



HAL
open science

Submerged single-photon LiDAR imaging sensor used for real-time 3D scene reconstruction in scattering underwater environments

Aurora Maccarone, Kristofer Drummond, Aongus Mccarthy, Ulrich K Steinlehner, Julián Tachella, Diego Aguirre Garcia, Agata Pawlikowska, Robert A Lamb, Robert K Henderson, Stephen Mclaughlin, et al.

► To cite this version:

Aurora Maccarone, Kristofer Drummond, Aongus Mccarthy, Ulrich K Steinlehner, Julián Tachella, et al.. Submerged single-photon LiDAR imaging sensor used for real-time 3D scene reconstruction in scattering underwater environments. *Optics Express*, 2023, 31 (10), pp.16690. 10.1364/OE.487129 . hal-04297444

HAL Id: hal-04297444

<https://hal.science/hal-04297444v1>

Submitted on 29 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Submerged single-photon LiDAR imaging sensor used for real-time 3D scene reconstruction in scattering underwater environments

AURORA MACCARONE,^{1,*}  KRISTOFER DRUMMOND,^{1,2} AONGUS MCCARTHY,¹  ULRICH K. STEINLEHNER,¹ JULIAN TACHELLA,³ DIEGO AGUIRRE GARCIA,⁴ AGATA PAWLKOWSKA,² ROBERT A. LAMB,² ROBERT K. HENDERSON,⁵  STEPHEN MCLAUGHLIN,¹ YOANN ALTMANN,¹  AND GERALD S. BULLER¹

¹*School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK*

²*Leonardo MW Ltd., Crewe Road North, Edinburgh, EH5 2XS, UK*

³*Laboratoire de Physique, CNRS, ENSL, Lyon 69364, France*

⁴*Verisk, Av. de Europa, 2, 28922 Alcorcón, Madrid, Spain*

⁵*School of Engineering, University of Edinburgh, Edinburgh EH9 3FF, UK*

**A.Maccarone@hw.ac.uk*

Abstract: We demonstrate a fully submerged underwater LiDAR transceiver system based on single-photon detection technologies. The LiDAR imaging system used a silicon single-photon avalanche diode (SPAD) detector array fabricated in complementary metal-oxide semiconductor (CMOS) technology to measure photon time-of-flight using picosecond resolution time-correlated single-photon counting. The SPAD detector array was directly interfaced to a Graphics Processing Unit (GPU) for real-time image reconstruction capability. Experiments were performed with the transceiver system and target objects immersed in a water tank at a depth of 1.8 meters, with the targets placed at a stand-off distance of approximately 3 meters. The transceiver used a picosecond pulsed laser source with a central wavelength of 532 nm, operating at a repetition rate of 20 MHz and average optical power of up to 52 mW, dependent on scattering conditions. Three-dimensional imaging was demonstrated by implementing a joint surface detection and distance estimation algorithm for real-time processing and visualization, which achieved images of stationary targets with up to 7.5 attenuation lengths between the transceiver and the target. The average processing time per frame was approximately 33 ms, allowing real-time three-dimensional video demonstrations of moving targets at ten frames per second at up to 5.5 attenuation lengths between transceiver and target.

Published by Optica Publishing Group under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. Introduction

Light Detection and Ranging (LiDAR) has become the technique of choice to obtain high resolution three dimensional images in several application areas, including security and defense [1], automotive [2,3], and offshore engineering [4]. Although several detection approaches can be used in a LiDAR system, single-photon detection techniques have seen an increasing interest over the recent years due to their uniquely high optical sensitivity and excellent surface-to-surface resolution.

Typically, single-photon 3D imaging systems are based on a time-of-flight (ToF) approach and the Time-Correlated Single-Photon Counting (TCSPC) technique. The TCSPC technique

is a statistical sampling method that measures the timing of a photon event with respect to the outgoing laser pulse, typically with picosecond timing resolution [5,6]. The TCSPC approach has emerged as a good candidate for LiDAR measurements, where the time-of-flight of target return photons is used to obtain distance measurements. By repeating this ToF measurement multiple times with high repetition rate laser sources (typically > 1 MHz), the timing (and therefore distance) estimate can be significantly less than the system timing jitter. This approach has been successfully demonstrated in several challenging environments where its inherent high optical sensitivity and surface-to-surface depth resolution are highly advantageous. This includes applications such as long-range 3D imaging [7–10], 3D profiling of objects behind camouflage [11] or a scattering material [12], 3D imaging through fog or smoke [13,14], and in presence of high photon flux [15,16]. In recent years, the TCSPC technique has also been used in preliminary demonstrations of high-resolution three-dimensional single-photon imaging in highly scattering underwater environments [17]. The first demonstration was performed by using a scanning monostatic transceiver unit and an individual single-photon avalanche diode (SPAD) detector module. This transceiver scanned the area of interest with a narrow optical field of view (0.5 mrad) and measured distance and reflectivity pixel by pixel in a serial manner. The use of a narrow field of view spatially filtered out much of the forward scattering that is a major cause of the reduced spatial resolution of the image obtained in a scattering underwater environment [18]. This transceiver configuration achieved 3D profiling up to 9.2 attenuation lengths (AL) between transceiver and target using sub-milliwatt average optical powers [17,19], where 1 AL is equivalent to the distance travelled by the light that corresponds to its intensity being reduced to $1/e$ of the initial value. However, this result came at the expense of long exposure times needed to acquire the timing information for a number of pixels, typically several minutes when the acquisition time per pixel is in the order of 100 ms.

The use of SPAD detector arrays enabled shorter acquisition times for complete images and simplified optical configurations. For example, the single-photon imaging system described in [20] was based on a linear array of 16×1 SPAD detectors manufactured in custom silicon fabrication technology [21]. This configuration required the scanning of the target area in one dimension and allowed acquisition times per line as low as $1 \mu\text{s}$. Another implementation of single-photon imaging for underwater applications was based on rectangular SPAD detector arrays, as described in [22]. Here, the optical system was based on a CMOS SPAD detector planar array with a large pixel format (192×128 pixels), with TCSPC timing electronics integrated into each pixel. The large pixel format allowed the acquisition of timing information for the full optical field in parallel, greatly reducing the overall exposure time to below 1 second for stationary targets, while using an average optical power of up to 8 mW. The use of this SPAD detector array enabled the demonstration of 3D profile videos of moving targets at a visualization frame rate of 10 Hz, in clear water and in highly scattering environments up to 6.7 AL between system and target.

The mentioned demonstrations of underwater 3D imaging using SPAD detector arrays have all been performed in controlled laboratory conditions with the optical setup placed outside the water tank, and with the data analysis being performed offline, i.e. after the data acquisition. Several underwater applications require real-time processing and visualization, such as inspection, intervention (e.g. grasping), or target tracking. Different computational methods have been developed to obtain the 3D profile of the target area in real-time, i.e. in the range from microseconds to milliseconds [23]. For example, Risholm et al. [24] developed a real-time imaging system for underwater applications based on a range-gated sweeping technique. The system included a pulsed laser with active Q-switch, and the receiver was a high-resolution CMOS camera. The distance estimation algorithm was based on a peak finder of the sweep curve, and real-time capabilities were achieved by processing the data on the camera field-programmable gate array (FPGA), which allowed the authors to achieve a video frame rate of 10 Hz in real-time.

