Low-Cost SPAD Sensing for Non-Line-Of-Sight Tracking, Material Classification and Depth Imaging

CLARA CALLENBERG, University of Bonn, Germany
ZHENG SHI, Princeton University, USA
FELIX HEIDE, Princeton University, USA
MATTHIAS B. HULLIN, University of Bonn, Germany

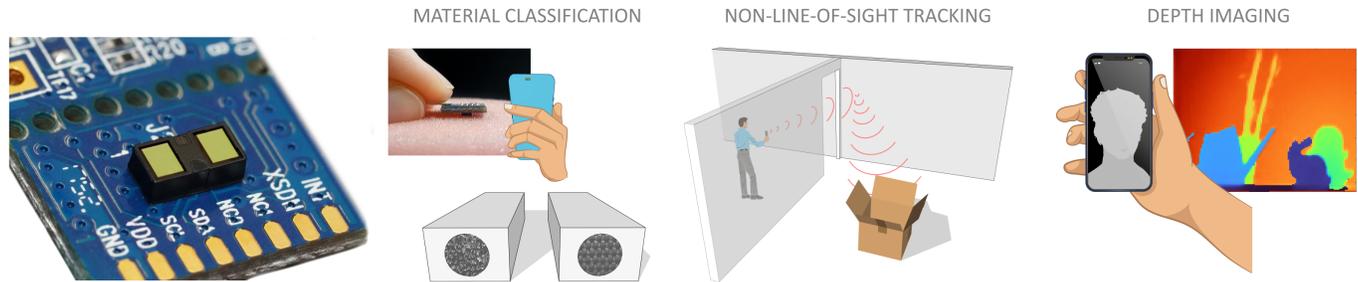MATERIAL CLASSIFICATION   NON-LINE-OF-SIGHT TRACKING   DEPTH IMAGING

Fig. 1. We propose the use of cheap, small off-the-shelf distance sensors (far left) for a variety of computational imaging and vision tasks, and demonstrate and evaluate their capabilities in emerging sensing applications like (from left to right) material classification, non-line-of-sight tracking, and depth imaging.

Time-correlated imaging is an emerging sensing modality that has been shown to enable promising application scenarios, including lidar ranging, fluorescence lifetime imaging, and even non-line-of-sight sensing. A leading technology for obtaining time-correlated light measurements are single-photon avalanche diodes (SPADs), which are extremely sensitive and capable of temporal resolution on the order of tens of picoseconds. However, the rare and expensive optical setups used by researchers have so far prohibited these novel sensing techniques from entering the mass market. Fortunately, SPADs also exist in a radically cheaper and more power-efficient version that has been widely deployed as proximity sensors in mobile devices for almost a decade. These commodity SPAD sensors can be obtained at a mere few cents per detector pixel. However, their inferior data quality and severe technical drawbacks compared to their high-end counterparts necessitate the use of additional optics and suitable processing algorithms. In this paper, we adopt an existing evaluation platform for commodity SPAD sensors, and modify it to unlock time-of-flight (ToF) histogramming and hence computational imaging. Based on this platform, we develop and demonstrate a family of hardware/software systems that, for the first time, implement applications that had so far been limited to significantly more advanced, higher-priced setups: direct ToF depth imaging, non-line-of-sight object tracking, and material classification.

Authors' addresses: Clara Callenberg, University of Bonn, Germany, callenbe@cs.uni-bonn.de; Zheng Shi, Princeton University, USA, zhengshi@princeton.edu; Felix Heide, Princeton University, USA, fheide@princeton.edu; Matthias B. Hullin, University of Bonn, Germany, hullin@cs.uni-bonn.de.

CCS Concepts: • **Computing methodologies → Computational photography**; Neural networks; • **Hardware → Sensor applications and deployments**.

1 INTRODUCTION

Time-correlated imaging, or the recording of the optical response of a scene to transient illumination, allows to analyze the temporal dimension of light transport, a feature that is not accessible in pure intensity imaging. Time-correlated optical measurements have established themselves as a valuable source of information for scene understanding [Rablau 2019] and analysis in computer graphics, computer vision, scientific imaging, healthcare and life sciences (e.g., fluorescence lifetime imaging), consumer electronics, defense, robotics and autonomous driving, and even locating hidden objects outside the direct line of sight, or "looking around a corner" [Velten et al. 2012].

The approaches available for recording time-correlated measurements are rich and varied, but most require bulky and expensive hardware and are too fragile to be used outside of lab settings. A notable exception is the emerging technology of single-photon avalanche diodes (SPADs). SPADs are single-photon sensitive devices [Zappa et al. 2007] that can be fabricated in CMOS technology [Charbon 2008] and, when combined with precise time-tagging, provide time-resolved images at picosecond resolution [Gariepy et al. 2015]. As direct time-of-flight sensors, SPADs have become the workhorse for a wide range of emerging fields [Altmann et al. 2018; Richardson et al. 2009], such as pulsed light detection and ranging
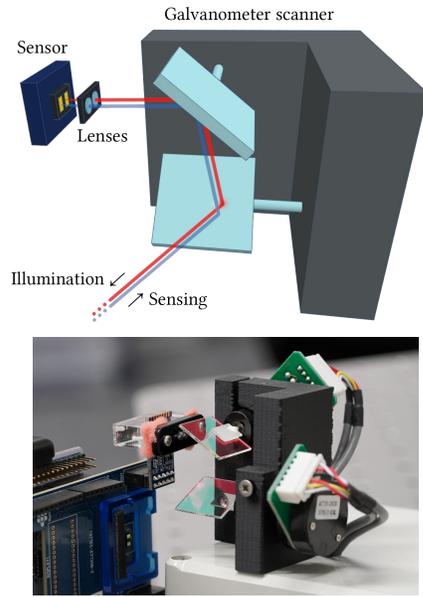
Fig. 2. Top: Setup schematic for depth and NLOS imaging: Two plano-convex lenses collimate the light emitted from the VL53L1X's light source and the SPAD sensor's field of view. Two rotatable mirrors allow deflection of the beam. Bottom: Lenses and galvanometer scanner pair in front of the sensor and light source in the lab.
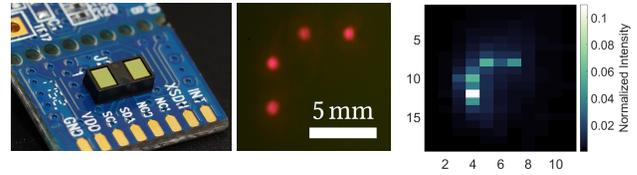

Fig. 3. Left: The VL53L1X time-of-flight sensor module by STMicroelectronics on a commercially available breakout board. Center: Profile of collimated laser beam at 0.4 m from the lens. Right: Point spread function of the full system shown in Figure 2, obtained using a small retroreflective target.

(lidar) in autonomous vehicles [Schwarz 2010], non-line-of-sight (NLOS) sensing [Chen et al. 2020; Heide et al. 2019; Lindell et al. 2019; Liu et al. 2019], as well as fluorescence lifetime imaging microscopy (FLIM) [Henderson et al. 2018] and extremely high dynamic range imaging [Ingle et al. 2019]. The sensors employed in all these works are custom-made research-grade devices that need to be combined with an ultrafast laser source, which is bulky and expensive (at least tens of thousands of US Dollars in total) and therefore out of reach for most real-life applications.

In parallel to the research on implementing novel applications using SPADS, however, the technology has already been fully democratized: small form-factor SPAD-based ranging systems are available at a price of USD 3, that integrate a pulsed laser source and a time-correlated single photon counter (TCSPC) along with an I^2C interface. Billions of consumer-grade mobile devices use them as low-cost proximity sensors [Rangwala 2020; STMicroelectronics 2019; Yoshida 2018], for instance to turn off the display when the phone is placed on the ear [Baxter 2015]. In this work, we investigate the suitability of such extremely cheap SPAD sensors for implementing computational imaging applications such as the ones discussed above. This is not a given since, despite the shared core technology, consumer-grade SPADs differ significantly from their high-end counterparts in terms of their feature set and performance. In lab settings without mass-market economies in mind, or in high-end industrial prototyping (robotics, autonomous mobility), cost is a far lesser concern and each of the components can be selected for optimum performance. Where research systems offer thousands of histogram bins with low temporal jitter on the order of a few tens of picoseconds, and sometimes detector arrays with relatively

high spatial resolutions [Henderson et al. 2018; Liu et al. 2019], consumer SPADs offer single-point or low resolution measurements with coarse temporal binning. Equally severe constraints are imposed by operation safety (consumer devices must be safe under all imaginable circumstances, even when pointed directly at the user's eye as is often the case with cell phones), and the tight power budget of mobile devices. Consumer SPAD systems therefore operate their lasers at the very minimum of what is required for close-range sensing, which is not enough for less light-efficient scenarios like non-line-of-sight sensing. On the API side, consumer SPADs are highly integrated devices that cannot be synchronized to external devices, and by default output heavily digested range data instead of raw timestamped photon events. This renders them unsuitable for most computational imaging applications.

