. Note that in both cases the calibrated reconstruction closely resembles the ground truth reconstruction, which implies that the calibration was successful.

sient 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. [Kle+18]. 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 Figure 6.12, while Figure 6.11 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 Figure 6.10). 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.

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

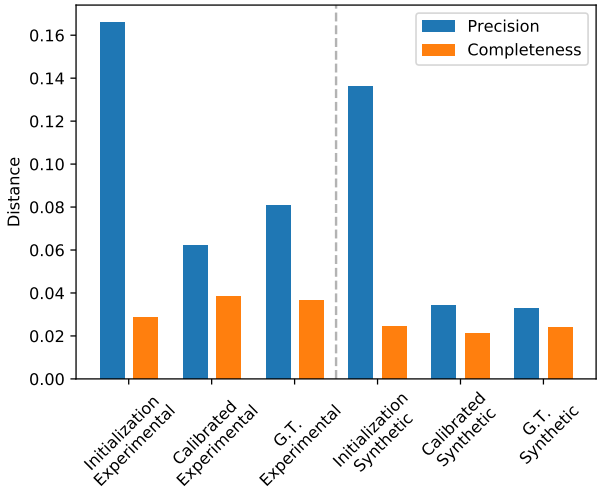


Figure 6.12: Bi-directional distance between reconstructions and reference obtained from different setups: *Initialization*: the ground truth setup perturbed with a noise level $\sigma = 10$ cm, *Calibrated*: the setup obtained by the presented calibration method, *G.T.*: the ground truth / measured 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.

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 6.5.1 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.

6.6 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 6.2).

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.

Supplemental material

6.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.

6.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 6.13 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.

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.

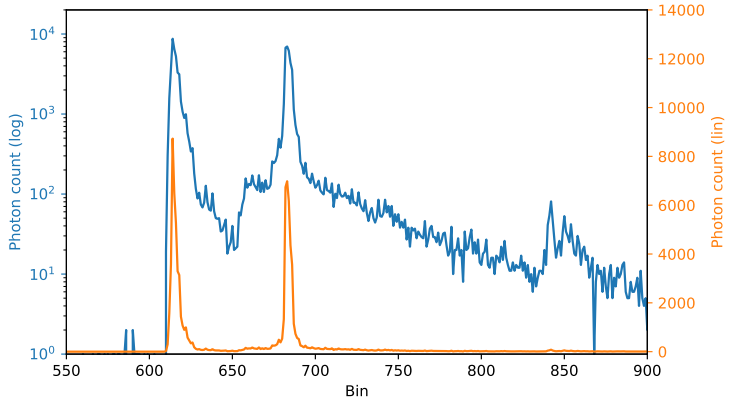
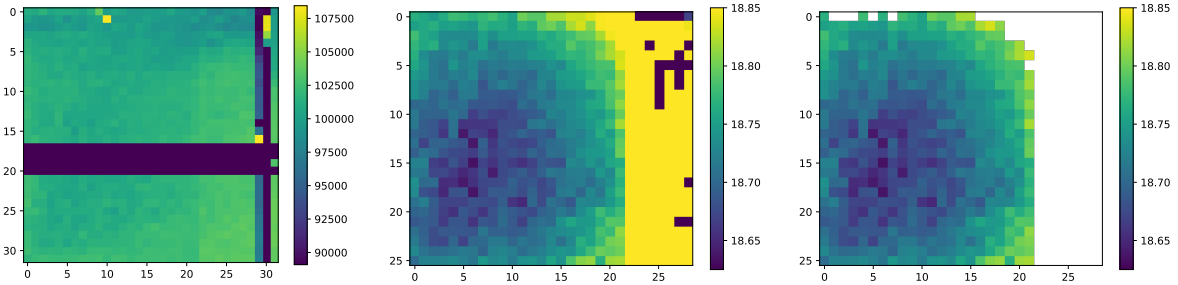


Figure 6.13: Histogram recorded by a single camera pixel. Both scales show the same data. The two peaks are well visible in the linear scale (orange). Scattering in the scene produces some background noise after the primary peak, which is visible in the logarithmic scale (blue).



a) Integrated b) Unfiltered c) Filtered

Figure 6.14: (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.

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 6.14a shows a raw image of the camera, integrated over time. The dead rows and columns 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 Figure 10.

6.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 6.14 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.

6.B Object reconstruction

We reconstruct the hidden objects in Section 6.5.2 using the phasor-field virtual wave optics algorithm by Liu et al. [Liu+19]. 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 Figure 6.13), 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 back projection 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.

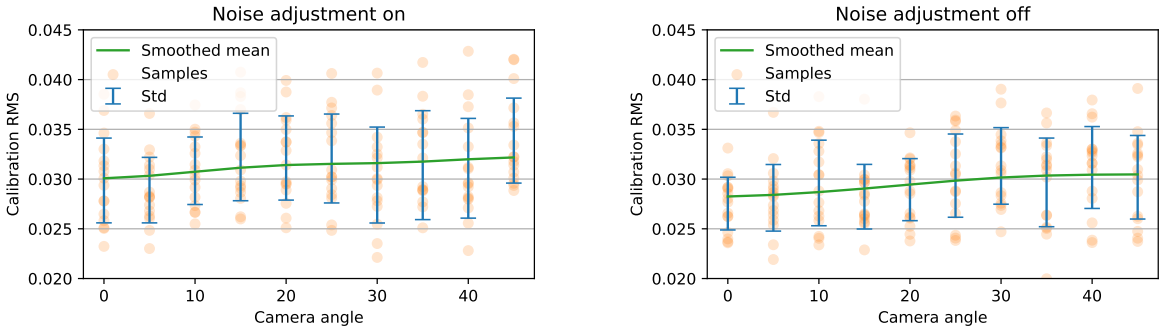


Figure 6.15: Calibration error with respect to the angle between camera and reflector wall. *Left*: The time-of-flight noise is adjusted to the pixel size. *Right*: All pixels have the same mean noise.

6.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.

6.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 6.4.1.

The camera position is rotated around the Z-axis with the rotation origin as the center of the reflector wall (see Figure 6.3) in angles between 0° (view direction normal to the wall, as in Figure 6.3) 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 6.4.1 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 Figure 6.15. 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).

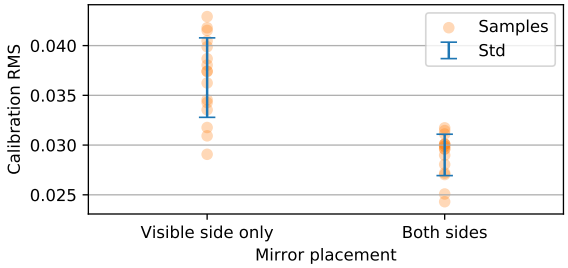


Figure 6.16: Calibration results for mirror placement constrained to the visible part of the scene and mirror placement in the visible and hidden part of the scene. In both cases 6 mirrors are used.

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.

6.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 Figure 6.3) and a setup with free mirror placement.

The setup is based on the synthetic setup from Section 6.4.1. 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 6.C.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 Figure 6.16. 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.