Another example of real-time three-dimensional underwater imaging is reported in [25,26], where the authors describe a system based on structured light. The system comprised a CCD camera and a continuous wave laser used in conjunction with a diffractive optical element used to form 25 laser lines on the target area. This approach enabled the 3D estimation based on triangulation to be performed for all the laser lines (i.e. the full field of view) simultaneously. In this case, the video rate was limited by the acquisition frame rate of the camera, which was up to 30 frames per second depending on the selected acquisition time for each frame.

Advances in silicon CMOS SPAD detector arrays led to timing devices with large pixel number (>65,000 [27]) and with rapid data acquisition (up to 500,000 frames per second [28]). However, such rapid data rates greatly affected the data processing, as a higher volume of data means longer processing times are required to estimate the distance and intensity profiles of targets, limiting the real-time imaging capabilities. Recently, real-time three-dimensional single-photon imaging has been achieved by developing bespoke computational models implemented on a graphics processing unit (GPU). The first example of real-time processing of multi-surface single-photon data was shown by Tachella et al. [29], who used existing point cloud denoising tools within a plug-and-play framework to achieve rapid and robust range estimation. With this approach, the authors demonstrated a video rate of 50 Hz, achieving processing times of 20 ms when using data from a 32×32 pixels InGaAs/InP SPAD detector array camera. Another example of real-time single-photon imaging is reported in [14], where the authors developed an algorithm based on advanced statistical modelling, non-linear parameter estimation, and spatio-temporal correlation to demonstrate three-dimensional imaging with processing times of 90 ms when using the same 32×32 pixels InGaAs/InP SPAD detector array camera as in [29]. Another algorithm that exploits the data statistics is described in [30], where the authors firstly identify the informative pixels containing target returns, and reject the pixels containing only noise. This approach aims to reduce the data volume, enabling real-time 3D reconstruction. The authors demonstrated the algorithm in underwater scattering environments, achieving processing times in the range of 10 ms, and 3D reconstruction up to 4.6 AL. However, in these examples, the models were demonstrated offline due to limitations in the camera output implementations, which meant that whilst the data were processed on a GPU the SPAD detector arrays were not directly interfaced with the GPU for live reconstruction.

In this paper, we present the first fully submerged underwater transceiver system based on single-photon detection, to the best of our knowledge. The system operates at the central wavelength of 532 nm, and is based on a Si-CMOS SPAD detector array interfaced to a GPU-equipped workstation for real-time, live imaging capabilities. Moreover, we demonstrated real-time implementations of two recently developed algorithms, namely the RT3D algorithm originally proposed in [29] and the ensemble algorithm recently proposed in [31] for robust distance estimation in highly scattering underwater environments. We demonstrate imaging of stationary targets up to 7.5 AL between transceiver and target, and three-dimensional videos of moving targets up to 5.5 AL at a video rate of 10 frames per second, with average processing times of approximately 33 ms for the RT3D and ensemble algorithms. This computational processing performance enabled real-time single-photon 3D imaging of moving targets with low latency.

2. System and experiment description

The SPAD detector array used for these experiments had 192×128 pixels with individual (per-pixel) timing electronics capable of measuring the arrival time of the detector events with histogram timing bins of 33 ps duration. The SPAD detector array was fabricated in 40 nm CMOS technology, and the details of the architecture of the device are reported in [32]. The fill-factor was 13% and the photon detection probability (PDP) of the SPADs was approximately 26% at the operational wavelength of 532 nm [33]. Defining the photon detection efficiency

(PDE) as the product of the fill-factor multiplied by the PDP [34], the PDE was approximately 3.4%. The output data provided by this SPAD detector array is in the form of a binary frame, which means that each individual pixel outputs in a single binary frame either a null result, or the timing information for one detected event only. Hence, a maximum of one event per pixel can be recorded for a given exposure period. For these experiments, the exposure time of each binary frame was set to 1 ms. However, a delay of approximately 1 ms between two consecutive binary frames was limited by the sensor firmware used with the SPAD array, which meant an overall acquisition binary frame rate of 500 Hz was achieved. Future work will include investigation of multi-event TDC [35], and further development is being made on the firmware of this SPAD detector array, with the aim of improving the data acquisition rate.

The underwater transceiver system is shown schematically in Fig. 1(a). The illumination was provided by a master oscillator fiber amplifier pulsed laser with central wavelength of 532 nm (VisUV, PicoQuant, Germany), operating at a repetition rate of 20 MHz. The output light was fiber-coupled into a 550 μm diameter core optical fiber, which was connected to a fiber collimation package (FCP) inside the underwater transceiver. The focal length of the collimating lens was 15 mm. The light was then directed through a N-BK7 ground glass diffuser (D) and a 75 mm focal length lens (L_1) to uniformly illuminate the target area. In these demonstrations, the size of the water tank (4 m long, 3 m wide, 1.8 m deep) limited the maximum target distance to approximately 3 m. The light scattered back from the target was collected by a 100 mm focal length lens (L_2) operating at $f/2.8$ (Canon EF 100 mm $f/2.8$ L Macro IS USM), which imaged the target return onto the sensor. In addition, two bandpass filters (BP) were used in front of the SPAD detector array to reduce the detection of ambient light which contribute to the background level. The SPAD detector array was controlled via a field-programmable gate array (FPGA) board (Opal Kelly XEM6310-LX150), which was connected to a computer via USB3.0. The computer used had an Intel Xeon CPU E5-1607 v2 @3.00 GHz processor and an NVIDIA TITAN Xp GPU processor.

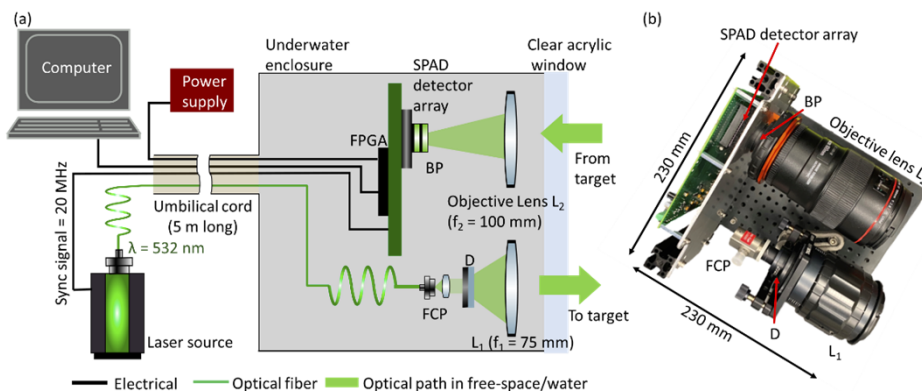


Fig. 1. (a) Schematic of the underwater transceiver. The optical setup included a fiber collimation package (FCP), an optical diffuser (D), lenses (L_1 and L_2), band pass filters (BP), and the SPAD detector array. The optical setup was placed in a watertight enclosure which was connected to the equipment outside the tank via an umbilical cord. (b) Photograph of the optical setup based on the SPAD detector array.