The goal of this study is to implement typical application scenarios like range imaging, material classification, and object tracking around a corner, on a popular type of consumer SPAD (STMicroelectronics VL53L1X), and to identify possible avenues to improve the performance of such systems. In particular, these are our contributions:

- We propose to use an off-the-shelf sensor evaluation kit as a low-cost alternative to high-end SPAD sensors, and equip the board with a custom firmware to output raw photon count histograms.
- We introduce hardware add-ons such as collimating optics and galvanometer scanners to meet the needs of a selection of key applications for time-resolved imaging. We further propose reconstruction pipelines based on inverse filtering, deep learning, and other computational sensing paradigms that are capable of handling the low-resolution time-tagged measurements produced by our system.
- We validate the proposed platform for some of the most iconic application modes of time-resolved imaging, namely non-line-of-sight object tracking, material classification, and depth imaging.
- We propose cost-neutral feature additions to the sensor hardware that would greatly improve their interfacing to external hardware, and their suitability as a general-purpose sensing platform for time-resolved light transport.

At a total system cost of USD 150 including all parts, our prototype system in the most expensive configuration is two to three orders of magnitude lower in cost than existing time-tagged research instrumentation. Software and data are provided as supplementary material. We hope that our work will help unlock a wide range of fascinating sensing applications on hardware that millions of users are already carrying in their pockets.

2 RELATED WORK

This paper builds upon a substantial body of prior art, both from a technological and an application-centered point of view.

Single-photon avalanche diodes. SPADs are emerging as a promising mass-market sensor technology capable of detecting electron-hole pairs generated by single photons incident on the sensor. When combined with time-correlation electronics, these sensors allow for accurate time-tagging of such individual photon detection events, or time-correlated single-photon counting (TCSPC). SPADs are reverse-biased photodiodes that are operated well above their breakdown voltage [Burri et al. 2016]. Every photon incident on a SPAD has a probability of triggering an electron avalanche, the so-called photon detection efficiency (PDE). The resulting avalanche event is time-stamped, providing a temporal resolution of tens to hundreds of picoseconds. Compared to other single-photon photodetectors such as photomultiplier tubes and multichannel plates, SPADs are small and versatile, and are able to work at fast rates under ambient lighting conditions without requiring a high bias voltage. SPADs and avalanche photodiodes (APDs) have been successfully employed for a wide range of TCSPC applications [O'Connor 2012] in optical telecommunication, fluorescence lifetime imaging, and remote sensing systems (e.g., LIDAR). While SPAD sensors can be fabricated in CMOS technology [Burri et al. 2016], research instrumentation has been prohibitively costly at tens of thousands of USD, outside the range of most practitioners and in particular computer vision and graphics researchers without optical laboratories. In this work, we introduce a scalable research platform for time-tagged photon counting that is affordable and easy to use, since it is based on smartphone proximity sensors.

Time-of-flight (ToF), transient and depth imaging. Transient imaging captures the impulse response of light transport in a scene, thereby completely characterizing light transport as a linear time-invariant system. The idea of transient imaging was originally proposed by Abramson [1978] as "light-in-flight recording", using a holographic technique to reconstruct the propagation of a picosecond light pulse over time. With recent developments in ultrafast sensing technology, there exists now a variety of hardware options for transient imaging, turning it into an emerging imaging modality with manifold applications in computer graphics and computer vision [Jarabo et al. 2017]. Transient images and related time-of-flight techniques have been used for fast and robust depth sensing and foreground-background segmentation [Lindell et al. 2018; O'Toole et al. 2014; Peters et al. 2015] where they often outperform passive methods on scenes with complex geometry and untextured regions [Meuleman et al. 2020; Smolyanskiy et al. 2018], while not requiring a wide stereo baseline. In addition, the availability of time-of-flight histogram data allows for insightful visualization of light transport [Velten et al. 2013], material estimation [Naik et al. 2011; Su et al. 2016], fluorescence lifetime microscopy [Zickus et al. 2020], and even the reconstruction of objects beyond the direct line of sight.

Technologies used for time-resolved imaging differ widely in cost as well as individual advantages and drawbacks. At the high end of the price spectrum, streak cameras offer very high temporal resolution, but require additional instrumentation and computation for imaging two-dimensional scenes [Gao et al. 2014; Liang et al. 2017; Velten et al. 2013]. At much lower prices, amplitude modulated continuous wave (AMCW) ToF sensors, specifically intended for depth imaging at relatively high spatial resolutions, have been shown to be suitable for time-resolved imaging [Heide et al. 2013; Kadambi et al. 2013]. For a more detailed comparison of this work with amplitude modulated ToF sensors, see Section 7.

Non-line-of-sight (NLOS) tracking. Conventional cameras capture scenes that are in their direct line of sight, but computational sensing techniques have been recently proposed to reconstruct objects that are obscured from direct view, using secondary effects like indirect reflections [Velten et al. 2012] or partial shadows [Bouman et al. 2017] as an information source. Such ability to see occluded parts of the scene would have numerous obvious benefits in traffic safety, search and rescue, healthcare (endoscopy), and defense, but has yet to find its way into practical applications. While a large body of work has been dedicated to the challenge of reconstructing detailed scene geometry [Arellano et al. 2017; Buttafava et al. 2015; Grau Chopite et al. 2020; Heide et al. 2019; Iseringhausen and Hullin 2020; Liu et al. 2019; O'Toole et al. 2018; Tsai et al. 2019; Velten et al. 2012], some applications do not require a full 3D reconstruction. Often, it could be sufficient to be able to detect objects and track their motion. Thanks to a greatly reduced number of degrees of freedom, this problem can be addressed with less detailed input data and even steady-state (intensity, no time of flight) input images under passive [Bouman et al. 2017] or active [Chen et al. 2019; Klein et al. 2016] illumination, and it has led to the first industry demonstrators to integrate robust non-line-of-sight sensing technology [Scheiner et al. 2020]. Nonetheless, with an expensive bill of components, these demonstrators are unlikely to converge to mass-market products. To our knowledge, this work marks the first reported instance of using low-budget SPAD sensors for non-line-of-sight tracking.

Material classification. Indoor and outdoor scenes for robotic or scene understanding tasks almost always contain a diverse set of materials. Being able to robustly identify materials can be beneficial in many computer vision tasks such as acquisition, object recognition and segmentation. Classifying materials based on optical measurements is still a largely unsolved problem in computer vision, as a result of the ambiguity in appearance measurements. For example, polystyrene foam and white paper can appear very similar in conventional intensity RGB images, which makes material classification challenging based on their visual appearances. However, the interaction of light with many materials gives rise to a unique temporal point spread function (TPSF), which can be resolved with time-correlated detectors. A recent line of work [Su et al. 2016; Tanaka et al. 2017] proposes classification methods that use temporally resolved measurements to identify the materials via subsurface scattering. Comparing to reflectance-based methods that rely on single-view RGB images [Caputo et al. 2005; Liu et al. 2010; Varma and Zisserman 2008], temporally-resolved approaches are more robust to changes in illumination, and they are not as easily fooled by replicas such as printed pictures of the target materials. In this work, we demonstrate that it is, in fact, possible to achieve

extremely reliable distinction between five whitish materials using a cellphone-class SPAD sensor.

3 SYSTEM DESCRIPTION

The centerpiece of our system is the VL53L1X time-of-flight sensor module by STMicroelectronics. The 12-pin package, priced around USD 3 for large volumes, has a footprint of 15 mm^2 and integrates a 940 nm light source and a 16 × 16 SPAD array sensor with a field of view of 27° imaged by a miniature lens. Ambient light influx is reduced by an appropriate filter. From this device, we obtain 24-bin histograms of time-of-flight data with a bin width of 1.3 ns, thus spanning a range of ca. 4.7 meters from the sensor in direct view. The data can be read as one combined measurement from the sensor as a whole, or from a region of interest (ROI) that can be any rectangle of sensor pixels with a minimum size of 4×4. Read-out of individual pixels is not possible.

For our prototype, we rely on the sensor evaluation kit P-NUCLEO-53L1A1 which adds a 32-bit microcontroller and USB interface. The system hosts up to three sensor units, one premounted and two more on satellite boards that connect to pin headers to either side of the main sensor. The stock firmware reads preprocessed range data from the sensor, and forwards it to the host via a virtual serial port. We replace it with a custom firmware that reads and forwards raw photon count histograms.

For some of the applications demonstrated in this paper, we use additional optical equipment for increased flexibility of the system.