CHAPTER 7

Conclusion and outlook

NLoS imaging has advanced into a broad research field of global interest as becomes apparent by the large number of publications in the recent years. The overview in Chapter 3 covers only a fraction of them and more publications are expected to appear in the future. Clearly, NLoS imaging is a vibrant topic with much anticipation from the industry. Our own impact to the field is summarized as follows:

In Chapter 4 we developed a novel reconstruction approach that allows cheap real-time object tracking for the first time. It was also the first NLoS reconstruction algorithm that does not rely on transient images and we introduced analysis-by-synthesis as a reconstruction modality. In Chapter 5 we presented a benchmark for NLoS imaging problems that includes a reference data sets and domain specific evaluation metrics. With these, the multitude of existing methods become comparable and the benchmark results can help to select the right algorithm for newly designed products. In Chapter 6 we presented an universal calibration approach that allows quicker setting up a NLoS imaging system in new environments and makes them less constrained to laboratories. It works on a large variety of setups and does not require additional hardware.

7.1 Impact, limitations, and future work

During the years in which this work was conducted the field of NLoS imaging progressed and broadened substantially. Therefore we now give a brief overview of the impact, limitations, and future work for each of our publications from the point of view of the current state of the art.

The main limitations of our analysis-by-synthesis-based object tracking are a too simplistic scene model, the reduced stability if orientation reconstruction is enabled, the need for calibration, and the lack of strong convergence guarantees.

The need for calibration is shared with the majority of other publications, but it has been addressed by our later work. While the evaluation in the published paper only shows results for an intensity camera due to paper length restrictions, the tracking framework is largely hardware-agnostic and can also process transient input data. This requires the forward renderer to output time profiles which can be computed along with the intensity values with little computational overhead. During the development we used this to successfully

perform object tracking with an AMCW lidar in our laboratories. The temporal dimension can be used as additional optimization constraint which in principal is beneficial for the reconstruction quality, but overall the higher resolution and better signal-to-noise ratio of the intensity camera led to superior results. For newer transient imaging hardware this would likely not be the case and in addition our calibration method could be readily applied to a setup using them.

With respect to the scene model a follow-up paper was published by our group [IH20]. It includes occlusion in the forward model and uses an advanced Gaussian-isosurface-based optimization approach that is capable of full three-dimensional reconstruction and is especially robust to noise.

As with most numerical optimization problems, guarantees to find the optimal solution are extremely hard to achieve. Moreover the uniqueness of NLoS reconstructability has not been proven in a mathematical rigorous way (i.e. the existence of an injective function that maps scene descriptions to measurements for a given light transport model). Surfaces facing away from the relay wall are usually regarded non-reconstructible, however this might not hold true for light transport models involving higher order reflections. There has been some studies on the type of measurements required to break symmetries [Ped+17] and some general (albeit not mathematical rigorous) considerations on reconstructability [LBV19], but further research is still required.

The main limitation of our benchmark is that it its current state the reference data set does not include confocal transient images. As discussed in Section 3.1, this measurement modality has become increasingly popular after our benchmark was published. Since some reconstruction algorithms rely on confocal measurements, they can not be evaluated on our current data sets.

One of the biggest challenges in the development of the benchmark was the unification of different setups and scanning patterns. For meaningful comparison, different reconstruction algorithms should run on the same input data. This prevents the support of a large number of scanning patterns since it would inherently split the evaluation results into many subcategories. However, given the popularity of confocal scanning, providing the synthetic renderings for two different scanning patterns would be well justified and enable the comparison of most of the existing approaches.

In addition, the data set should be extended to include more complex scenes to account for the improved capabilities of newer reconstruction algorithms. Adding a background signal or other types of noise would provide means to evaluate the out-of-lab performance of different approaches. Similarly, more complex materials should be included, as in the current state only one non-diffuse material is available. The scenes also contain only a single object each which drastically lowers the amount of interreflections in the scene. In contrast to additional scanning patterns, extending the variety of scenes would not lead to a split in the evaluation but rather help the estimate the robustness of all algorithms.

Another problem is the lack of open-source implementations of published methods. Therefore it would be beneficial to create a community driven repository that implements various algorithms in a unified framework and makes them reusable for future research.

Given this limitations, our benchmark is not yet a reference on the performance of the latest work. Nonetheless the data set provided has proven to be a valuable contribution to the community and was used in publications such as Lindell et al. 2019 [LWO19]. The same

is true for the evaluation metrics [Kle+20].

The main limitations of our calibration method are the number of mirrors required in the scene and, to a lesser degree, the need for a rough initialization.

There are multiple ways in which this could be significantly improved on. So far, only time values are considered in the optimization. Including the intensity values of the peaks as an additional constraint would lead to a better conditioned optimization problem and likely improve results. These intensity values would mostly depend on the distance of the mirror (through the inverse-square law), which is already measured by the temporal measurements, and the angle between the reflector wall and the mirror (through Lambert's cosine law), which adds valuable additional information.

Furthermore it would be possible to place multiple mirrors in the scene at the same time instead of taking consecutive measurements. This would result in an additional combinatorial problem of deciding which peaks originated from which mirror as now multiple peaks appear in each measurement. However, the number of total mirrors is still low and a rough initialization does already exist.

Lastly, augmenting the mirrors with specialized tracking patterns (such as ChArUco Corners [Gar+14]) would result in these patterns being projected onto the relay wall. If a full-field image sensor is used, features of the markers could be detected in the measurements which would reveal for each pixel, where it was reflected on the mirror plane. Since this projection is a low-dimensional homography, only very few features would need to be detected for this, which could make this approach suitable even for the relative low resolution of today's transient imaging hardware. When the projection is known, the position of the mirror could be directly determined with simple geometric reasoning. With known mirror positions, the optimization is more constrained and can be solved more robustly and faster. Our method then would also most likely work with fewer mirror positions and even less requirements for the initialization of camera and laser points (while mirror positions would not need to be initialized at all).

7.2 Closing remarks

In this thesis, we presented a number of contributions to the filed of NLoS imaging. We put a focus on solving practical problems and our work has been build upon in publications of other groups. NLoS imaging is still an emerging technology and we are excited to see the progress that will be achieved by future research and the release of commercial products that result from it.

List of Figures

List of Tables

List of Abbreviations

ADMM Alternate Direction Method of Multipliers.

AMCW Amplitude Modulated Continuous Wave.

BRDF Bidirectional Reflectance Distribution Function.

DoF Degrees of Freedom.

Lidar LIght Detection And Ranging.

NLoS Non-Line-of-Sight.

PMD Photonic Mixer Devices.

Radar RAdio Detection And Ranging.

RMS Root Mean Square.

Sonar SOund NAvigation Ranging.

SPAD Single Photon Avalanche Diode.

Surfel Surface element.

ToF Time-of-Flight.

Bibliography

[Abr78] Nils Abramson. "Light-in-flight recording by holography". In: *Optics Letters* 3.4 (Oct. 1978), pp. 121–123. DOI: 10.1364/OL.3.000121. URL: http://ol.osa.org/abstract.cfm?URI=ol-3-4-121. The Optical Society (OSA) (cit. on pp. 9, 35, 36).

[Abr83] Nils Abramson. "Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena". In: *Applied Optics* 22.2 (1983), pp. 215–232. The Optical Society (OSA) (cit. on p. 36).

[Adi+15] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. "Capturing the Human Figure Through a Wall". In: *ACM Transactions on Graphics (SIGGRAPH Asia)* 34.6 (Oct. 2015), 219:1–219:13. ISSN: 0730-0301. DOI: 10.1145/2816795.2818072. ACM (cit. on pp. 33, 36, 50).