The optical setup and the SPAD detector array (photograph shown in Fig. 1(b)) were mounted in a watertight cylindrical enclosure (300 mm diameter by 500 mm long) made of anodized aluminum. A clear acrylic endcap acted as a window for the outgoing laser illumination and the scattered return photons. The watertight cylindrical enclosure was attached to a frame (made from aluminum extrusions) which enabled the enclosure to be positioned and kept in place

when submerged in the water tank. A 5 m long flexible hose tubing (38 mm diameter) acted as a watertight umbilical cord and contained the optical fiber and electrical cables necessary for connecting the system to the equipment outside the tank, e.g. laser source, computer, power supply for the FPGA.

Before and after each measurement, the transmittance of the water in the tank was measured to estimate the level of attenuation of the propagating medium, and consequently the number of attenuation lengths between system and target. These measurements were performed with a custom-built transmissometer shown in the schematic in Fig. 2, which measured the transmission of water at the operational wavelength of 532 nm over a fixed distance (equivalent to the distance $2d_1 + d_2$ shown in Fig. 2).

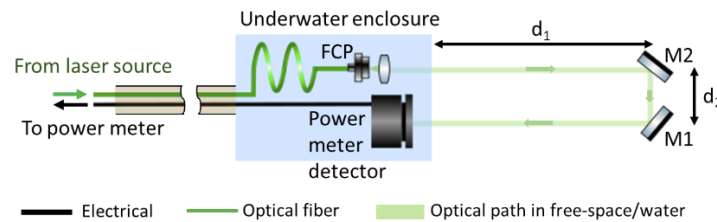


Fig. 2. Schematic of the custom-built transmissometer. The system included a fiber collimation package (FCP, focal length = 11 mm), a calibrated power meter detector (silicon photodiode detector), and two mirrors (M1 and M2).

The transmissometer optical setup was placed in a watertight clear acrylic enclosure connected to an umbilical cord, which contained a fiber coupled to the laser source and an electrical cable for the signal to the power meter detector's control module, both placed outside the water tank. The laser source used for the transmissometer was the same used for the imaging experiments. The laser light was fiber-coupled into an optical fiber with a 50 μm diameter core, and collimated using a fiber collimation package (FCP), to be directed in water through the clear acrylic optical window. Then, the light followed the path shown in green in Fig. 2, where $d_1 = 270$ mm and $d_2 = 50$ mm. The average optical power of the light reflected back to the transmissometer was measured with a silicon photodiode detector connected to the optical power meter control module. The level of scattering in water was varied by using a commercial medicine, Mucogel, which greatly affects the level of scatterer concentration in water without significantly affecting the absorption of light in water [36]. Several concentrations of Mucogel were used in these experiments, and the settling times for each concentration was measured to be more than 30 minutes, which was much longer than the acquisition times required for the measurements (up to 40 seconds for the longest scan). The water was mixed with a small portable boat propeller in order to achieve as homogeneous a concentration of Mucogel as possible.

First, a transmittance measurement was performed in air to measure the average optical power in the absence of attenuation in the transmission medium (P_A). Then, the transmissometer was lowered in water to a position as close as possible to the line of sight between the underwater transceiver and the target to measure the average optical power in the presence of obscurants (P_W). The attenuation coefficient α of the underwater environment was then calculated from the Lambert-Beer law [37]:

$$P_W = P_A e^{-\alpha(2d_1 + d_2)} \quad (1)$$

where P_A and P_W have been corrected in order to take into account the estimated Fresnel reflectivity losses at the interfaces of air, acrylic, and water. The number of attenuation lengths between the system and target was obtained from the product of the attenuation coefficient and the distance of the target in water. Each power measurements had an error of 5.5% of the reading, which meant an overall error on the attenuation length estimation equal to ± 0.4 AL.

Figure 3 shows the schematic of the overall experiment. The underwater transceiver and the transmissometer were submerged in the water tank at a depth of 1.8 m. The target was placed at a distance of approximately 3 m from the clear window of the transceiver. The transmissometer measurements were taken 1.3 m away from the target and approximately 0.3 m away from the line of sight. The main parameters of the experiment are summarized in Table 1.

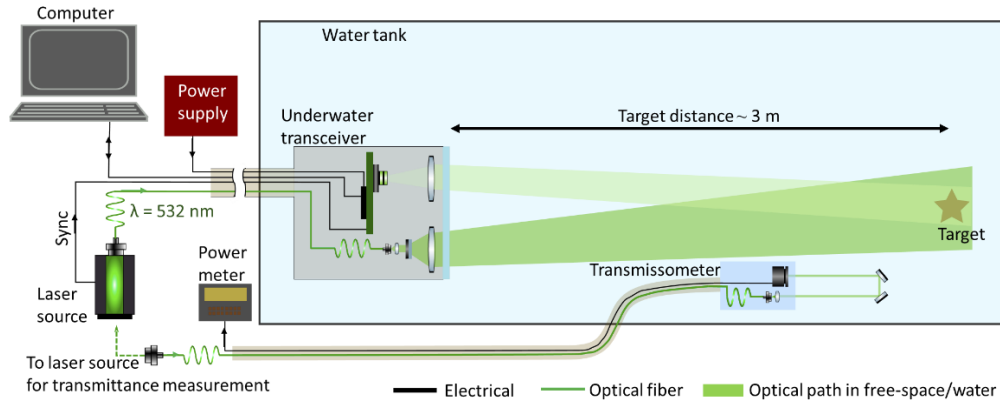


Fig. 3. Schematic of the experiment showing the submerged transceiver, external control and source modules, target position and transmissometer.

Table 1. Main parameters and components used in the experiments.

Parameter	Value/Comment
Laser system	VisUV PicoQuant
Illumination wavelength	532 nm
Laser repetition rate	20 MHz
Average optical power	up to 52 mW
Detector array	CMOS Single-Photon Avalanche-Diode (SPAD) detector array
Pixel format	192 × 128
Single-photon detection probability at operating wavelength of 532 nm	26%
Pixel fill-factor	13%
Total recorded histogram length	1540 bins
Timing bin duration	33 ps
Binary frame acquisition time	1 ms
Overall system temporal jitter	~ 290 ps FWHM
Target stand-off distance	~ 3 m
Angular Field of View	23 mrad × 17 mrad (52 × 38 mm at 3 m distance in water)
Average dark count rate	28 counts per pixel per second
Average background light (system operating in clear water)	297.5 cps per pixel