Addition 1: Glasses. The sensor by default is configured for weakly directional forward sensing. For some purposes, it is more desirable to work with narrower illumination and viewing beams. We equip the sensor with a pair of small plano-convex lenses (f = 6 mm, ⌀ = 3 mm) mounted at an adjustable distance from the exit/entrance apertures to fine-tune collimation (Figure 2). Using a bare image sensor placed at roughly 40 cm from the emitter, we characterize the profile of the laser beam as shown in Figure 2 (top right). It is clearly visible that the emitter employs four laser diodes arranged in an arc shape.

Addition 2: Galvo scanners. With the illumination and sensing beams collimated through the lens pair, we mounted a pair of gal-vanometer scanners with hot mirrors to scan the sensing path in two angular dimensions (Figure 2). We used the cheapest unbranded product "20kpps Laser Galvanometer Set" that is available through the most common global sales platforms, and equipped it with larger infrared mirrors and a custom 3D-printed mount. An Arduino micro-controller board provides the analog input signals for these scanners.

Being able to control the sensing beam like this, we characterized the system's point spread function by using a small retroreflective target (Figure 3, right).

4 MATERIAL CLASSIFICATION

We use the VL53L1X to classify different materials based on their temporal and spatial response to the illumination emitted by the device. When placing the sensor right onto the surface of a material, the infrared light from the VL53L1X light source penetrates the material, is scattered inside, and part of it is reflected back to the SPAD sensor as illustrated in Figure 4. Depending on the structure of the
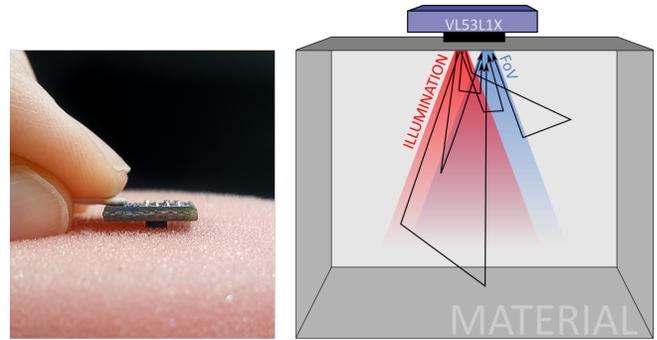

Fig. 4. For the material measurement, the VL53L1X is placed in direct contact of the sample. The active illumination penetrates the material and is scattered inside as illustrated on the right. Depending on the material's structure, the signal measured by the SPAD sensor varies spatially and temporally.

material, the signal measured by the sensor can vary temporally and spatially. By training a neural network, characteristics of different materials can be learned and they can later be distinguished by holding the sensor to an object.

For the material classification, we use contact measurements without any additional equipment. This makes this application particularly suitable for usage in small confined spaces, as well as for scenarios where the sensor is integrated into (consumer) devices like smartphones or cameras. A possible use case could be the distinction between a real finger and a dummy to improve the security of fingerprint sensors.

For the measurement, we read out the whole sensor area in ROIs of 4 × 4 pixels, which yields 16 independent ROI measurements arranged on the sensor in a 4 by 4 grid. This configuration allows the maximum number of independent ROIs on the sensor and constitutes a good compromise between captured information and acquisition time.

We record data for five different materials — foam, paper, skin, towel, and wax — by holding the sensor to the material 40 times in different positions and orientations and recording 25 histograms on all 16 ROIs. This procedure takes about 10 minutes per material. For evaluation purposes, we repeated the measurement for each material in the presence of ambient illumination in the form of a 100 W incandescent lighbulb that was placed in a distance of ∼30-40 cm from the sample, as well as for different color variants of each material (lightgreen foam, colored paper, another person's skin, darkblue towel and red wax). Due to the low temporal resolution of the SPAD sensor and the very short range of a few millimeters, most of the information is contained in the first few time bins of each histogram. We truncate the measurements to 16 time bins from the original 24, then reshape the data to matrices of size 16×16 with one spatial and one temporal dimension. Renditions of this measurement data for the five different materials are shown in Figure 5.

A two-dimensional principal component analysis of the data shows a certain clustering of the materials - almost independent of the presence of ambient light -, but poor separation between the categories. The PCA and the first two eigenvectors are shown in Figure 6.
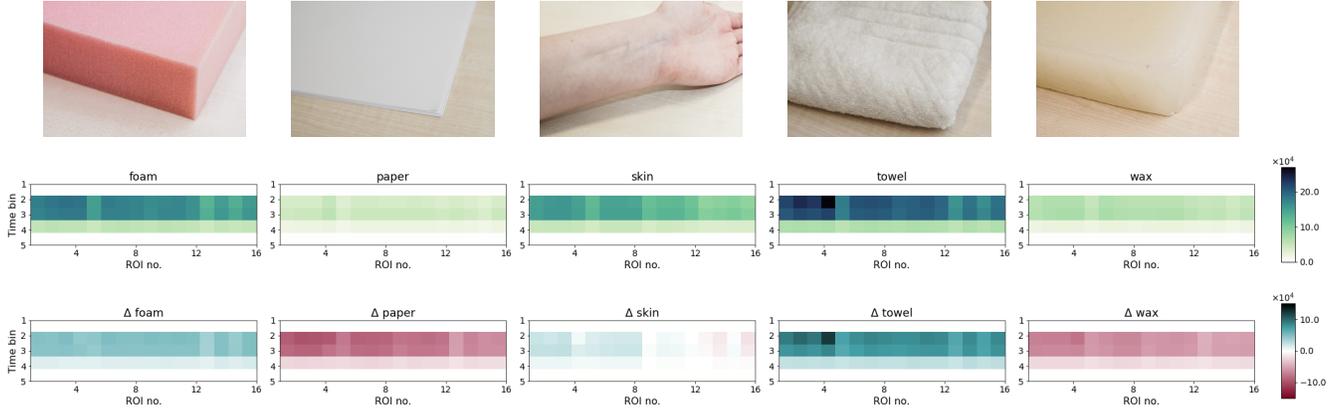
Fig. 5. Measurements for five different materials. Top: Photograph of material. Center: SPAD histogram data averaged over 1000 measurements. Bottom: Deviation of each material from the mean of all materials. *Center and bottom: Since no visible information is contained in later time bins, these plots only show bins 1 to 5.*
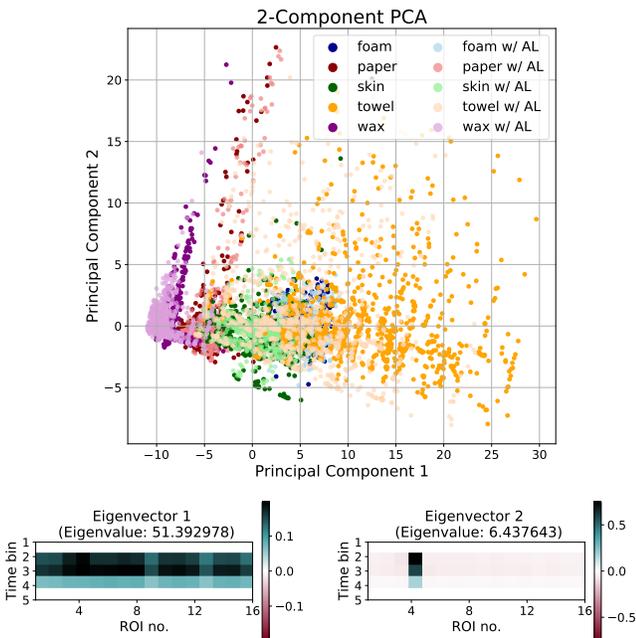

Fig. 6. Top: Two-dimensional PCA of the material data with and without ambient light (AL). Bottom: The first two eigenvectors with their corresponding eigenvalues, reshaped to the input data format. As in Figure 5, only the first five time bins are shown.
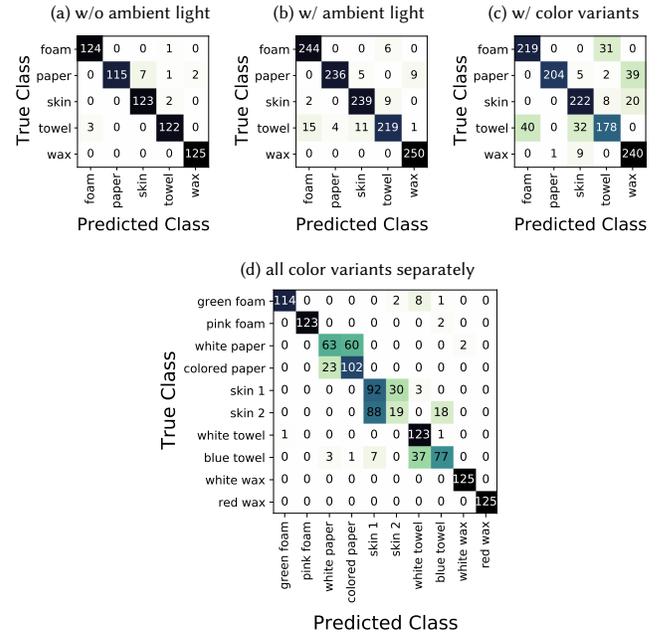

Fig. 7. Confusion matrices for the classification of five material classes on a test dataset. (a) Measurements without ambient light, (b) measurements with and without ambient light in the training and test datasets, (c) data of two color variants for each material in the training and test datasets, (d) evaluation of classification of all color variants treated as separate classes. Numbers are the absolute count of data points in the test sets. The number of training data samples was kept consistent across (a)-(c) for better comparability.