[AGJ17] Victor Arellano, Diego Gutierrez, and Adrian Jarabo. "Fast back-projection for non-line of sight reconstruction". In: *Optics Express* 25.10 (May 2017), pp. 11574–11583. DOI: 10.1364/OE.25.011574. The Optical Society (OSA) (cit. on pp. 26, 51, 59, 65, 70, 75).

[Ahn+19] Byeongjoo Ahn, Akshat Dave, Ashok Veeraraghavan, Ioannis Gkioulekas, and Aswin C. Sankaranarayanan. "Convolutional Approximations to the General Non-Line-of-Sight Imaging Operator". In: *IEEE International Conference on Computer Vision (ICCV)* (2019) (cit. on p. 27).

[AK13] Fadel Adib and Dina Katabi. "See Through Walls with WiFi!" In: *SIGCOMM Comput. Commun. Rev.* 43.4 (Aug. 2013), pp. 75–86. ISSN: 0146-4833. DOI: 10.1145/2534169.2486039. ACM (cit. on pp. 33, 36, 50).

[AMO15] Sameer Agarwal, Keir Mierle, and Others [sic]. *Ceres Solver*. http://ceres-solver.org. 2015 (cit. on p. 40).

[And06] Pierre Andersson. "Long-range three-dimensional imaging using range-gated laser radar images". In: *SPIE Optical Engineering* (Mar. 2006). DOI: 10.1117/1.2183668. International Society for Optics and Photonics (SPIE) (cit. on p. 11).

[BA87] Peter J. Burt and Edward H. Adelson. "Readings in Computer Vision: Issues, Problems, Principles, and Paradigms". In: ed. by Martin A. Fischler and Oscar Firschein. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987. Chap. The Laplacian Pyramid As a Compact Image Code, pp. 671–679. ISBN:

0-934613-33-8. URL: http://dl.acm.org/citation.cfm?id=33517.33571 (cit. on p. 58).

[Bar+18] Manel Baradad, Vickie Ye, Adam B. Yedidia, Frédo Durand, William T. Freeman, Gregory W. Wornell, and Antonio Torralba. "Inferring Light Fields From Shadows". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018) (cit. on p. 31).

[BBC17] Samuel Burri, Claudio Bruschini, and Edoardo Charbon. "LinoSPAD: A Compact Linear SPAD Camera System with 64 FPGA-Based TDC Modules for Versatile 50 ps Resolution Time-Resolved Imaging". In: *Instruments* (2017). DOI: 10.3390/instruments1010006 (cit. on p. 12).

[Ber+12] Jacopo Bertolotti, Elbert G. van Putten, Christian Blum, Ad Lagendijk, Willem L. Vos, and Allard P. Mosk. "Non-invasive imaging through opaque scattering layers". In: *Nature* (2012). DOI: 10.1038/nature11578 (cit. on p. 33).

[BH04] Jens Busck and Henning Heiselberg. "Gated viewing and high-accuracy three-dimensional laser radar". In: *Applied Optics* 43 (2004). DOI: 10.1364/AO.43.004705. The Optical Society (OSA) (cit. on p. 11).

[Bij+20] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. "Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2020) (cit. on p. 33).

[BK19] Jeremy Boger-Lombard and Ori Katz. "Passive optical time-of-flight for non line-of-sight localization". In: *Nature Communications* (2019). DOI: 10.1038/s41467-019-11279-6 (cit. on p. 32).

[Bou+17] Katherine L. Bouman, Vickie Ye, Adam B. Yedidia, Frédo Durand, Gregory W. Wornell, Antonio Torralba, and William T. Freeman. "Turning Corners into Cameras: Principles and Methods". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 2270–2278 (cit. on pp. 31, 75).

[Bur+14] Samuel Burri, Yuki Maruyama, Xavier Michalet, Francesco Regazzoni, Claudio Bruschini, and Edoardo Charbon. "Architecture and applications of a high resolution gated SPAD image sensor". In: *Optics Express* (2014). DOI: 10.1364/OE.22.017573. The Optical Society (OSA) (cit. on p. 12).

[Bur08] John Burnet. *Early Greek Philosophy*. London: Adam and Charles Black, 1908 (cit. on p. 2).

[Bus05] Jens Busck. "Underwater 3-D optical imaging with a gated viewing laser radar". In: *SPIE Optical Engineering* (2005). DOI: 10.1117/1.2127895 (cit. on p. 33).

[But+15] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. "Non-line-of-sight imaging using a time-gated single photon avalanche diode". In: *Optics Express* 23.16 (2015), pp. 20997–21011. The Optical Society (OSA) (cit. on pp. 12, 25, 26, 36, 38, 51, 60, 65, 74, 75).

[Car+17] Piergiorgio Caramazza, Alessandro Boccolini, Daniel Buschek, Matthias B. 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". In: *arXiv preprint* (2017). arXiv: 1709.07244 (cit. on p. 51).

[Car+18] Piergiorgio Caramazza, Alessandro Boccolini, Daniel Buschek, Matthias B. Hullin, Catherine F. Higham, Robert Henderson, Roderick Murray-Smith, and Daniele Faccio. "Neural network identification of people hidden from view with a single-pixel, single-photon detector". In: *Scientific Reports* (Aug. 2018). DOI: 10.1038/s41598-018-30390-0 (cit. on pp. 28, 75, 76).

[ÇG14] Y. A. Çengel and A. J. Ghajar. *Heat and Mass Transfer: Fundamentals and Applications*. McGraw-Hill Education, 2014. ISBN: 9789814595278 (cit. on p. 46).

[Cha+17a] Susan Chan, R. E. Warburton, G. Gariepy, Y. Altmann, S. McLaughlin, J. Leach, and D. Faccio. "Fast tracking of hidden objects with single-pixel detectors". In: *Electronics Letters* 53.15 (July 2017), pp. 1005–1008. ISSN: 0013-5194. DOI: 10.1049/el.2017.0993. Institution of Engineering and Technology (cit. on pp. 51, 60).

[Cha+17b] Susan Chan, Ryan E. Warburton, Genevieve Gariepy, Jonathan Leach, and Daniele Faccio. "Non-line-of-sight tracking of people at long range". In: *Optics Express* 25.9 (May 2017), pp. 10109–10117. DOI: 10.1364/OE.25.010109. The Optical Society (OSA) (cit. on pp. 26, 30).

[Che+19] Wenzheng Chen, Simon Daneau, Fahim Mannan, and Felix Heide. "Steady-state Non-Line-of-Sight Imaging". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019) (cit. on pp. 28, 30, 75).

[Cho+20] Javier Grau Chopite, Matthias B. Hullin, Michael Wand, and Julian Iseringhausen. "Deep Non-Line-of-Sight Reconstruction". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2020). arXiv: 2001.09067 [cs.CV] (cit. on pp. 28, 75, 76).

[Dor+11] A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter. "Separating true range measurements from multi-path and scattering interference in commercial range cameras". In: *SPIE Three-Dimensional Imaging, Interaction, and Measurement* 7864 (2011), pp. 37–46. DOI: 10.1117/12.876586 (cit. on p. 12).

[Eno06] Jay M. Enoch. "History of Mirrors Dating Back 8000 Years". In: *Optometry and Vision Science* 83.10 (2006), pp. 775–781. American Academy of Optometry (cit. on p. 1).

[FC99] David D. Ferris Jr. and Nicholas C. Currie. *Survey of current technologies for through-the-wall surveillance (TWS)*. 1999. DOI: 10.1117/12.336988 (cit. on p. 50).