During a TCSPC measurement, the bin position of each event detected was stored in a timing histogram for the corresponding pixel, with a 33 ps timing bin width used in all the experiments reported in this paper. When using a laser pulse repetition rate of 20 MHz, this meant a histogram

of 1540 timing bins was recorded. Since each individual pixel can detect at most one photon event per exposure cycle a number of binary frames were aggregated together to build the timing histogram for each pixel for further analysis. As is typical with SPAD detector arrays, there were timing shifts between pixels, meaning relative delays between pixels were observed for the same time-of-flight measurement conditions. Therefore, individual instrumental responses were recorded for each pixel and used in the analysis described below. For example, Fig. 4(a) shows the instrumental response of an individual pixel recorded in unfiltered tap water, equivalent to a level of attenuation less than 0.5 AL between system and target. The target was a flat Lambertian reference target, with nominal reflectance of 99% (Spectralon panel, Labsphere, USA), and orientated at approximately normal incidence to the optical axis of the transmit channel. The instrumental response was obtained by aggregating together 20000 binary frames, with an acquisition time per binary frame equal to 1 ms. The temporal instrumental response of the system was estimated as the average value of the Full-Width at Half-Maximum (FWHM) for the full array, which meant approximately 290 ps FWHM. Typically, some pixels exhibit a dark count rate significantly higher than the average of the array [38]. These pixels are known as hot pixels and were identified by performing a long acquisition time measurement of 20000 binary frames in full dark conditions. Those pixels that exhibited a dark count level higher than 50% of the selected number of binary frames were classified as hot pixels and their data were removed prior to analysis.

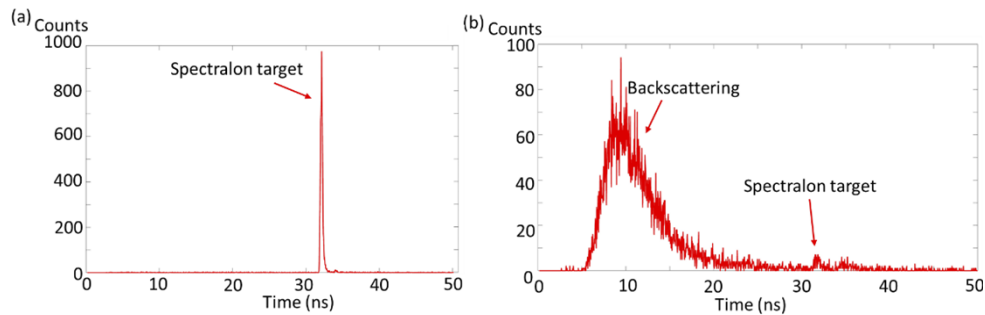


Fig. 4. Timing histograms showing the photon return from a flat Lambertian panel in different scattering conditions. Both timing histograms were obtained by aggregating 20000 binary frames, with 1 ms acquisition time per binary frame. (a) Histogram from an individual pixel of the SPAD detector array in unfiltered tap water ($< 0.5\text{ AL}$), obtained using an average optical power of 0.2 mW. (b) Histogram of the same individual pixel recorded in highly scattering water equivalent to 8.3 AL between system and target, measured by using a much higher average optical power of 50 mW. Note the different y-axis scales used in this Figure.

Figure 4(b) illustrates how the timing histogram of the same pixel can change in the presence of scattering. The measurement was performed by using the same reference target as Fig. 4(a) but in an attenuation level equivalent to 8.3 AL between system and target, with average optical power equal to 50 mW. As in Fig. 4(a), 20000 binary frames and 1 ms acquisition time per binary frame was used. In this case the collected return from the target is highly attenuated and difficult to detect, while the backscattering dominates in the histogram. Range estimation methods, for example the methods described below, when applied to the entire histograms are likely to fail due to the higher peak caused by the backscattered light leading to the wrong distance value. Hence, only the ToFs corresponding to a short timing interval centered around the expected object range are processed. By default, this interval is set to 400 bins, although we will also report results with other settings in Section 4.

3. Data analysis

3D profiles of the targets were obtained using three algorithms: the classical cross-correlation method (or maximum likelihood distance estimation); the RT3D algorithm described in [29]; and the ensemble method introduced in [31]. Two processing approaches were used to implement and compare these algorithms: offline processing and online, real-time processing. Offline processing was used to investigate 3D profiles of stationary targets, which allowed us to vary the number of binary frames used to build the histograms (typically up to 5000 binary frames), and the admissible target range, to analyze their impact on the reconstruction performance. In contrast, online real-time processing was used to investigate scenarios with moving targets and consisted of aggregating at most 50 binary frames (corresponding to 100 ms periods including read-out time) in order to avoid motion blur. Two-dimensional images of the distance and intensity profiles (where the color of the image pixels represents the target distance with respect to a reference and intensity respectively), as well as three-dimensional point clouds of the targets were obtained by collecting the distance and intensity information for the entire SPAD detector array after discarding the hot pixels. For fair comparisons, RT3D [29] was constrained to reconstruct at most one surface per pixel.

The observation model adopted by the three reconstruction methods considered in this paper is described in the next section.

3.1. Single-photon LiDAR observation model

Let us consider a set of K photon ToFs $\mathbf{y} = \{y_k\}_{k=1}^K$ where $y_k \in (0, T)$ and where T is the width of the temporal window considered. Indices representing pixel dependency are omitted to simplify notation. The probability density function (PDF) of the photon ToF y_k , is given by [39,40]:

$$f(y_k|d, \omega) = \omega h_0 \left(y_k - \frac{2d}{c} \right) + (1 - \omega) \mathcal{U}_{(0,T)}(y_k) \quad (2)$$

where d is the range of the target within the admissible range gate and c is the speed of light in the transmission medium, such that $y_d = \frac{2d}{c}$ is the characteristic ToF associated with the target. The function h_0 represents the normalized impulse response function (IRF) of the pixel and the second term $\mathcal{U}_{(0,T)}$ represents the background distribution. This needs not necessarily be uniform and can in fact be a non-uniform distribution, however for the purpose of this work, the distribution is assumed to be a uniform distribution. The variable ω represents the probability that a recorded ToF value is associated with the photon which has been emitted from the laser source, reflected off the target and returned to the photon detector. The parameter ω can also be interpreted as the signal-to-background ratio (*SBR*) of the pixel, which depends on the target reflectivity and background level. When K photons are detected, and the dead-times of the SPAD detector can be neglected, the photon ToFs are mutually independent (given d and ω) and the joint likelihood can be expressed as

$$f(\mathbf{y}|d, \omega) = \prod_{k=1}^K f(y_k|d, \omega) \quad (3)$$

3.2. Cross-correlation

This approach consisted of performing a discrete pixel-wise cross-correlation, $(g \star H)[n]$, between the recorded histogram $H[n]$ with N temporal bins, and the discrete logarithmic function, $g[m] = \log(f(y_k|d, \omega)[m])$, of the discretized probability density function $f(y_k|d, \omega)[m]$ of the photon ToF, y_k , with $M < N$ temporal bins, as:

$$(g \star H)[\tau] \triangleq \sum_{t=0}^M g[t]H[t + \tau] \quad (4)$$

where τ is the displacement value. The target is centered within the histogram, and the cross-correlation values are calculated using displacement values within the range $\tau_r =$

$[0 + \frac{M}{2}, N - \frac{M}{2}]$ in order to relieve computational miscalculations which arise when the peak of the PDF is located at or near the edges of the histogram $H[t]$. The time-of-flight, y_d , of the photons associated with the relative distance of the target for each pixel was given by the highest value of the cross-correlation:

$$y_d = \arg \max_{\tau \in \tau_r} ((g \star H)[\tau]) \quad (5)$$

which was used to calculate the target relative distance, d , for the corresponding pixel:

$$d = \frac{c y_d}{2} \quad (6)$$

3.3. Real-time 3D reconstruction using plug-and-play denoisers

This method uses the approach described by Tachella et al. [29], herein referred to as the RT3D method. This algorithm basically iterates between distance, intensity and background updates, such that each update applies a gradient step with respect to a likelihood term similar to (3) which promotes data-consistent estimates, followed by a different denoising method which promotes regular estimates. The distance update uses a scalable point cloud denoiser [41]. The intensity update uses the manifold metrics defined by the point cloud to smooth the locally the intensity and a simple denoiser based on low-pass filtering is used for the background update.