In order to classify the different materials, we train a convolutional neural network (CNN) with two 3×3 convolution layers and two fully connected layers on data from 35 of the 40 collected positions while 5 serve as test data. We perform and evaluate the training for different constellations of data for the five materials: without ambient light, with added ambient light, and with added color variants for each material. To ensure comparability in the results, the total number of measurements in the training was kept

consistent - results are shown in Table 1 and Figure 7(a)-(c). Additionally, we evaluated the performance for all material variants separately, as shown in Figure 7(d).

While the addition of ambient light has almost no effect on the performance, adding materials with different reflection properties

Table 1. Classification accuracy on the test dataset with and without ambient light, as well as with color variants of each material.

	foam	paper	skin	towel	wax
w/o ambient light	99.2 %	92.0 %	98.4 %	97.6 %	100.00 %
w/ ambient light	97.6 %	94.4 %	95.6 %	87.6 %	100.00 %
w/ color variants	87.6 %	81.6 %	88.8 %	71.2 %	96.00 %
w/ color variants (2× training size)	99.6 %	97.6 %	97.2 %	95.2 %	100.00 %

(and slightly different structure) decreases the performance quality which can, however, be almost completely avoided by adding more training data (see Tab. 1). As can be seen in the confusion matrix in Figure 7(d), especially paper, skin and towel are mostly confused among the variants of one type of material.

While the results for training and testing on a single sensor are very good and the classification performs very well even in a live application where the sensor is held to different materials and the classification runs at interactive rates (as demonstrated in the supplemental video), it has to be considered that there is a certain hardware variation between individual copies of the sensor and that classification accuracy decreases significantly when performed on a sensor instance that has not been used for the acquisition of the training data. However, due to the relatively short time needed for the acquisition of training data and the training itself, as well as the individuality of potential use cases that probably require tailoring to the particular situation and used materials, we do not consider this to be a substantial drawback. Future work could focus on a generalization of the method that allows for the calibration of a particular sensor instance in order to make the data consistent across different devices.

5 TRACKING OBJECTS "AROUND THE CORNER"

Observing objects hidden from the direct line of sight is a common application of time-resolved imagers. We show that the VL53L1X can be used to track an object "around the corner" by illuminating a wall facing the hidden area and recording the echoing light signal that is reflected from the target object. To this end, we train a neural network to recognize the target position from the SPAD data of four measurements on the wall, as described in detail in the following subsections.

5.1 Data Acquisition and Processing

We propose two possible setups for the non-line-of-sight tracking task. In the first configuration, the bare light source and sensor are pointed at the relay wall and the four corners of the SPAD sensor are read out as ROIs of size 5×5 pixels (see Figure 8, left). The second configuration employs the "glasses" and galvo-mirror system described in section 3 to focus the illumination and SPAD view to four points on the wall (Figure 8, right). This way, better spatial separation between the probe positions can be achieved, and the luminous efficacy - and therefore the signal-to-noise ratio in the measurements - is increased.

A square-shaped piece of cardboard laminated with retroreflective film serves as the target. It is placed facing the relay wall at different

positions in the hidden volume by a Universal Robots UR10 robotic arm for optimal precision. We perform all experiments with a big (50 cm × 50 cm) and a small (30 cm × 30 cm) target.

We record the SPAD signal for 800 random target positions with 10 histograms each while averaging two consecutive measurements to mitigate intensity fluctuation in the data (training data acquisition takes about 5 hours). In the configuration using the galvo-mirrors, we additionally normalize the histogram by the total intensity collected in the first five bins, which only contain signal caused by light reflected back directly from the mirror system. This way, fluctuations and deviations caused by internal factors of the sensor and illumination system can be accurately compensated for.

Assuming a completely diffuse reflection of the illumination at the relay wall, the light intensity falls off quadratically from thereon. We correct for this effect by multiplying each bin content with the square of the corresponding bin number in order to keep the signal intensity of light reflected off the target consistent across the whole measurement volume. This has shown to greatly improve the reconstruction accuracy.

5.2 Position Reconstruction

Figure 9 shows an example measurement of four corners in the "no mirror" setup for one target position, averaged over 30 measurements to reduce noise. The direct peak from the wall is clearly visible in all four histograms, while the indirect peak position is not obvious to the naked eye. Due to the indirect peak being this low, we propose to use a neural network to determine the position of the target in the hidden space. Below, we outline four different approaches to reconstruct the target position, three of which rely on training a neural network, while the last is a 'classical' approach that does not depend on previously recorded data.

1. Direct position prediction: We train a neural network (a multilayer perceptron with five hidden layers of size 50) to directly predict the target position from the four histograms. It takes the four histograms as input and reconstructs the coordinates of the target from these.

2. Distance prediction and multilateration: This approach consists of two stages. First, for each of the four probe points, its distance to the target is estimated from the measured histogram. To do this, an MLP with five hidden layers of size 16 and a final dense output layer of size 1 is trained to yield the distance from the wall point to the target from the histogram it takes as input. In this manner, the distances of all four points on the wall to the target can be predicted and used to multilaterate the target position from the four wall points by solving a simple optimization problem using the L-BFGS-B algorithm [Byrd et al. 1995].

3. Distance prediction and multilateration with histogram shift: Before processing the histogram, we find the 'direct peak' that corresponds to the reflection on the relay wall by calculating the weighted mean (center of mass) of the histogram. We then shift the histogram such that it starts with the center of the direct peak. For sub-bin precision of this method, we upsample the histogram by a factor of ten and resubsample it after the shift. Then we proceed as described in approach 2. By making the histogram independent of the distance between the SPAD sensor and the wall and thus training
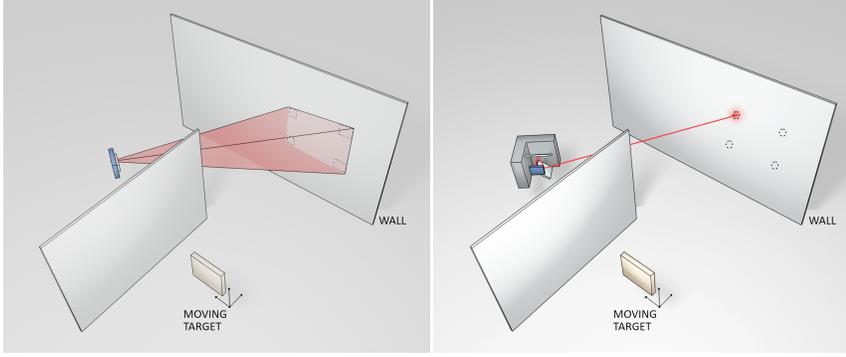
Fig. 8. Two variants of NLOS tracking setting. Left: The SPAD package's integrated light source directly illuminates the wall in front of the occluded target. The histograms from the four corners of the SPAD image are read out. Right: The illumination and SPAD view is collimated using the *glasses* and deflected to four points on the wall using a mirror galvanometer.
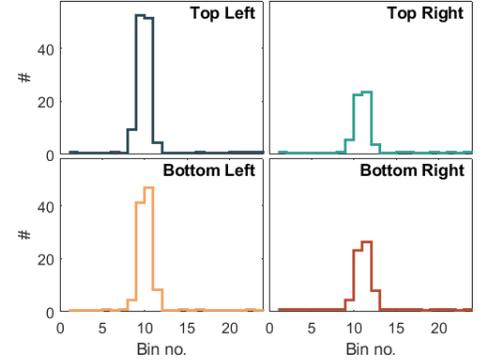
Fig. 9. Histograms for all four corners measured with the setting without mirrors for one target position, averaged over 30 measurements. The peaks from the reflection on the wall are clearly visible.