[Fen+88] Shechao Feng, Charles Kane, Patrick A. Lee, and A. Douglas Stone. "Correlations and fluctuations of coherent wave transmission through disordered media". In: *Physical Review Letters* 61.7 (Aug. 1988) (cit. on p. 32).

[FR88] Isaac Freund and Michael Rosenbluh. "Memory Effects in Propagation of Optical Waves through Disordered Media". In: *Physical Review Letters* 61.20 (Nov. 1988) (cit. on p. 32).

[Gar+14] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez. "Automatic generation and detection of highly reliable fiducial markers under occlusion". In: *Pattern Recognition* 47.6 (2014), pp. 2280–2292. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2014.01.005 (cit. on p. 97).

[Gar+15] Genevieve 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". In: *Nature Communications* 6 (2015) (cit. on pp. 36, 73, 75).

[Gar+16] Genevieve Gariepy, Francesco Tonolini, Robert Henderson, Jonathan Leach, and Daniele Faccio. "Detection and tracking of moving objects hidden from view". In: *Nature Photonics* 10.1 (2016) (cit. on pp. 26, 30, 35, 36, 38, 44, 51, 60).

[GBH70] Richard Gordon, Rober Bender, and Gabor T. Herman. "Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography". In: *Journal of Theoretical Biology* (Dec. 1970). DOI: 10.1016/0022-5193(70)90109-8 (cit. on p. 27).

[Gil66] Lester F. Gillespie. "Apparent Illuminance as a Function of Range in Gated, Laser Night-Viewing Systems". In: *Journal of the Optical Society of America (OSA)* 56 (1966). DOI: 10.1364/JOSA.56.000883 (cit. on p. 11).

[Gki+15] Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd Zickler. "Micron-scale Light Transport Decomposition Using Interferometry". In: *ACM Transactions on Graphics* 34.4 (July 2015), 37:1–37:14. ISSN: 0730-0301. DOI: 10.1145/2766928 (cit. on p. 75).

[Gor+84] Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg, and Bennett Battaile. "Modeling the Interaction of Light Between Diffuse Surfaces". In: *ACM SIGGRAPH Computer Graphics* 18.3 (Jan. 1984), pp. 213–222. ISSN: 0097-8930. DOI: 10.1145/964965.808601 (cit. on p. 46).

[GTJ09] K. Goda, K. K. Tsia, and B. Jalali. "Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena". In: *Nature* 458.7242 (2009), pp. 1145–1149 (cit. on p. 36).

[Gup+12] Otkrist Gupta, Thomas Willwacher, Andreas Velten, Ashok Veeraraghavan, and Ramesh Raskar. "Reconstruction of hidden 3D shapes using diffuse reflections". In: *Optics Express* 20.17 (Aug. 2012), pp. 19096–19108. DOI: 10.1364/OE.20.019096. The Optical Society (OSA) (cit. on pp. 7, 26).

[Ham08] Hamamatsu. *Guide to Streak Cameras*. 2008. URL: https://web.archive.org/web/20201129004529/https://www.hamamatsu.com/resources/pdf/sys/SHSS0006E_STREAK.pdf (cit. on p. 11).

[Has21] Hasan Ibn al-Haytham. *Book of Optics (Kitāb al-Manāẓir)*. 1011-1021 (cit. on p. 2).

[HCZ00] P. Y. Han, G. C. Cho, and X.-C. Zhang. “Time-domain transillumination of biological tissues withterahertz pulses”. In: *Optics Letters* 25.4 (2000) (cit. on p. 33).

[Hei+13] Felix Heide, Matthias B. Hullin, James Gregson, and Wolfgang Heidrich. “Low-budget Transient Imaging using Photonic Mixer Devices”. In: *ACM Transactions on Graphics (SIGGRAPH)* 32.4 (2013), 45:1–45:10 (cit. on pp. 12, 36, 75).

[Hei+14] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B. Hullin. “Diffuse Mirrors: 3D Reconstruction from Diffuse Indirect Illumination Using Inexpensive Time-of-Flight Sensors”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2014) (cit. on pp. 7, 12, 25, 29, 35, 36, 38, 51, 52, 60, 74, 75).

[Hei+18] Felix Heide, Matthew O’Toole, Kai Zhang, David B. Lindell, Steven Diamond, and Gordon Wetzstein. “Non-line-of-sight Imaging with Partial Occluders and Surface Normals”. In: *ACM Transactions on Graphics* abs/1711.07134 (2018). arXiv: 1711.07134 (cit. on pp. 51, 52).

[Hei+19] Felix Heide, Matthew O’Toole, Kai Zang, David B. Lindell, Steven Diamond, and Gordon Wetzstein. “Non-line-of-sight Imaging with Partial Occluders and Surface Normals”. In: *ACM Transactions on Graphics* 38.3 (May 2019). ISSN: 0730-0301. DOI: 10.1145/3269977 (cit. on pp. 26, 73, 75).

[Hev47] Johannes Hevelius. *Selenographia sive lunae descriptio*. Gedani: Hünefeld, 1647. DOI: 10.3931/e-rara-238 (cit. on pp. 1, 2).

[HGJ17] Quercus Hernandez, Diego Gutierrez, and Adrian Jarabo. “A Computational Model of a Single-Photon Avalanche Diode Sensor for Transient Imaging”. In: *arXiv preprint* (2017). arXiv: 1703.02635 (cit. on p. 12).

[Huy09] Du Q. Huynh. “Metrics for 3D Rotations: Comparison and Analysis”. In: *Journal of Mathematical Imaging and Vision* 35.2 (2009), pp. 155–164. DOI: 10.1007/s10851-009-0161-2 (cit. on p. 57).

[ICG86] David S. Immel, Michael F. Cohen, and Donald P. Greenberg. “A Radiosity method for Non-Diffuse Environments”. In: *Association for Computing Machinery* 20.4 (1986) (cit. on p. 15).

[IH20] Julian Iseringhausen and Matthias B. Hullin. “Non-Line-of-Sight Reconstruction using Efficient Transient Rendering”. In: *ACM Transactions on Graphics (TOG)* 39.1 (2020) (cit. on pp. 27, 73, 75, 76, 86, 96).

[Ise+17] Julian Iseringhausen, Bastian Goldlücke, Nina Pesheva, Stanimir Iliev, Alexander Wender, Martin Fuchs, and Matthias B. Hullin. “4D Imaging through Spray-On Optics”. In: *ACM Transactions on Graphics (SIGGRAPH)* 36.4 (2017). DOI: 10.1145/3072959.3073589 (cit. on p. 32).

[Jar+14] Adrian Jarabo, Julio Marco, Adolfo Mu noz, Raul Buisan, Wojciech Jarosz, and Diego Gutierrez. “A Framework for Transient Rendering”. In: *ACM Transactions on Graphics (SIGGRAPH Asia)* 33.6 (2014), p. 177 (cit. on p. 13).

[Jar+17] Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. "Recent advances in transient imaging: A computer graphics and vision perspective". In: *Visual Informatics* 1.1 (2017), pp. 65–79. ISSN: 2468-502X. DOI: 10.1016/j.visinf.2017.01.008 (cit. on pp. 9, 51, 75).

[JEH15] Kishore Jaganathan, Yonina Eldar, and Babak Hassibi. "Phase Retrieval: An Overview of Recent Developments". In: *arXiv preprint* (Oct. 2015) (cit. on p. 32).