3.4. Joint surface detection and distance estimation algorithm

The third method used is the Bayesian approach as described in Drummond et al. [31] which performs surface detection and relative distance estimation collectively while rapidly providing conservative posterior uncertainties. This is achieved by discretizing ω , which can then only take a finite number of values. This allows the surface detection problem to be recast as a hypothesis testing problem (based on the marginal posterior distribution of ω) and the distance estimation problem to be reformulated as an ensemble estimation problem. Hence this method is referred to as Ensemble Method. Note that this method not only provides detection maps (indicating where surfaces are present or not) and 3D maps, but also uncertainty maps, providing confidence indicators about the quality of the estimated distance, intensity and background profiles. Here we used the algorithm with its basic settings (e.g., with $M = 30$ values for ω and $\omega_0 = 0.05$ for the probability of presence measurements).

It is important to mention that the cross-correlation and Ensemble methods do not use spatial information and process all the pixels independently, in contrast to RT3D which enforces spatially coherent solutions. Cross-correlation is considered as it is a very fast method, RT3D has been designed for fast reconstruction but does not provide uncertainty estimates about the reconstructed point clouds and the Ensemble Method has the potential to be fast (real-time compatible although not demonstrated in [31]) and provides posterior uncertainties.

3.5. Metrics

While the three methods above estimate 3D and intensity profiles, in this work we are primarily interested in the quality of the estimated 3D profiles, i.e. the quality of the reconstructed (intensity-free) point clouds, as the level of scattering increases. The metrics used to measure the quality of the estimated 3D profiles are the root mean square error (RMSE), structure similarity index measure (SSIM) [42], and multi-scale structure similarity index measure (MS-SSIM) [43]. These metrics compare the similarity between a distorted measured image, where the distortion arises due to noise, with a non-distorted base image which in this work was obtained in unfiltered tap water (i.e. in absence of scattering), and provide a value indicating the perceived quality of the measured image. Hence, the metrics provide an indication on how much the measured images have degraded as a result of increasing the level of scattering. The MS-SSIM criterion was compared with the more classical SSIM criterion, as MS-SSIM has been shown to

provide equally good or better approximations to perceived image quality than SSIM for images which are distorted by isolated pixels corrupted by large levels of gaussian noise, known as salt-and-pepper-noise [43].

The 3D profiles estimated from the 3 algorithms at different attenuation lengths and different number of binary frames are compared to a reference 3D image obtained in clear water and by aggregating 5000 binary frames.

The RMSE was calculated using the standard RMSE equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (d_i - \hat{d}_i)^2}{n}} \quad (7)$$

where d_i is the relative distance recorded in pixel i of the base image and \hat{d}_i is the relative distance recorded in pixel i of the measured image, and n is the total number of pixels used in the RMSE calculation. Similarly, the SSIM and MS-SSIM criteria are also computed using the same reference 3D image.

4. Results

The objects that were imaged in the experiments are shown in the photographs in Fig. 5, along with their dimensions. All target surfaces were painted with a primer layer in order to avoid specular reflections, which can saturate the SPAD detector array. Detailed 3D features were investigated using a brass wall plate elbow connector for 15 mm diameter pipes (Fig. 5(a)). We also considered a 3D-printed target (Fig. 5(b)), which had simpler 3D features as it was made of three pillars with heights of 10, 20 and 30 mm, as shown in Fig. 5(c). Both these targets were mounted on a black anodized aluminum breadboard, and were used in experiments with stationary objects, while the target shown in Fig. 5(d) and (e) was used to investigate imaging of a moving object. This third target is a steel equal tube T-connector for 7 mm diameter pipes, and it was mounted in the two orientations shown in Fig. 5(d) and (e) as seen from the system. The T-connector was attached to a 2 m long rail, which was used to move the target across the field of view of the system (x and y), and along the z-direction.

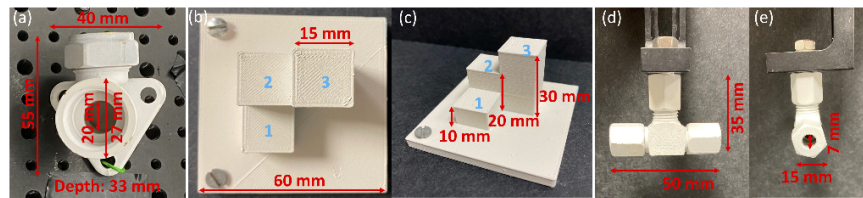


Fig. 5. Photographs of the targets used during the experiments and their dimensions. Measurements with stationary targets were performed by using (a) a brass pipe connector, (b) a custom 3D printed target with three square pillars providing four flat surfaces at different heights, shown from the side in (c). The target used for the experiments with moving objects is shown in (d) and (e). All targets were painted with a primer in order to avoid specular reflections.

4.1. Static scenes

The brass pipe connector (Fig. 5(a)) was analyzed first as its shape allows us to visualize better how its fine detail is not observed as the scatter is increased. For this investigation, the histograms were built by aggregating 5000 binary frames for varying levels of attenuation between the system and the target. Expressed as one-way attenuation length between transceiver and target, the attenuation of the underwater environment was varied in the range < 0.5 AL to 7.5 AL. The

average optical power used to perform the measurements was adjusted based on the level of scattering in water, in order to maximize the return from the target but without saturating the SPAD detector array. The optical power measurements were performed at the output of a 550 μm fiber coupled to the laser source, and the selected values are reported in Table 2 for different attenuation lengths, along with the concentration of scattering agent used to achieve the estimated attenuation level.

Table 2. Average optical power used during experiments at different scattering levels corresponding at different concentrations of scattering agent in water.