Table 2. RMSE (root mean square error) of the reconstructed target position in all spatial dimensions for all methods and target configurations. We also evaluate the reconstruction methods' ability to generalize across target sizes, where the model is trained on data recorded with the big target (if training is required) but tested on small target data, listed as "zero-shot" performance

Reconstruction Approach	w/ Mirror Performance (RMSE)			w/o Mirror Performance (RMSE)		
	Big target	Small target	Zero-shot	Big target	Small target	Zero-shot
Direct Position	0.160	0.255	0.288	0.269	0.316	0.319
Distance-Multilateration	0.299	0.304	0.307	0.397	0.421	0.423
HistShift-Distance-Multilateration	0.304	0.316	0.312	0.413	0.453	0.445
Peak-Finding-Multilateration	0.498	0.516	0.516	0.508	0.535	0.535

the neural net to predict distances to the target from any point on the wall, the tracking method becomes invariant to the positioning of the SPAD relative to the relay wall: For multilateration, the position of the probe points can be calculated from the positions of the four direct peaks in the histograms (yielding the distance between the SPAD and the wall) and the angles between the four beam directions, which are a known system property.

4. Peak finding and multilateration. For a non-learning-based comparison, we determine the position of the indirect peak from the latter part of the histogram (behind the direct peak) as the weighted mean (center of mass) and use the distance between this and the direct peak – converted from bin widths to spatial distances – for multilateration of the target position. This approach does not require any training and thus generalizes across target sizes and SPAD/wall locations.

We evaluate all methods quantitatively and demonstrate their "zero-shot" ability across the different target sizes. All configurations and the corresponding reconstruction accuracies are shown in Table 2.

As is expected due to the higher light efficiency, the setup using lenses and mirrors consistently yields better results for all evaluated methods. Since the estimation of the indirect peak with classical methods is almost impossible, the 'peak finding and multilateration' approach, however, performs poorly in all configurations. The best accuracy is achieved by training a neural net to directly predict the target coordinates from the four given histograms, which we
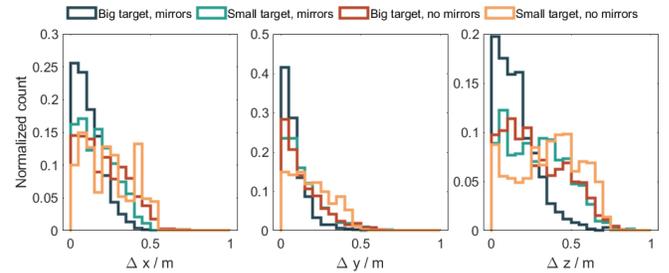

Fig. 10. Distances in all spatial dimensions between the true and predicted target position in the test dataset for the 'direct position prediction' method.

therefore consider our proposed method. It performs particularly well with the big target and mirrors, but also yields fair results with the small target and even without the additional mirrors, but just the bare VL53L1X device.

Predicting the distance between the probe points on the wall and the target results in lower accuracy, but the performance drops only slightly when additionally shifting the histogram to make the model independent of the SPAD and wall position. All methods generalize reasonably well across target sizes.

Figure 10 shows the full distribution of distances in all spatial dimensions between the true and the predicted target position in the test dataset for the proposed method, where *x* denotes the dimension to the left and right (parallel to the wall), *y* is the distance from the
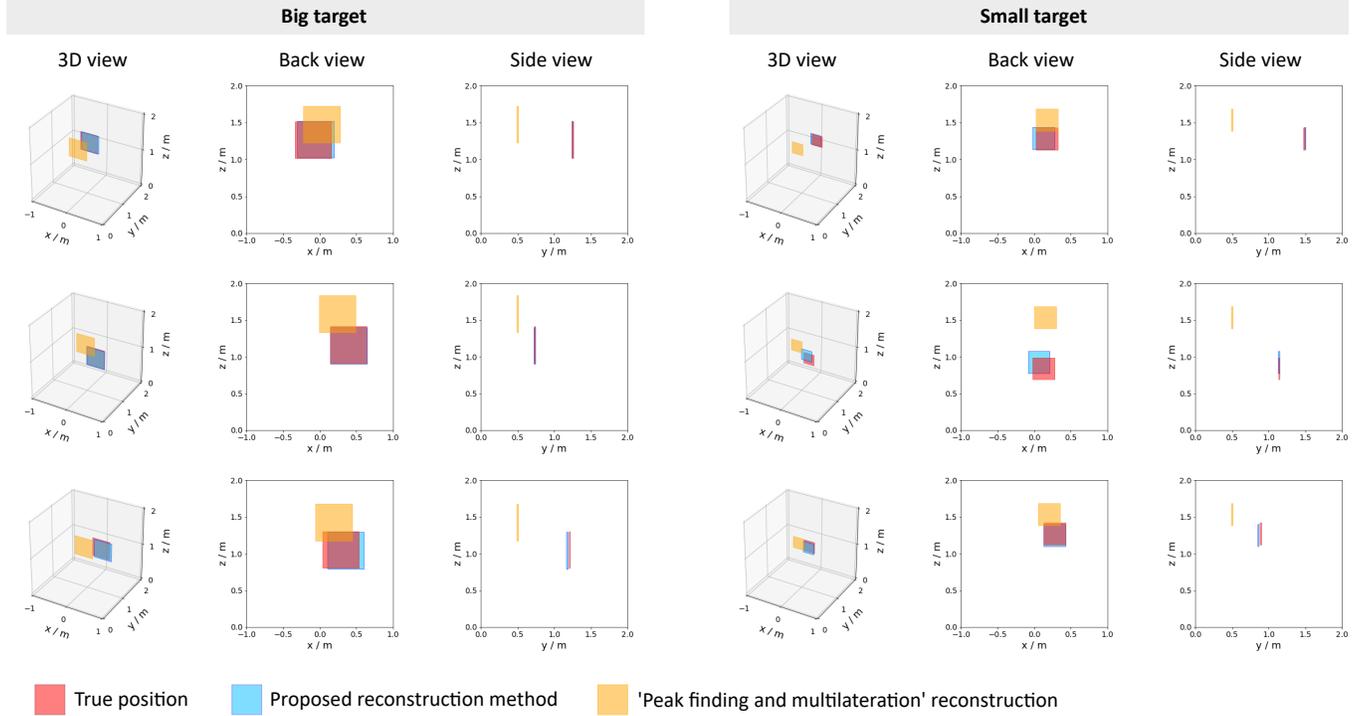
Fig. 11. Exemplary reconstructions for both target sizes. Shown are the true target position, the reconstruction using the proposed method (direct prediction of position), and the peak finding and multilateration method for comparison. For better visibility, three different projections are shown for each target position: a three-dimensional view, a view from behind the target in viewing direction to the wall, and a side view. The relay wall is located in the x-z-plane at $y = 0$. Note that the red and the blue squares overlap almost completely in some plots which is why they appear as one purple square.

wall, and z is the height from the floor. As visible in these plots, the method performs especially well at predicting the distance from the wall. The height from the floor apparently poses the greatest challenge, which could be attributed to the spacing of the probe points on the wall – due to the oblique projection of the sensor's field of view onto the wall, the left and right corner points are set wider apart than the top from the bottom, providing a larger baseline in x-direction. The height reconstruction accuracy also benefits the most from the better light efficiency through the bigger target and the collimating lenses.

A set of exemplary reconstructions with the proposed method and the 'peak finding and multilateration' approach for comparison is shown in Figure 11.

6 DEPTH IMAGING

The VL53L1X can yield a spatially resolved transient image by scanning all possible 4×4 ROIs on the 16×16 pixel sensor, which yields a 13×13 pixel measurement. This measurement, however, features a substantial blur due to the overlapping ROIs and the poor optical quality of the imaging lens. We therefore instead use the imager setup shown in Figure 2. We rotate the mirrors to 128×128 positions and capture a single all-sensor measurement per mirror position to acquire the final transient data cube (acquisition time ~30 minutes). In order to correct for internal automatic intensity corrections of the sensor, we normalize each pixel's histogram by the amount of

light that is backscattered from the lens and its mounting which is collected in the first five temporal bins of each measurement.

Additionally, using a small retroreflector we measure the point spread function of the system as shown in Figure 3 (right) and employ the fast deconvolution method by Krishnan and Fergus [2009] to deconvolve each temporal slice of the data cube with the measured PSF.

To keep intensity values relatively consistent to each other along the full depth of the measurement volume, we compensate the intensity falloff of the light travelling from the illuminated point in the scene back to the sensor by multiplying each temporal bin n_i with i^2.

Figure 12 (a)–(c) shows three different scenes scanned with our setup, where column (b) shows the raw data (each pixel's intensity summed over all time bins) and column (c) shows the data after applying the above-mentioned corrections.