[Jin+14] Chenfei Jin, Zitong Song, Siqi Zhang, Jianhua Zhai, and Yuan Zhao. "Look through a small hole using three laser scatterings". In: *Optics Letters* (Nov. 2014). The Optical Society (OSA) (cit. on p. 51).

[Kab76] Wolfgang Kabsch. "A solution for the best rotation to relate two sets of vectors". In: *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 32.5 (1976), pp. 922–923 (cit. on p. 80).

[Kad+13] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. "Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles". In: *ACM Transactions on Graphics (SIGGRAPH Asia)* 32.6 (2013), p. 167 (cit. on pp. 12, 36).

[Kad+16] Achuta Kadambi, Hang Zhao, Boxin Shi, and Ramesh Raskar. "Occluded Imaging with Time-of-Flight Sensors". In: *ACM Transactions on Graphics* 35.2 (Mar. 2016), 15:1–15:12. ISSN: 0730-0301. DOI: 10.1145/2836164 (cit. on pp. 7, 29, 33, 36, 38, 44, 51, 60).

[Kaj86] James T. Kajiya. "The Rendering Equation". In: *Association for Computing Machinery* 20.4 (1986) (cit. on pp. 13–15).

[Kal+93] L. L. Kalpaxis, L. M. Wang, P. Galland, X. Liang, P. P. Ho, and R. R. Alfano. "Three-dimensional temporal image reconstruction of an object hidden in highly scattering media by time-gated optical tomography". In: *Optics Letters* 18.20 (1993). DOI: 10.1364/OL.18.001691. The Optical Society (OSA) (cit. on p. 11).

[Kat+14] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations". In: *Nature Photonics* 8.10 (2014), pp. 784–790 (cit. on pp. 25, 32, 33, 35, 50).

[KBC13] Ahmed Kirmani, Arrigo Benedetti, and Philip A. Chou. "SPUMIC: Simultaneous Phase Unwrapping and Multipath Interference Cancellation in Time-of-Flight Cameras using Spectal Methods". In: *IEEE International Conference on Multimedia & Expo (ICME)* (July 2013) (cit. on p. 12).

[Kel+07] Maik Keller, Jens Orthmann, Andreas Kolb, and Valerij Peters. "A Simulation Framework for Time-Of-Flight Sensors". In: *International Symposium on Signals, Circuits and Systems* 1 (2007), pp. 1–4 (cit. on p. 13).

[Kir+09] Ahmed Kirmani, T. Hutchison, J. Davis, and R. Raskar. "Looking around the corner using transient imaging". In: *IEEE International Conference on Computer Vision (ICCV)* (2009), pp. 159–166 (cit. on pp. 7, 13, 24, 25, 49, 51, 52, 60, 73, 75).

[Kle+16] Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B. Hullin. "Tracking objects outside the line of sight using 2D intensity images". In: *Scientific Reports* 6.32491 (Aug. 2016). DOI: 10.1038/srep32491. Nature Publishing Group (cit. on pp. 3, 4, 25, 27, 35, 51–53, 60, 75, 76).

[Kle+17a] Jonathan Klein, Stefan Hartmann, Michael Weinmann, and Dominik L. Michels. "Multi-Scale Terrain Texturing using Generative Adversarial Networks". In: *IEEE Conference on Image and Vision Computing New Zealand (IVCNZ)* (2017) (cit. on p. 6).

[Kle+17b] Jonathan Klein, Christoph Peters, Martin Laurenzis, and Matthias B. Hullin. "Non-line-of-sight MoCap". In: *ACM SIGGRAPH Emerging Technologies* (2017) (cit. on p. 5).

[Kle+18] Jonathan Klein, Martin Laurenzis, Dominik L. Michels, and Matthias B. Hullin. "A Quantitative Platform for Non-Line-of-Sight Imaging Problems". In: *British Machine Vision Conference (BMVC)* (2018). URL: https://nlos.cs.uni-bonn.de/paper (cit. on pp. 3, 4, 49, 87).

[Kle+20] Jonathan Klein, Martin Laurenzis, Matthias B. Hullin, and Julian Iseringhausen. "A Calibration Scheme for Non-Line-of-Sight Imaging Setups". In: *Optics Express* (2020). DOI: 10.1364/OE.398647. The Optical Society (OSA) (cit. on pp. 4, 28, 73, 97).

[KLH16] Jonathan Klein, Martin Laurenzis, and Matthias B. Hullin. "Transient Imaging for Real-Time Tracking Around a Corner". In: *SPIE Electro-Optical Remote Sensing* 9988 (2016) (cit. on pp. 3, 5, 10, 19).

[KLH17] Jonathan Klein, Martin Laurenzis, and Matthias B. Hullin. "Wenn eine Wand kein Hindernis mehr ist". In: *photonik* (May 2017). URL: http://www.photonik.de/wenn-eine-wand-kein-hindernis-mehr-ist/150/21005/350453 (cit. on p. 5).

[KLS96] Reinhard Klein, Gunther Liebich, and Wolfgang Straßer. "Mesh Reduction with Error Control". In: *IEEE Conference on Visualization* (1996), pp. 311–318 (cit. on p. 55).

[KM17] C. R. Karanam and Y. Mostofi. "3D Through-Wall Imaging with Unmanned Aerial Vehicles Using WiFi". In: *IEEE Conference on Information Processing in Sensor Networks (IPSN)* (2017), pp. 131–142 (cit. on pp. 8, 33).

[KPM20] Jonathan Klein, Sören Pirk, and Dominik L. Michels. "Domain Adaptation with Morphologic Segmentation". In: *arXiv preprint* (2020). arXiv: 2006.09322 (cit. on p. 6).

[KS88] Avinash C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988. ISBN: 978-0898714944. URL: http://www.slaney.org/pct/pct-toc.html (cit. on p. 26).

[KSS12] Ori Katz, Eran Small, and Yaron Silberberg. "Looking around corners and through thin turbid layers in real time with scattered incoherent light". In: *Nature Photonics* 6.8 (2012), pp. 549–553 (cit. on pp. 32, 33, 35, 75).

[La +17] Marco La Manna, Fiona Kine, Eric Breitbach, Jonathan Jackson, and Andreas Velten. *Error Backprojection Algorithms for Non-Line-of-Sight Imaging*. Tech. rep. TR1850. http://digital.library.wisc.edu/1793/76968. University of Wisconsin-Madison, 2017 (cit. on p. 52).

[La +18] Marco La Manna, F. Kine, E. Breitbach, J. Jackson, T. Sultan, and Andreas Velten. "Error Backprojection Algorithms for Non-Line-of-Sight Imaging". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018), pp. 1–1. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2018.2843363 (cit. on p. 26).

[La +20] Marco La Manna, Ji-Hyun Nam, Syed Azer Reza, and Andreas Velten. "Non-line-of-sight-imaging using dynamic relay surfaces". In: *Optics Express* 28.4 (Feb. 2020), pp. 5331–5339. DOI: 10.1364/OE.383586. The Optical Society (OSA) (cit. on pp. 30, 75, 78).

[Lau+12] Martin Laurenzis, Frank Christnacher, David Monnin, and Thomas Scholz. "Investigation of range-gated imaging in scattering environments". In: *SPIE Optical Engineering* (May 2012). DOI: 10.1117/1.OE.51.6.061303 (cit. on pp. 8, 33).