Attenuation lengths between transceiver and target	<0.5	3.6 ± 0.4	5.5 ± 0.4	7.1 ± 0.4	7.5 ± 0.4
Concentration of scattering agent in water	0%	0.002%	0.003%	0.004%	0.0046%
Average optical power (mW)	0.96	22	32	47	52

For display and comparison purposes, the final 3D images are shifted and gated between 0 mm and 50 mm, such that any points below 0 mm are set to 0 mm and points with a relative distance greater than 50 mm are set to 50 mm. For the Ensemble Method, the 3D images were obtained by setting the threshold for the probability of presence of a surface detection $w_0 = 0.05$ and we discarded any points with a posterior probability of presence smaller than 0.5 (see [31]). In order to display the final 2D image results for all methods, the pixels containing the discarded points were set to 0 and displayed as black pixels. The results are displayed in Fig. 6, showing that the cross-correlation method performs well at low attenuation lengths up to 3.6 AL, however the results contain significantly more noise at higher attenuation lengths (> 3.6 AL) as this method includes no denoising or automatic surface detection procedures (but no surface is displayed in empty pixels). RT3D provides smooth images at low attenuation lengths but tends to miss surfaces at high attenuation lengths probably due to the smoothing arising from the use of the point cloud denoiser [41]. The Ensemble Method offers a suitable tradeoff here as it does not seem to remove too many points at lower attenuation lengths (< 7.1 AL) and is able to denoise the results at higher attenuation lengths (≥ 7.1 AL) without distorting the target 3D image due to spatial smoothing. Although the Ensemble Method does not necessarily improve the 3D profile of the target, it offers advanced noise reduction and surface detection with conservative inference error results [31]. Thus, this method is used as a primary method for the remainder of the investigation with static scenes.

We now compare the 3D target reconstructions between the brass connector and the pillar target (shown in Fig. 5(b)). For this investigation, the histograms were built by aggregating different numbers of binary frames (5000, 1000, 200, 50 binary frames) for each of the attenuation levels equivalent to <0.5 AL, 3.6 AL, 5.5 AL, 7.1 AL, 7.5 AL. Again, the histograms were time gated so that the target was centered in a temporal window of 400 bins, and the final images are gated between 0 mm and 50 mm. The estimated 3D target reconstruction for the brass connector are displayed in Fig. 7, and for the pillar target are displayed in Fig. 8. The average optical power used to perform the measurements are the same as in Table 2. We can see that as the attenuation length increases, or the number of binary frames aggregated decreases, the visual quality of the 3D reconstruction decreases, and it becomes more difficult to discern the features of the target.

In addition to the qualitative results displayed in Figs. 7 and 8, the 3D image reconstruction was evaluated by using the SSIM, MS-SSIM and RMSE metrics introduced in Section 3.5. Due to minor target alignment limitations during the experiments, the target position can be slightly different between the different recordings. Thus, the images are registered to the reference 3D images and cropped (65×123 pixels, i.e. $28.19 \text{ mm} \times 26.67 \text{ mm}$) prior to computing the three criteria. The SSIM, MS-SSIM and RMSE results are shown in Fig. 9. The calculations were performed by aggregating a varying number of binary frames at increasing numbers of attenuation lengths.

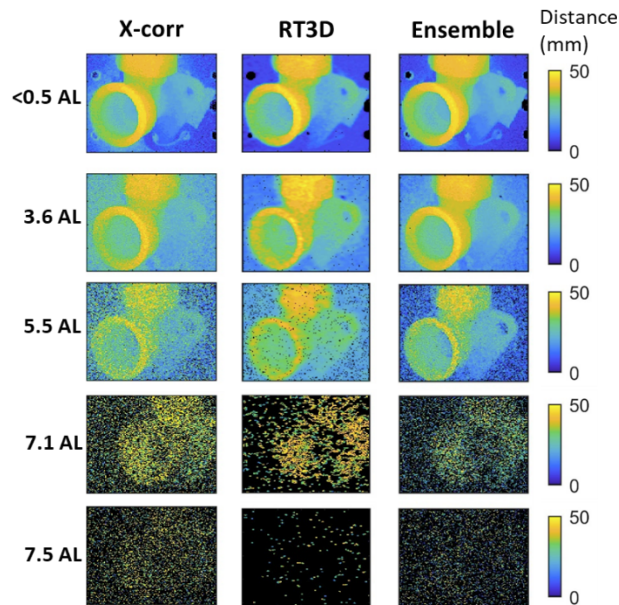


Fig. 6. Brass pipe connector 3D profiles obtained aggregating 5000 binary frames and gating the histograms at 400 temporal bins. The figure compares 3D profiles at different attenuation lengths for cross-correlation method, the RT3D method [29] and Ensemble method [31]. The color bar represents the distance along the axis from the transceiver, with an arbitrary origin chosen near the reference target position.

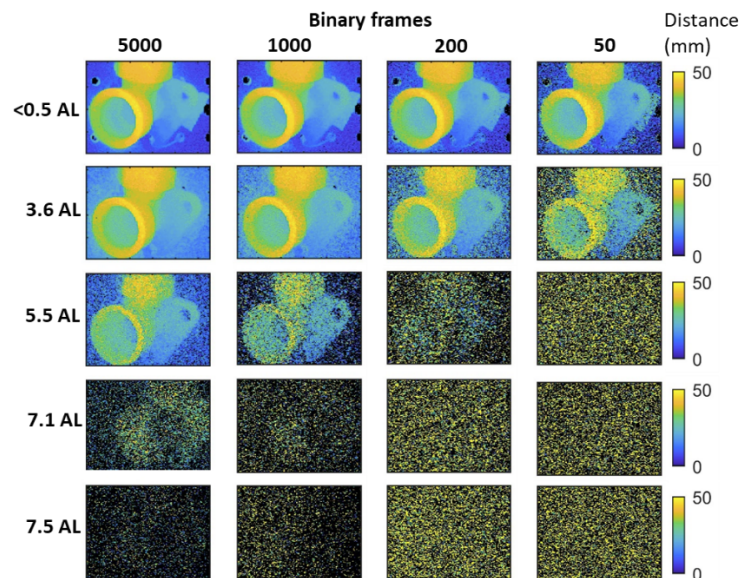


Fig. 7. Brass pipe connector 3D profiles using the ensemble method [31], where the histograms are gated at 400 temporal bins and built by aggregating different numbers of binary frames, at different attenuation lengths. The color bar represents the distance along the axis from the transceiver, with an arbitrary origin chosen near the reference target position.

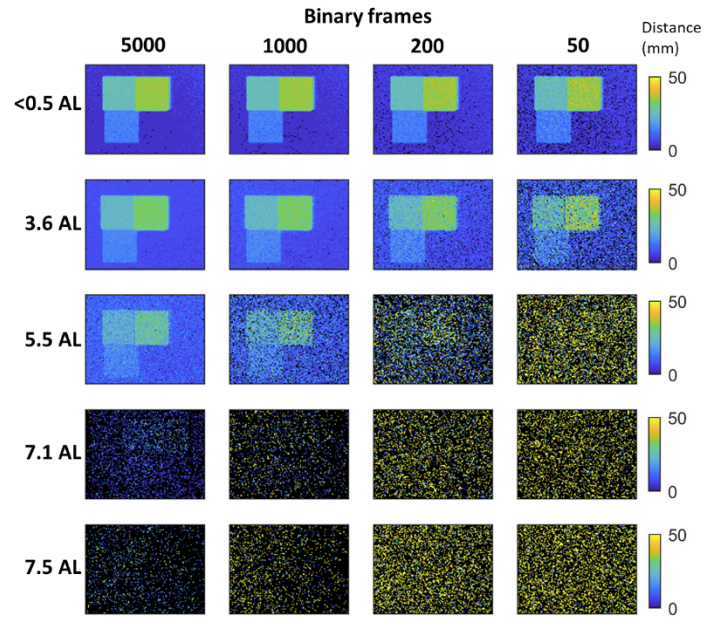


Fig. 8. 3D reconstructions of the pillar target using the ensemble method [31], where the histograms are gated at 400 temporal bins and built by aggregating different numbers of binary frames for different attenuation lengths. The color bar represents the distance along the axis from the transceiver, with an arbitrary origin chosen near the reference target position.