From the $128 \times 128 \times 19$ 'corrected' data cube, we calculate detailed depth maps in two different ways. In the first approach, we calculate the weighted mean of the histogram to use as the given pixel's depth value d as

$$d = \frac{\sum_i i \cdot n_i}{\sum_i n_i}$$

where n_i is the intensity of bin i. This way we achieve sub-bin accuracy in the depth estimation, allowing even smooth depth gradients to be faithfully reproduced despite the large bin width that corresponds to ~40 cm (or ~20 cm in depth due to forth and back travel of

the light). Results are shown in Figure 12(d), where the second scene has been recorded in the presence of ambient light (fluorescent ceiling lights). While this method produces detailed depth images, it lacks the ability to distinguish between fore- and background contributions and, instead of separating them, yields a mixture of both. This is especially relevant for highly specular surfaces that contribute to the measurement with direct and indirect reflections, where the latter take a longer time to arrive back at the sensor, as well as depth edges of objects where pixels contain contributions from the foreground object and the background. Additionally, this method suffers from a bias in very near and very far distances due to noise and secondary reflections dragging the center of mass of the histogram to the center of the distribution. This effect is furthermore dependent on the albedo of the imaged surface, as low peaks (in comparison to background noise) in the histogram will have less impact on the resulting mean than high peaks.

To mitigate these effects, we use a second approach where we fit Gaussian functions to the histogram of each pixel. While this method takes longer to compute, it yields sharper and more reliable results as shown in Figure 12(e). Especially the highly specular surface of the bucket in the third scene is much more accurately reconstructed than with the first method. Note that both methods produce false results for objects that are beyond the measurement range, as visible in the second scene where the background behind the shelves is too far away to be measured correctly.

The Gaussian fit method also allows a variety of active tweaking of the produced depth images. For example, varying sensitivity to back- or foreground can easily be implemented by changing the threshold for which of two detected peaks to be used in the depth map. For semitransparent objects, contributions from the object itself and the background can be easily separated and selected independently of each other.

A calibration measurement of the relation between the obtained distance in units of bin numbers and the true distance is shown in Figure 13 for both methods. From the fitted linear relation, distances can be easily and accurately converted from bin numbers to meters for each method. This, however, does not account for the above-mentioned bias of the weighted mean-method towards differences in surface albedo.

In order to analyze the VL53L1X's response and robustness to different parameters like ambient light, object reflectivity, object distance and integration time, we imaged a scene with three flat targets arranged at different depths from the sensor under different ambient lighting conditions. Each target consists of five patches that appear at different brightness at the 940 nm illumination of the light source. Ambient light sources (a 940 nm LED and a 100 W incandescent light bulb) have been placed next to the VL53L1X, in a distance of ca. 90 cm from the closest target, directly illuminating the scene. Ambient daylight (third row) was indirectly illuminating the target through an open window in a few meters distance from the scene. An exemplary histogram for one pixel, averaged over 10 measurements each, with different sources of ambient light is given in Figure 14: only the 100 W incandescent light bulb and, to some extent, daylight produce levels of ambient light that noticeably impact the measured histograms. Especially the former results in a decreased signal-to-noise ratio, which we observed to be more

severe with higher total influx of light on the sensor, indicating that there is a general saturation threshold which was, however, not reached in any of our experiments. The ambient light background level is flat with no pile-up effects and can therefore be easily subtracted from the signal, however leaving an increased level of noise in the histograms. This has almost no effect on depth maps created using the Gauss fit method, but does induce a bias towards larger distances in the weighted mean method. Results for the whole scene are shown in Figure 15: Column (a) shows the level of total brightness for each measured pixel (integrated over all time bins) - note that the scale in the last row is adjusted to the significantly higher light intensity. Columns (b) and (c) show the depth maps of the scene acquired with the aforementioned methods. As illustrated in columns (d) and (e), the additional ambient light almost exclusively affects the weighted mean method (and more, the brighter the target), while in the Gauss fit depth maps, only slight deviations around depth edges are visible.

Since the VL53L1X does not allow adjustment of the integration time for each histogram, we acquired ten measurements for each ambient light situation and analyzed the variance of the depth values acquired from them as plotted in columns (f) and (g). Again, the weighted mean method shows larger variance as it is more susceptible to the increased noise level. In the Gauss fit depth maps, there is almost no variance except at depth edges where peak fitting is less precise due to contributions from back- and foreground. In conclusion, there is no significant benefit from averaging multiple histograms, and thus extended acquisition times, for this method. Especially given these limitations, we consider the weighted mean method a "quick-and-dirty" approach for a fast but noisy depth image while the Gauss fit method provides the accurate results.

7 COMPARISON TO AMCW TOF

Due to their relatively low cost in comparison to high-end SPAD setups, amplitude modulated continuous wave (AMCW) ToF devices can be considered the closest alternative technology for the tasks shown in this paper. We therefore provide a short overview of the similarities and differences of the two technologies. In general, cheap SPAD sensors like the VL53L1X are still orders of magnitude cheaper than correlation-based ToF devices, making them especially suitable for applications where multiple sensors are involved, as well as more widespread in existing hardware where they are currently used, for instance, as proximity sensors, while correlation ToF sensors are still a rather specialized feature of a small number of device models. The designated application of correlation ToF sensors is depth imaging, which they are therefore best suited for, without requiring scanning of the scene. While ToF sensors directly measure a single depth value per pixel, SPAD sensors yield a response histogram for each pixel, providing complex light transport information. Heide et al. [2013] and Kadambi et al. [2013] have shown that such temporally resolved information can be recorded using photonic mixer device (PMD) measurements, which however requires hours of post processing time. Even non-line-of-sight imaging with PMD sensors has been shown by Heide et al. [2014]; their setup uses six 250 mW laser diodes as illumination sources, while the VL53L1X features an eye-safe class 1 laser source.
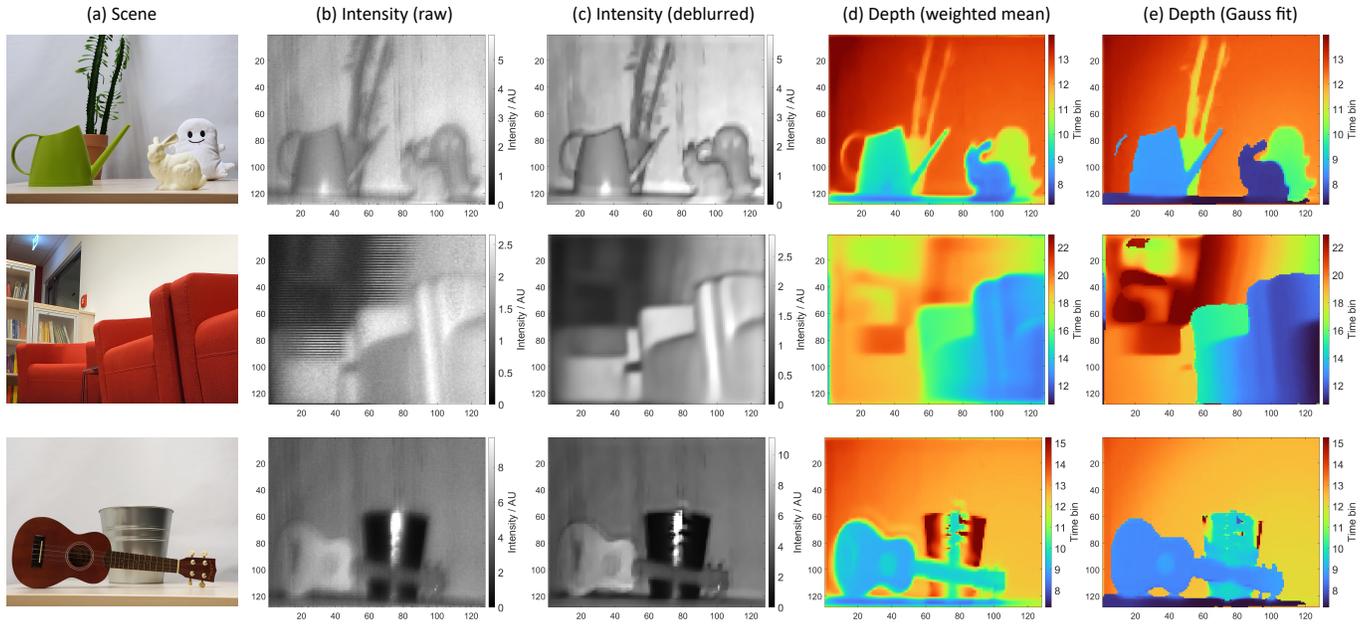
Fig. 12. Measurements taken with our scanning setup of three different real scenes shown in column (a). Column (b) shows the measured intensity of each pixel as raw data. After applying several corrections to the data, we obtain clearer and less noisy data as shown in (c). From this data, depth maps are calculated in two different ways, once as a weighted mean as shown in column (d), and once by fitting Gaussian distributions to the measured histograms (column (e)).
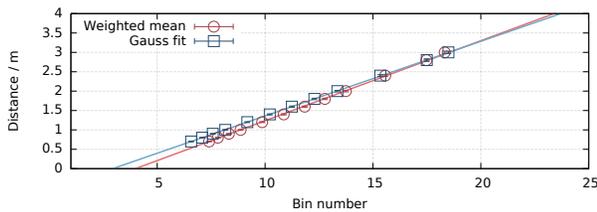

Fig. 13. Calibration of the relation between bin number and true distance for both methods used to determine the position of the depth peak.
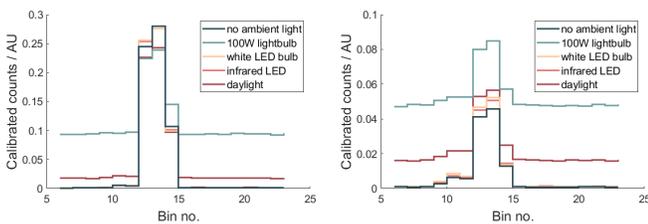

Fig. 14. Exemplary histograms for a bright (left) and a dark (right) scene pixel (after intensity calibration using the first five bins of the raw histogram) for different ambient light scenarios (no ambient light, a 100 W incandescent lightbulb, a 230 lm white LED lightbulb, a 940 nm infrared LED, and indirect daylight through an open window).