[Lau+15a] Martin Laurenzis, Frank Christnacher, Jonathan Klein, Matthias B. Hullin, and Andreas Velten. "Study of single photon counting for non-line-of-sight vision". In: *SPIE* 9492 (2015), 94920K-94920K–8. DOI: 10.1117/12.2179559 (cit. on p. 5).

[Lau+15b] Martin Laurenzis, Jonathan Klein, Emmanuel Bacher, and Nicolas Metzger. "Multiple-return single-photon counting of light in flight and sensing of non-line-of-sight objects at shortwave infrared wavelengths". In: *Optics Letters* 40.20 (Oct. 2015), pp. 4815–4818. DOI: 10.1364/OL.40.004815. The Optical Society (OSA) (cit. on p. 5).

[Lau+16] Martin Laurenzis, Jonathan Klein, Emmanuel Bacher, Nicolas Metzger, and Frank Christnacher. "Sensing and reconstruction of arbitrary light-in-flight paths by a relativistic imaging approach". In: *SPIE* (2016) (cit. on p. 5).

[Lau+19] Martin Laurenzis, Jonathan Klein, Emmanuel Bacher, and Stephane Schertzer. "Approaches to solve inverse problems for optical sensing around corners". In: *SPIE Security + defense: Emerging Imaging and Sensing Technologies for Security and Defence IV* (2019) (cit. on p. 4).

[LBV19] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. "Analysis of Feature Visibility in Non-Line-Of-Sight Measurements". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019) (cit. on p. 96).

[LC87] William Lorensen and Harvey Cline. "Marching Cubes: A High Resolution 3D Surface Construction Algorithm". In: *ACM SIGGRAPH Computer Graphics* (1987), pp. 163–169. DOI: 10.1145/37401.37422 (cit. on p. 56).

[LCM07] Martin Laurenzis, Frank Christnacher, and David Monnin. "Long-range three-dimensional active imaging with superresolution depth mapping". In: *Optics Letters* 32 (2007). DOI: 10.1364/OL.32.003146. The Optical Society (OSA) (cit. on p. 11).

[Lei+19] Xin Lei, Liangyu He, Yixuan Tan, Ken Xingze Wang, Xinggang Wang, Yihan Du, Shanhui Fan, and Zongfu Yu. "Direct Object Recognition Without Line-Of-Sight Using Optical Coherence". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019) (cit. on p. 32).

[LHK15] Martin Lambers, Stefan Hoberg, and Andreas Kolb. "Simulation of Time-of-Flight Sensors for Evaluation of Chip Layout Variants". In: *IEEE Sensors Journal* (Mar. 2015). DOI: 10.1109/JSEN.2015.2409816 (cit. on p. 13).

[Liu+19] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. "Non-line-of-sight imaging using phasor-field virtual wave optics". In: *Nature* 572.7771 (2019), pp. 620–623 (cit. on pp. 29, 75, 86, 91).

[LKB16] Martin Laurenzis, Jonathan Klein, and Emmanuel Bacher. "Relativistic effects in imaging of light in flight with arbitrary paths". In: *Optics Letters* 41.9 (May 2016), pp. 2001–2004. The Optical Society (OSA) (cit. on pp. 5, 9).

[LKC17] Martin Laurenzis, Jonathan Klein, and Frank Christnacher. "Transient light imaging laser radar with advanced sensing capabilities: reconstruction of arbitrary light in flight path and sensing around a corner". In: *SPIE Laser Radar Technology and Applications* (May 2017). DOI: 10.1117/12.2261961 (cit. on p. 5).

[LT98] Peter Lindstrom and Greg Turk. "Fast and Memory Efficient Polygonal Simplification". In: *IEEE Conference on Visualization*. VIS '98 (1998), pp. 279–286 (cit. on p. 55).

[LV14] Martin Laurenzis and Andreas Velten. "Nonline-of-sight laser gated viewing of scattered photons". In: *SPIE Optical Engineering* 53.2 (2014), pp. 023102–023102. DOI: 10.1117/1.OE.53.2.023102 (cit. on pp. 11, 25, 26, 35, 36, 51, 75).

[LVK17] Martin Laurenzis, Andreas Velten, and Jonathan Klein. "Dual-mode optical sensing: three-dimensional imaging and seeing around a corner". In: *SPIE Optical Engineering* (2017) (cit. on p. 5).

[LW20] David B. Lindell and Gordon Wetzstein. "Three-dimensional imaging through scattering media based on confocal diffuse tomography". In: *Nature Communications* (2020). DOI: 10.1038/s41467-020-18346-3 (cit. on p. 33).

[LWK19] David B. Lindell, Gordon Wetzstein, and Vladlen Koltun. "Acoustic Non-Line-Of-Sight Imaging". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019) (cit. on p. 30).

[LWO19] David B. Lindell, Gordon Wetzstein, and Matthew O'Toole. "Wave-based non-line-of-sight imaging using fast f-k migration". In: *ACM Transactions on Graphics (SIGGRAPH)* 38.4 (2019), p. 116 (cit. on pp. 20, 29, 75, 78, 96).

[Mae+19] Tomohiro Maeda, Guy Satat, Tristan Swedish, Lagnojita Sinha, and Ramesh Raskar. "Recent Advances in Imaging Around Corners". In: *arXiv preprint* (2019) (cit. on p. 23).

[Mar63] Donald W. Marquardt. "An algorithm for Least-Squares Estimation of Nonlinear Parameters". In: *Journal of the society for Industrial and Applied Mathematics* 11.2 (1963), pp. 431–441. DOI: 10.1137/0111030 (cit. on p. 40).

[McC08] M. W. McCall. "Electromagnetics: from Covariance to Cloaking". In: *Applications of Mathematics in Engineering and Economics, American Institute of Physics (AIP)* 1067.1 (Nov. 2008). DOI: 10.1063/1.3030822 (cit. on p. 33).

[Met+20] Christopher A. Metzler, Felix Heide, Prasana Rangarajan, Muralidhar Madabhushi Balaji, Aparna Viswanath, Ashok Veeraraghavan, and Richard G. Baraniuk. "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging". In: *Optica* (2020). DOI: 10.1364/OPTICA.374026 (cit. on pp. 32, 75).

[MLW19] Christopher A. Metzler, David B. Lindell, and Gordon Wetzstein. "Keyhole Imaging: Non-Line-of-Sight Imaging and Tracking of Moving Objects Along a Single Optical Path at Long Standoff Distances". In: *arXiv preprint* (2019) (cit. on p. 30).

[MNK13] S. Meister, R. Nair, and D. Kondermann. "Simulation of Time-of-Flight Sensors using GlobalIllumination". In: *Vision, Modeling & Visualization* (2013) (cit. on p. 13).

[Mus+19] Gabriella Musarra, Ashley Lyons, Enrico Conca, Yoann Altmann, Frederica Villa, F. Zappa, Miles John Padgett, and D. Faccio. "Non-Line-of-Sight Three-Dimensional Imaging with a Single-Pixel Camera". In: *Physical Review Applied* 12 (1 July 2019), p. 011002. DOI: 10.1103/PhysRevApplied.12.011002 (cit. on pp. 30, 74).

[Nai+11] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala. "Single view reflectance capture using multiplexed scattering and time-of-flight imaging". In: *ACM Transactions on Graphics* 30.6 (2011), p. 171 (cit. on pp. 11, 25, 27, 60, 75, 76).