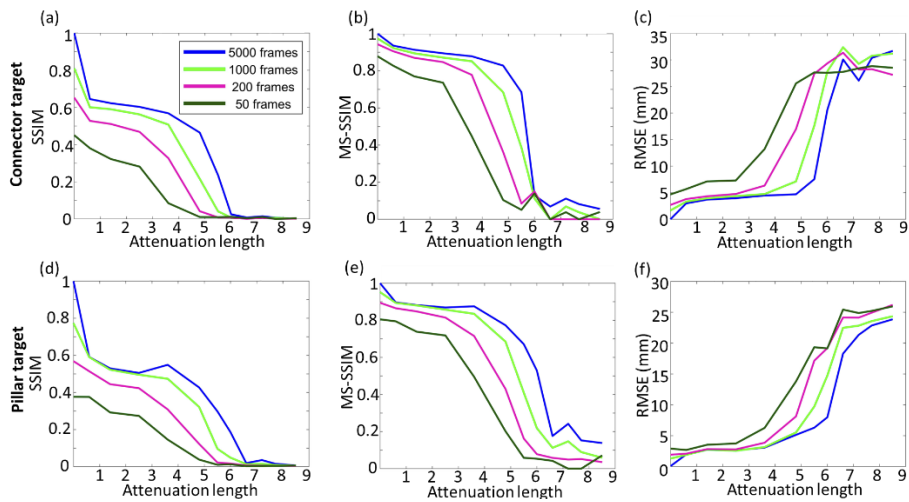


Fig. 9. (a) SSIM, (b) MS-SSIM, and (c) RMSE (in mm) results for the brass connector target. (d) SSIM, (e) MS-SSIM, (f) RMSE (in mm) results for the pillar target. The analysis was performed by aggregating different numbers of binary frames.

As explained previously, the primary motivation for using MS-SSIM in addition to the SSIM is its robustness in providing better approximations to perceived image quality than SSIM for images which are degraded by large sparsely pixelated noise. These results show that the MS-SSIM remains high for longer than the SSIMs results for both targets as the AL increases. Moreover, the MS-SSIM results are more in line with the visual quality of the 3D profiles in Fig. 7 and Fig. 8 than the SSIM curves. Indeed, it confirms that SSIM is more affected by small random errors. We observe that the results of both targets follow the same trend for all the metric measurements. As the attenuation length increases, the sensor detects more scattering events which lead to the wrong estimation of the distance for an increasing number of pixels, hence the SSIM/MS-SSIM values decrease and the RMSE values increase. Similarly, as the number of frames used to construct the histogram data are decreased, the SSIM/MS-SSIM values decrease and the RMSE values increase. From the results of both targets, we can see that the image quality is degraded to where the target starts to become unrecognizable at approximately 4-5 AL for data constructed of 50 binary frames. As the number of binary frames used to construct the data is increased, the target becomes unrecognizable at approximately 6-7 AL for 5000 frames. Due to the similarity of the results for both targets, the connector target is used for the remainder of this section of the paper as this is the more complicated of the two targets.

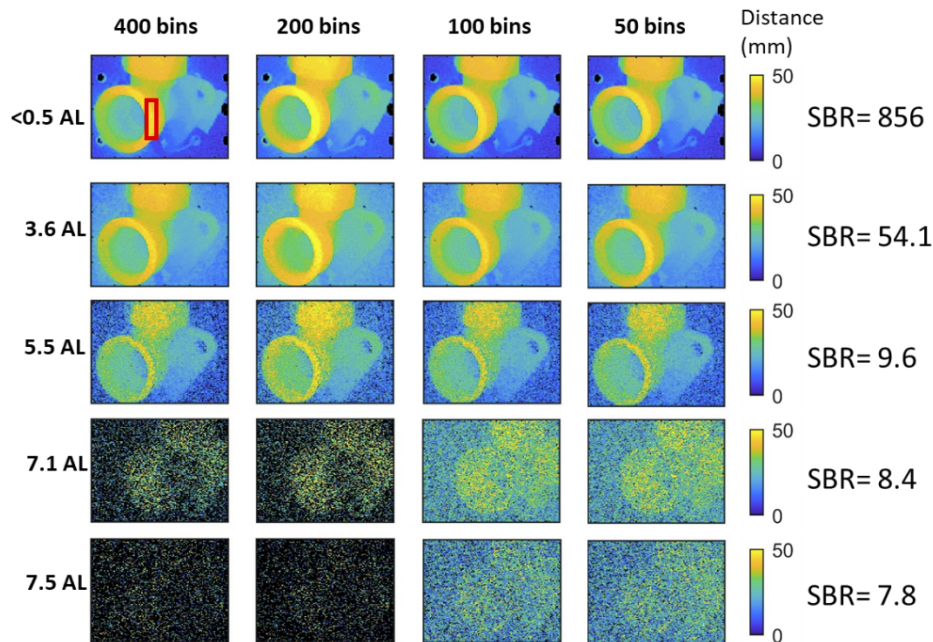


Fig. 10. Brass pipe connector 3D profiles obtained processing different lengths of timing bin window, at different attenuation lengths. The results were obtained by aggregating 5000 binary frames.

The previous analysis was performed over a timing window of 400 temporal bins centered on the expected target reflection, and taken from the full recorded 1540 bins shown in the examples illustrated in Fig. 4. By using a smaller set of bins, the large scattering peak shown in Fig. 4(b) can be removed from the analysis, removing ambiguity from the estimate of target depth. The timing window of 400 bins corresponds to a distance of 1.48 m in water. This distance was chosen in order to have a large timing window which would be of more interest when the target position is not known. However, when the expected target position can be bounded more precisely (e.g., if a sonar is used to obtain a coarse range estimate), the data can be processed using fewer

timing bins in order to exclude more scattering events in the analysis without removing any signal return. Therefore, the same data were processed by aggregating 5000 binary frames, and varying the number of bins in the timing window where the processing is performed. When processing the data over fewer temporal bins with the ensemble method (while correlating over the full 400 temporal bins to produce the posterior distributions), we only considered the posterior distribution values over the admissible distance range and renormalized these values to produce the final distance estimates. This allowed the estimation to be more robust as we can reduce the effects of noise when considering all the information from the 400 temporal bins. The results are illustrated in Fig. 10, and show that as the timing window is reduced, more features of the target can be reconstructed. For example, when the data are processed over 50 bins, imaging can be performed at up to 7.1 AL and the target can be detected at 7.5 AL. Additionally, Fig. 10 shows the Signal-to-Background (SBR) defined as the ratio between the number of photons in the highest bin in the peak and the average background per bin. The calculation was performed by estimating the average SBR in the area corresponding to the most reflective part of the target connector within the red rectangle shown in the first 3D profile of Fig. 10, equivalent to 16×35 pixels.