Su et al. [2016] have demonstrated material classification using raw PMD measurements on similar material samples as used in our paper, but from larger distances. Since in our approach the whole device, including the sensor and light source, has such a small form factor and can be placed directly on the material surface, spatial and

angular characteristics of the scattered light can be exploited. We achieve much better classification accuracies than Su et al. and do not require a combination of multiple samples or manual segmentation of the recorded data.

8 DISCUSSION AND FUTURE WORK

Although cheap SPADs deliver low-quality data and are narrowly optimized for short-distance, single-point ranging applications, we have been able to demonstrate that, by configuring these sensors to output raw photon counts, they can be opened to a wider range of computational sensing tasks. They are readily available in existing consumer devices. With nonstandard programming, they could instantly enable new features for a wide audience without the need for additional hardware. Furthermore, the affordability of these chips even qualifies them for use in larger quantities (arrays), which would be prohibitively expensive for most research-grade SPAD systems.

The limitations of our prototype system (an off-the-shelf sensor evaluation kit with custom firmware) are numerous. Although the sensor can capture at 60 histograms per second, overhead from API and serial communication reduces the effective capture rate to about 9 histograms per second. Despite the relatively fine grid of individual detector pixels, they cannot be read out individually but only in blocks. We assume that the manufacturers want to ensure a minimum photon count and hence signal-to-noise ratio. For our purposes, however, the main effect is a reduced spatial resolution. While the spatial resolution is potentially improvable with additional hardware [Callenberg et al. 2021], the temporal resolution of 24 bins per histogram, too, is barely enough for most applications. Additional temporal blur due to laser pulse length and detection
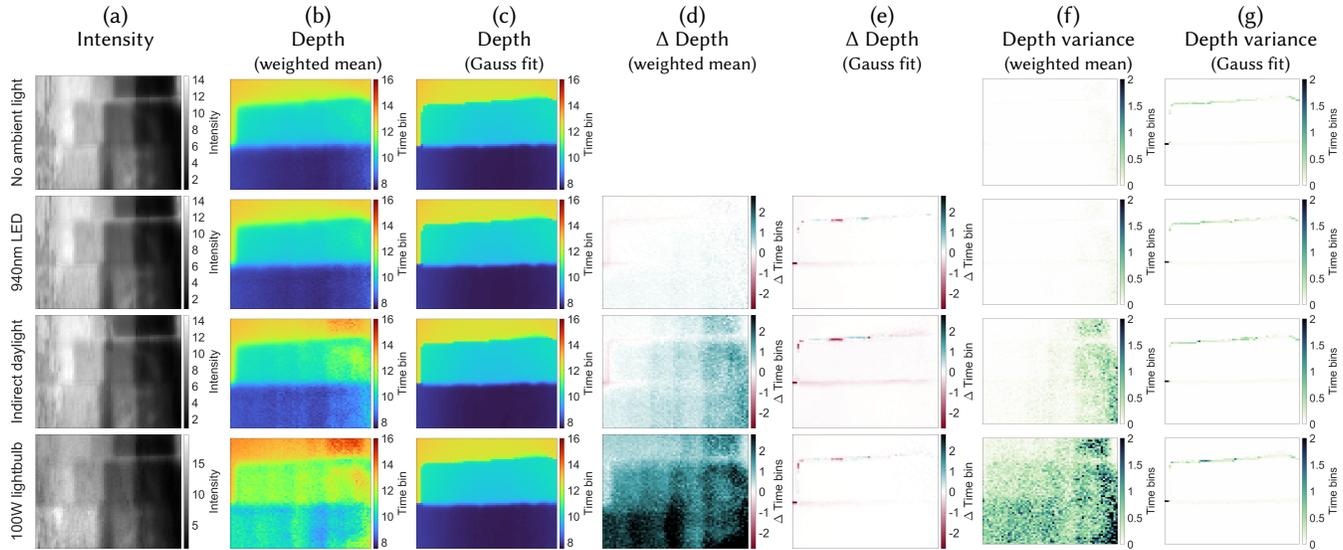
Fig. 15. Measurements of three targets in different depths, each with five patches of different reflectivity at the illumination wavelength, for different levels / sources of ambient light (no ambient light, a 940 nm infrared LED, a 100 W incandescent lightbulb, and indirect daylight through an open window). (a) Total intensity integrated over all time bins - please note the adjusted colorscale in the last row. (b) Depth map acquired with the 'weighted mean' method. (c) Depth map acquired with the 'Gauss fit' method. (d)/(e) Difference of the depth maps at different ambient light levels to the depth maps acquired without any ambient light (first row). (f)/(g) Variance of the calculated depth over 10 measurements.

jitter further reduces the capability to discriminate components with similar time of flight. Finally, the light output of our system is limited by the strict requirements of low power consumption and eye safety in mobile applications. Most experimental work on non-line-of-sight sensing, on the other hand, uses optical powers that are at least 3–4 orders of magnitude higher.

To facilitate the broader adoption of the proposed sensor platform, we hope that the sensor manufacturer alleviates these limitations by making minor technical modifications to the sensor design and its API. An API change could be official support of histogram readout, at the same level of configurability that is offered for the ranging mode. An increased number of histogram bins, even if temporal blur is the limiting factor and reduced photon count per bin leads to a higher shot noise level, would enable computational recovery of a higher-resolved signal. With the option of using region-of-interest sizes down to a single pixel, it would be possible to obtain noisy photon counts at full sensor resolution, which can be more useful for many applications than pre-binned counts. In the future, we hope that an even wider range of applications could be enabled by deeper hardware modifications to the sensor. A major difference between existing scientific time-tagged SPAD systems and the mass-market proximity sensor used in this work is the ability of syncronized triggered acquisition. The proposed system is fully free-running in the sense that it does not facilitate external triggered acquisition but only delivers repeated measurements using the internal source. As such, the user can not "start" or "stop" measurements but only read out continuous measurement streams. In the future, a trigger interface could enable the triggering of the sensor by external light sources, or vice versa. With changes only to the logic portion of the chip (introducing a time-to-digital converter per pixel), full image

data could be captured. Finally, we envision a bare-sensor version (without lens) that would allow for the use of higher-quality optics for better image quality.

With the ongoing establishment of SPAD sensors as a technology for time-resolved imaging, which is accompanied by the in-depth analysis and modelling of these sensor devices [Hernandez et al. 2017], we believe that the data quality even from cheap and simple sensors will improve rapidly and that they will open up affordable research using SPAD sensors across disciplines.

ACKNOWLEDGEMENTS

REFERENCES

Nils Abramson. 1978. Light-in-flight recording by holography. *Optics letters* 3, 4 (1978), 121–123.

Yoann Altmann, Stephen McLaughlin, Miles J Padgett, Vivek K Goyal, Alfred O Hero, and Daniele Faccio. 2018. Quantum-inspired computational imaging. *Science* 361, 6403 (2018).

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

David Patrick Baxter. 2015. Application using a single photon avalanche diode (SPAD). (June 16 2015). US Patent 9,058,081.

Katherine L Bouman, Vickie Ye, Adam B Yedidia, Frédo Durand, Gregory W Wornell, Antonio Torralba, and William T Freeman. 2017. Turning corners into cameras: Principles and methods. In *IEEE International Conference on Computer Vision (ICCV)*. 2289–2297.

Samuel Burri, Harald Homulle, Claudio Bruschini, and Edoardo Charbon. 2016. LinoSPAD: a time-resolved 256x1 CMOS SPAD line sensor system featuring 64

FPGA-based TDC channels running at up to 8.5 giga-events per second. *Proc. SPIE* 9899 (2016), 98990D–10.

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

Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing* 16, 5 (1995), 1190–1208.