[NT09] D. Needell and J. A. Tropp. "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples". In: *Applied and Computational Harmonic Analysis* 26 (May 2009), pp. 301–321. DOI: 10.1016/j.acha.2008.07.002 (cit. on p. 26).

[OLW18] Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. "Confocal non-line-of-sight imaging based on the light-cone transform". In: *Nature* 555.25489 (2018), pp. 338–341. DOI: 10.1038/nature25489 (cit. on pp. 25, 26, 30, 52, 53, 65, 73, 75, 76).

[ON94] Michael Oren and Shree K. Nayar. "Generalization of Lambert's Reflectance Model". In: *ACM SIGGRAPH Computer Graphics*. SIGGRAPH '94 (1994), pp. 239–246. DOI: 10.1145/192161.192213 (cit. on p. 15).

[Pan+11] R. Pandharkar, A. Velten, A. Bardagjy, E. Lawson, M. Bawendi, and R. Raskar. "Estimating Motion and size of moving non-line-of-sight objects in cluttered environments". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2011), pp. 265–272. ISSN: 1063-6919. DOI: 10.1109/CVPR.2011.5995465 (cit. on pp. 7, 11, 24, 26).

[Ped+17] 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: *IEEE International Conference on Computational Photography (ICCP)* (2017), pp. 1–12 (cit. on pp. 14, 29, 30, 51, 52, 75, 96).

[Pet+15] Christoph Peters, Jonathan Klein, Matthias B. Hullin, and Reinhard Klein. "Solving Trigonometric Moment Problems for Fast Transient Imaging". In: *ACM Transactions on Graphics (SIGGRAPH Asia)* 34.6 (Nov. 2015), 220:1–220:11. DOI: 10.1145/2816795.2818103 (cit. on pp. 5, 12, 36).

[PJH16] M. Pharr, W. Jakob, and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. Third Edition. Elsevier Science, 2016. ISBN: 978-0-12800-709-9. URL: http://www.pbr-book.org/ (cit. on pp. 13, 15, 17, 54, 64).

[PPP93] Frank L. Pedrotti, Leno M. Pedrotti, and Leno S. Pedrotti. *Introduction to Optics*. Third Edition. Pearson, 1993 (cit. on p. 16).

[PSV09] Xiaochuan Pan, Emil Y. Sidky, and Michael Vannier. "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" In: *Inverse Problems* 25.12 (2009), p. 123009 (cit. on p. 36).

[Pue+13] I. Puente, H. González-Jorge, J. Martínez-Sánchez, and P. Arias. "Review of mobile mapping and surveying technologies". In: *Measurement* 46.7 (2013), pp. 2127–2145. ISSN: 0263-2241. DOI: 10.1016/j.measurement.2013.03.006 (cit. on p. 45).

[PVG19] Adithya Kumar Pediredla, Ashok Veeraraghavan, and Ioannis Gkioulekas. "Ellipsoidal Path Connections for Time-Gated Rendering". In: *ACM Transactions on Graphics (TOG)* 38.4 (July 2019). ISSN: 0730-0301. DOI: 10.1145/3306346.3323016 (cit. on p. 13).

[QM85] Franco Quercioli and Giuseppe Molesini. "White light-in-flight holography". In: *Applied Optics* 24.20 (Oct. 1985), pp. 3406–3415. DOI: 10.1364/AO.24.003406. The Optical Society (OSA) (cit. on p. 36).

[Rez+19] Syed Azer Reza, Marco La Manna, Sebastian Bauer, and Andreas Velten. "Phasor field waves: A Huygens-like light transport model for non-line-of-sight imaging applications". In: *Optics Express* (2019). DOI: 10.1364/OE.27.029380 (cit. on p. 28).

[RGH09] Justin A. Richardson, Lindsay A. Grant, and Robert K. Henderson. "A Low Dark Count Single Photon Avalanche DiodeStructure Compatible with Standard NanometerScale CMOS Technology". In: *International Image Sensor Workshop (IISW)* (2009) (cit. on p. 12).

[Ric14] Mark A. Richards. *Fundamentals of Radard Signal Processing*. 2. Edition. McGraw-Hill Education, 2014 (cit. on p. 12).

[RR96] Rémi Ronfard and Jarek Rossignac. "Full-range Approximation of Triangulated Polyhedra". In: *Computer Graphics Forum* 15 (1996), pp. 67–76 (cit. on p. 55).

[Sat+17] Guy Satat, Matthew Tancik, Otkrist Gupta, Barmak Heshmat, and Ramesh Raskar. "Object classification through scattering media with deep learning on time resolved measurement". In: *Optics Express* 25.15 (July 2017). DOI: 10.1364/OE.25.017466 (cit. on p. 33).

[SC14] Malcolm Slaney and Philip A. Chou. *Time of Flight Tracer*. Tech. rep. Microsoft Research, Nov. 2014. URL: https://www.microsoft.com/en-us/research/publication/time-of-flight-tracer/ (cit. on p. 13).

[Sch+20] Nicolas Scheiner, Florian Kraus, Fangyin Wei, Buu Phan, Fahim Mannan, Nils Appenrodt, Werner Ritter, Jürgen Dickmann, Klaus Dietmayer, Bernhard Sick, and Felix Heide. "Seeing Around Street Corners: Non-Line-of-Sight Detection and Tracking In-the-Wild Using Doppler Radar". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2020) (cit. on pp. 20, 29).

[Sch+97] Rudolf Schwarte, Zhanping Xu, Horst-Guenther Heinol, Joachim Olk, Ruediger Klein, Bernd Buxbaum, Helmut Fischer, and Juergen Schulte. "New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD)". In: *SPIE Sensors, Sensor Systems, and Sensor Data Processing* (1997). DOI: 10.1117/12.287751 (cit. on p. 12).

[Sei+19] 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". In: *IEEE International Conference on Computational Photography (ICCP)* (May 2019), pp. 1–9. DOI: 10.1109/ICCPHOT.2019.8747342 (cit. on pp. 31, 75).

[SEL11] Ove Steinvall, Magnus Elmqvist, and Håkan Larsson. "See around the corner using active imaging". In: *SPIE* 8186 (2011), pp. 818605-818605–17. DOI: 10.1117/12.893605 (cit. on p. 35).

[Sen+05] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen R. Marschner, Mark Horowitz, Marc Levoy, and Hendrik Lensch. "Dual photography". In: 24.3 (2005), pp. 745–755 (cit. on pp. 35, 50).

[She+15] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. "Phase Retrieval with Application to Optical Imaging: A contemporary overview". In: *IEEE Signal Processing Magazine* 32.3 (2015), pp. 87–109. DOI: 10.1109/MSP.2014.2352673 (cit. on p. 32).

[Shr+16] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. "Computational Imaging with Multi-Camera Time-of-Flight Systems". In: *ACM Transactions on Graphics (SIGGRAPH)* (2016) (cit. on p. 51).

[SMG19] Charles Saunders, John Murray-Bruce, and Vivek K. Goyal. "Computational periscopy with an ordinary digital camera". In: *Nature* (2019) (cit. on p. 31).

[Smi+17] Brandon M. Smith, Pratham Desai, Vishal Agarwal, and Mohit Gupta. "CoLux: Multi-Object 3D Micro-Motion Analysis Using Speckle Imaging". In: *ACM Transactions on Graphics (SIGGRAPH)* 36.4 (July 2017). DOI: 10.1145/3072959.3073607 (cit. on p. 32).