C. Callenberg, A. Lyons, D. den Brok, A. Fatima, A. Turpin, V. Zickus, L. Machesky, J. Whitelaw, D. Faccio, and M. B. Hullin. 2021. Super-resolution time-resolved imaging using computational sensor fusion. *Scientific Reports* 11, 1 (Jan. 2021), 1689.

Barbara Caputo, Eric Hayman, and P Mallikarjuna. 2005. Class-specific material categorisation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Vol. 2. IEEE, 1597–1604.

Edoardo Charbon. 2008. Towards large scale CMOS single-photon detector arrays for lab-on-chip applications. *Journal of Physics D: Applied Physics* 41, 9 (2008), 094010.

Wenzheng Chen, Simon Daneau, Fahim Mannan, and Felix Heide. 2019. Steady-state non-line-of-sight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6790–6799.

Wenzheng Chen, Fangyin Wei, Kiriakos N Kutulakos, Szymon Rusinkiewicz, and Felix Heide. 2020. Learned feature embeddings for non-line-of-sight imaging and recognition. *ACM Trans. Graph.* 39, 6 (2020), 1–18.

Liang Gao, Jinyang Liang, Chiye Li, and Lihong V Wang. 2014. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature* 516, 7529 (2014), 74.

Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. 2015. Single-photon sensitive light-in-fight imaging. *Nature communications* 6 (2015), 6021.

Javier Grau Chopite, Matthias B. Hullin, Michael Wand, and Julian Iseringhausen. 2020. Deep Non-Line-of-Sight Reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Felix Heide, Matthias B. Hullin, James Gregson, and Wolfgang Heidrich. 2013. Low-Budget Transient Imaging using Photonic Mixer Devices. *ACM Trans. Graph.* 32, 4 (2013), 45:1–45:10.

Felix Heide, Matthew O'Toole, Kai Zang, David B Lindell, Steven Diamond, and Gordon Wetzstein. 2019. Non-line-of-sight imaging with partial occluders and surface normals. *ACM Trans. Graph.* 38, 3 (2019), 1–10.

Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. 2014. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3222–3229.

Robert K. Henderson, Nick Johnston, Haochang Chen, David Day Uei Li, Graham Hungerford, Richard Hirsch, David McLoskey, Philip Yip, and David J.S. Birch. 2018. A 192×128 Time Correlated Single Photon Counting Imager in 40nm CMOS Technology. *ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference* (2018), 54–57. https://doi.org/10.1109/ESSCIRC.2018.8494330

Quercus Hernandez, Diego Gutierrez, and Adrian Jarabo. 2017. A Computational Model of a Single-Photon Avalanche Diode Sensor for Transient Imaging. (2017). arXiv:physics.ins-det/1703.02635

Atul Ingle, Andreas Velten, and Mohit Gupta. 2019. High Flux Passive Imaging With Single-Photon Sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Julian Iseringhausen and Matthias Hullin. 2020. Non-line-of-sight reconstruction using efficient transient rendering. *ACM Trans. Graph.* 39, 1 (2020), 1–14.

Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. 2017. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics* 1, 1 (2017), 65–79.

A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. 2013. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. Graph.* 32, 6 (2013).

Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B Hullin. 2016. Tracking objects outside the line of sight using 2D intensity images. *Scientific reports* 6 (2016), 32491.

Dilip Krishnan and Rob Fergus. 2009. Fast Image Deconvolution using Hyper-Laplacian Priors. In *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Eds.). Curran Associates, Inc., 1033–1041.

Jinyang Liang, Cheng Ma, Liren Zhu, Yujia Chen, Liang Gao, and Lihong V Wang. 2017. Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse. *Science advances* 3, 1 (2017), e1601814.

David B Lindell, Matthew O'Toole, and Gordon Wetzstein. 2018. Single-photon 3D imaging with deep sensor fusion. *ACM Trans. Graph.* 37, 4 (2018), 113.

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

Ce Liu, Lavanya Sharan, Edward H Adelson, and Ruth Rosenholtz. 2010. Exploring features in a bayesian framework for material recognition. In *2010 ieee computer society conference on computer vision and pattern recognition*. IEEE, 239–246.

Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. 2019. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature* (2019), 1–4.

Andreas Meuleman, Seung-Hwan Baek, Felix Heide, and Min H Kim. 2020. Single-Shot Monocular RGB-D Imaging Using Uneven Double Refraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2465–2474.

Nikhil Naik, Shuang Zhao, Andreas Velten, Ramesh Raskar, and Kavita Bala. 2011. Single View Reflectance Capture Using Multiplexed Scattering and Time-of-flight Imaging. *ACM Trans. Graph.* 30, 6 (2011), 171:1–171:10.

Desmond O'Connor. 2012. *Time-correlated single photon counting*. Academic Press.

Matthew O'Toole, Felix Heide, Lei Xiao, Matthias B. Hullin, Wolfgang Heidrich, and Kiriakos N. Kutulakos. 2014. Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport. *ACM Trans. Graph.* 33, 4 (Aug. 2014).

Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. 2018. Confocal Non-line-of-sight imaging based on the light cone transform. *Nature* (2018), 338–341. Issue 555.

Christoph Peters, Jonathan Klein, Matthias B. Hullin, and Reinhard Klein. 2015. Solving Trigonometric Moment Problems for Fast Transient Imaging. *ACM Trans. Graph.* 34, 6 (Nov. 2015). https://doi.org/10.1145/2816795.2818103

Corneliu Rablau. 2019. LIDAR–A new (self-driving) vehicle for introducing optics to broader engineering and non-engineering audiences. In *Education and Training in Optics and Photonics*. Optical Society of America, 11143_138.

Sabbir Rangwala. 2020. The iPhone 12 – LiDAR At Your Fingertips. *Forbes* (12 November 2020). https://www.forbes.com/sites/sabbirrangwala/2020/11/12/the-iphone-12lidar-at-your-fingertips/

Justin A. Richardson, Lindsay A. Grant, and Robert K. Henderson. 2009. Low dark count single-photon avalanche diode structure compatible with standard nanometer scale CMOS technology. *IEEE Photonics Technology Letters* 21, 14 (2009), 1020–1022.

N. Scheiner, F. Kraus, F. Wei, B. Phan, F. Mannan, N. Appenrodt, W. Ritter, J. Dickmann, K. Dietmayer, B. Sick, et al. 2020. Seeing Around Street Corners: Non-Line-of-Sight Detection and Tracking In-the-Wild Using Doppler Radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2068–2077.

Brent Schwarz. 2010. LIDAR: Mapping the world in 3D. *Nat. Photonics* 4, 7 (2010), 1749–4885.

Nikolai Smolyanskiy, Alexey Kamenev, and Stan Birchfield. 2018. On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1007–1015.

STMicroelectronics. 2019. STMicroelectronics Ships 1 Billionth Time-of-Flight Module. (26 November 2019). https://www.st.com/content/st_com/en/about/media-center/press-item.html/t4210.html Press release.

Shuochen Su, Felix Heide, Robin Swanson, Jonathan Klein, Clara Callenberg, Matthias Hullin, and Wolfgang Heidrich. 2016. Material Classification Using Raw Time-Of-Flight Measurements. In *Proc. IEEE CVPR*.

Kenichiro Tanaka, Yasuhiro Mukaigawa, Takuya Funatomi, Hiroyuki Kubo, Yasuyuki Matsushita, and Yasushi Yagi. 2017. Material classification using frequency-and depth-dependent time-of-flight distortion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 79–88.

Chia-Yin Tsai, Aswin C Sankaranarayanan, and Ioannis Gkioulekas. 2019. Beyond Volumetric Albedo–A Surface Optimization Framework for Non-Line-Of-Sight Imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1545–1555.

Manik Varma and Andrew Zisserman. 2008. A statistical approach to material classification using image patch exemplars. *IEEE transactions on pattern analysis and machine intelligence* 31, 11 (2008), 2032–2047.

Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G. Bawendi, and Ramesh Raskar. 2012. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications* 3 (2012), 745. https://doi.org/10.1038/ncomms1747

Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. 2013. Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.* 32, 4 (2013), 44.

Junko Yoshida. 2018. ST & Apple, through Thick and Thin. *EETimes* (22 August 2018). https://www.eetimes.com/st-apple-through-thick-and-thin/

Franco Zappa, Simone Tisa, Alberto Tosi, and Sergio Cova. 2007. Principles and features of single-photon avalanche diode arrays. *Sensors and Actuators A: Physical* 140, 1 (2007), 103–112.

Vytautas Zickus, Ming-Lo Wu, Kazuhiro Morimoto, Valentin Kapitany, Areeba Fatima, Alex Turpin, Robert Insall, Jamie Whitelaw, Laura Machesky, Claudio Bruschini, et al. 2020. Fluorescence lifetime imaging with a megapixel SPAD camera and neural network lifetime estimation. *Scientific Reports* 10, 1 (2020), 1–10.