[SOG18] Brandon M. Smith, Matthew O'Toole, and Mohit Gupta. "Tracking Objects Outside the Line of Sight using Speckle Imaging". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018) (cit. on p. 32).

[SS69] H. Steingold and R. E. Strauch. "Backscatter Effects in Active Night Vision Systems". In: *Applied Optics* (1969). DOI: 10.1364/AO.8.000147. The Optical Society (OSA) (cit. on p. 11).

[SSD08] Adam Smith, James Skorupski, and James Davis. *Transient Rendering*. Tech. rep. UCSC-SOE-08-26. School of Engineering, University of California, Santa Cruz, 2008 (cit. on p. 13).

[Sta55] Thomas Stanley. *The history of philosophy*. 1655 (cit. on p. 2).

[Sta72] Orestes Stavroudis. *The optics of rays, wavefronts, and caustics*. Elsevier, 1972 (cit. on p. 29).

[STM20] STMicroelectronics. *ST has introduced a new generation of high-performance proximity and ranging sensors, based on FlightSense™ Time-of-Flight (ToF) technology.* https://web.archive.org/web/20201016134352/https://www.st.com/en/imaging-and-photonics-solutions/proximity-sensors.html. 2020 (cit. on pp. 12, 20).

[Sto78] R. H. Stolt. "Migration by Fourier transform". In: *Geophysics* 43.1 (1978), pp. 23–48. DOI: 10.1190/1.1440826 (cit. on p. 28).

[Su+16] Shuochen Su, Felix Heide, Robin Swanson, Jonathan Klein, Clara Callenberg, Matthias B. Hullin, and Wolfgang Heidrich. "Material Classification Using Raw Time-of-Flight Measurements". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016) (cit. on p. 5).

[Sum+11] A. Sume, M. Gustafsson, M. Herberthson, A. Janis, S. Nilsson, J. Rahm, and A. Orbom. "Radar Detection of Moving Targets Behind Corners". In: *Geoscience and Remote Sensing, IEEE Transactions on* 49.6 (June 2011), pp. 2259–2267. ISSN: 0196-2892. DOI: 10.1109/TGRS.2010.2096471 (cit. on p. 36).

[SZL92] William J. Schroeder, Jonathan A. Zarge, and William E. Lorensen. "Decimation of Triangle Meshes". In: *ACM SIGGRAPH Computer Graphics* 26.2 (July 1992), pp. 65–70. ISSN: 0097-8930. DOI: 10.1145/142920.134010 (cit. on p. 55).

[Thr+18] Christos Thrampoulidis, Gal Shulkind, Feihu Xu, William T. Freeman, Jeffrey H. Shapiro, Antonio Torralba, Franco N. C. Wong, and Gregory W. Wornell. "Exploiting Occlusion in Non-Line-of-Sight Active Imaging". In: *IEEE Transactions on Computational Imaging* 4.3 (2018), pp. 419–431 (cit. on pp. 31, 75).

[TQ04] H. W. Tang and X. Z. Qin. "Practical methods of optimization". In: *Dalian University of Technology Press, Dalian* (2004), pp. 138–149 (cit. on p. 78).

[Tru+19] Elena Trunz, Sebastian Merzbach, Jonathan Klein, Thomas Schulze, Michael Weinmann, and Reinhard Klein. "Inverse Procedural Modeling of Knitwear". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019). CVPR Oral, pp. 8630–8639 (cit. on p. 6).

[Tsa+17] 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". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017) (cit. on pp. 27, 51).

[TSG19] Chia-Yin Tsai, Aswin C. Sankaranarayanan, and Ioannis Gkioulekas. "Beyond Volumetric Albedo–A Surface Optimization Framework for Non-Line-Of-Sight Imaging". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 1545–1555 (cit. on pp. 18, 27, 29, 75).

[Uri83] Robert J. Urick. *Principles of Underwater Sound*. 3. Edition. McGraw-Hill Education, 1983 (cit. on p. 12).

[Vel+12] Andreas Velten, T. Willwacher, Otkrist Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar. "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging". In: *Nature Communications* 3 (2012), p. 745 (cit. on pp. 11, 25, 35, 36, 38, 49, 51, 60, 73, 75).

[Vel+13] 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". In: *ACM Transactions on Graphics* 32.4 (July 2013), 44:1–44:8. ISSN: 0730-0301. DOI: 10.1145/2461912.2461928 (cit. on pp. 9, 20, 51, 52).

[VRB11] Andreas Velten, Ramesh Raskar, and Moungi Bawendi. "Picosecond Camera for Time-of-Flight Imaging". In: *Imaging and Applied Optics* (2011), IMB4. DOI: 10.1364/ISA.2011.IMB4. Optical Society of America (OSA) (cit. on pp. 26, 35, 36).

[Wal+07] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. "Microfacet Models for Refraction through Rough Surfaces". In: *Eurographics Symposium on Rendering* (2007) (cit. on pp. 53, 55).

[Wan+91] L. Wang, P. P. Ho, C. Liu, G. Zhang, and R. R. Alfano. "Ballistic 2-D Imaging Through Scattering Walls Using an Ultrafast Optical Kerr Gate". In: *Science* 253 (June 1991), pp. 769–771. ISSN: 1095-9203. DOI: 10.1126/science.253.5021.769 (cit. on p. 33).

[War+16] 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* (2016), IT4E.2. DOI: 10.1364/ISA.2016.IT4E.2. The Optical Society (OSA) (cit. on p. 51).

[War92] Gregory J. Ward. "Measuring and Modeling Anisotropic Reflection". In: *ACM SIGGRAPH Computer Graphics* 26.2 (July 1992), pp. 265–272. ISSN: 0097-8930. DOI: 10.1145/142920.134078 (cit. on p. 15).

[Wil+95] George M. Williams Jr., Alice L. Reinheimer, C. Bruce Johnson, K. D. Wheeler, Norm D. Wodecki, Verle W. Aebi, and Kenneth A. Costello. "Back-illuminated and electron-bombarded CCD low-light-level imaging system performance". In: *SPIE Photoelectronic Detectors, Cameras, and Systems* 2551 (Sept. 1995). DOI: 10.1117/12.218632 (cit. on p. 11).

[Win+02] Gerald A. Winer, Jane E. Cottrell, Virginia Gregg, Jody S. Fournier, and Lori A. Bica. "Fundamentally misunderstanding visual perception. Adults' belief in visual emissions". In: *American Psychologist* (2002). DOI: 10.1037//0003-066x.57.6-7.417 (cit. on p. 2).

[Xin+19] Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. "A Theory of Fermat Paths for Non-Line-of-Sight Shape Reconstruction". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019) (cit. on pp. 29, 33).

[Yed+19] Adam B. Yedidia, Manel Baradad, Christos Thrampoulidis, William T. Freeman, and Gregory W. Wornell. "Using Unknown Occluders to Recover Hidden Scenes". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019) (cit. on p. 31).

[Yil01] Öz Yilmaz. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Society of Exploration Geophysicists, 2001. ISBN: 9781560801580. DOI: 10.1190/1.9781560801580 (cit. on p. 29).

[Zap+07] F. Zappa, S. Tisa, A. Tosi, and S. Cova. "Principles and features of single-photon avalanche diode arrays". In: *Sensors and Actuators A: Physical* 140 (Oct. 2007). DOI: 10.1016/j.sna.2007.06.021 (cit. on p. 12).