4.2. Dynamic scenes

To illustrate real-time, online processing, we used the aluminum T-connector shown in Fig. 5(d) and (e). The data for all attenuation lengths were taken with two different T-connector orientations: one shows the shortest side of the connector where the axis of the connecting hole is parallel to the line of sight of the system (as in Fig. 5(e), and it will be called parallel orientation); whilst the second orientation shows the full length of the connector which is perpendicular to the line of sight of the system (as in Fig. 5(d), and it will be called perpendicular orientation). The histograms were built by aggregating together 50 binary frames. We compare the results for cross-correlation, the RT3D method [29], and the Ensemble Method [31] for distance and intensity estimation. The histograms in the real time processing were gated so that the target would be centered in a temporal window of 400 bins. To obtain real-time scene reconstruction, GPU processing methods were used where each pixel was processed in parallel. Specifically for the Ensemble Method, the construction of the code was built in a modular fashion, where the method was broken into three different GPU kernels. The first is to obtain the posterior probability distribution values, the second normalizes these probability distribution values, and the final kernel produces the results from these normalized posterior probabilities. This layout not only prevents the GPU from crashing, but also helps with future work where we can combine posterior probabilities obtained from uniform and multiple different non-uniform distribution estimates and normalize these in the second kernel. Further improvement would be required to also calculate the results for each w value for each pixel in parallel to further improve the processing times. The times for processing each image frame were recorded for all three methods at several attenuation lengths up to 6.6 AL. The average results are presented for the parallel orientation of the connector in Table 3 and for the perpendicular orientation of the connector in Table 4. For the Ensemble Method, $M = 10$ for the discrete set of w values, for faster processing speeds. From the results we can see that as the attenuation length increases the processing times for all methods decrease. As the attenuation length is increased, the scattering of emitted photons increases and so fewer photons are returned to the LiDAR detector. Thus, the histograms produced for higher attenuation lengths are more sparse compared to histograms from measurements taken at lower attenuation lengths, and so require less time to process. We also observe that the cross-correlation method is much faster than both the RT3D and the Ensemble Method, which exhibit similar times to process the data where the Ensemble Method is very slightly the slower of the two methods. However, it is important to observe that the processing times required by both methods are less than the time needed to measure the 50 binary frames used to build the histograms, enabling low latency

three-dimensional reconstruction and visualization of the scene, as shown in the example in [Visualization 1](#). In addition, it was observed that for the Ensemble Method smaller values of M resulted in faster processing times. Thus, as the different posterior distributions for different w values were computed in parallel, the slightly slower processing times of the Ensemble Method compared with the RT3D method may be a result of overheads and/or libraries used.

Table 3. Mean processing times (in ms) to obtain an individual image frame, along with its standard deviation (in brackets) for different attenuation lengths. These results were obtained by processing the data obtained for the parallel orientation of the T-connector target (orientation shown in Fig. 5(e)).

Attenuation lengths	<0.5	0.6	1.4	2.5	3.6	4.8	5.5	6.0	6.6
<i>Cross-Correlation</i>	1.15 (0.11)	1.42 (0.16)	1.27 (0.19)	1.49 (0.24)	1.20 (0.20)	0.97 (0.18)	0.97 (0.13)	0.93 (0.12)	0.92 (0.07)
<i>RT3D Method [29]</i>	34.53 (1.25)	35.34 (1.20)	32.95 (1.53)	34.53 (1.38)	31.34 (1.39)	28.95 (1.18)	29.53 (1.38)	28.54 (2.28)	28.14 (1.44)
<i>Ensemble Method [31]</i>	32.76 (1.21)	35.78 (2.37)	33.72 (1.92)	37.08 (2.58)	33.05 (1.77)	29.51 (1.32)	30.92 (0.91)	30.12 (0.88)	29.47 (0.81)

Table 4. Mean processing times (in ms) to obtain an individual image frame, along with its standard deviation (in brackets) for different attenuation lengths. These results were obtained by processing the data obtained for the perpendicular orientation of the T-connector target (orientation shown in Fig. 5(d)).

Attenuation lengths	<0.5	0.6	1.4	2.5	3.6	4.8	5.5	6.0	6.6
<i>Cross-Correlation</i>	1.27 (0.19)	1.54 (0.19)	1.35 (0.20)	1.65 (0.27)	1.30 (0.15)	1.02 (0.11)	0.96 (0.09)	0.94 (0.09)	0.91 (0.09)
<i>RT3D Method [29]</i>	34.67 (1.98)	36.17 (2.80)	34.83 (1.23)	34.86 (1.39)	32.36 (1.74)	30.74 (1.28)	29.55 (1.12)	28.30 (2.77)	29.46 (4.34)
<i>Ensemble Method [31]</i>	35.23 (2.44)	37.19 (3.11)	34.51 (3.15)	38.56 (2.98)	34.22 (2.28)	30.07 (1.19)	30.7 (0.98)	30.54 (0.9)	29.7 (0.9)

Videos, displaying both relative distance and intensity profiles as well as 3D point-clouds, of the T-connectors parallel and perpendicular orientations were recorded for several attenuation lengths. When using the Ensemble Method, the images and point clouds were cleaned for the lower AL (≤ 4.8), by setting the limiter for probability of presence of a surface detection $w_0 = 0.25$, as described in [23], and ignored any points with a probability of detection less than 0.5 and intensity value less than 5. For 5.5, 6.0, 6.6 AL, the intensity limit was reduced to 3 as the intensity was too low at this higher scattering level resulting in too many points being removed.

[Visualization 2](#) and [Visualization 3](#) show the results obtained in clear water and at an attenuation level equivalent to 4.8 AL between the system and the target, respectively. However, as the attenuation length increases, the quality of the image reconstruction decreases, and the target is not recognizable for attenuation lengths greater than 5.5 AL (shown in [Visualization 4](#)), as predicted in the offline processing. However, we do still see clusters of points moving for the higher attenuation lengths. Thus, while the methods may not be adequate for image reconstruction at higher attenuation lengths, they could be suitable for target tracking. For comparison, [Visualization 5](#) show the results obtained at a low level of scattering equivalent to less than 0.5 AL between the system and the target, by processing the data with the cross-correlation method. In this case, the target is clearly visible and its details can be discerned in the final video, although the point cloud shows a higher level of noise with respect to the Ensemble Method. [Visualization 6](#) shows the results obtained with cross-correlation at a higher level of scattering equivalent to 5.5 AL, and in this case it is not possible to identify or detect the target.

5. Conclusions and discussion

This paper presents a fully submerged underwater single-photon LiDAR transceiver system based on a CMOS SPAD detector array, with real-time three-dimensional target reconstruction capability. The system was characterized under different levels of turbidity.

We have performed 3D imaging experiments at an underwater depth of approximately 1.8 meters, using stationary and moving targets. Three-dimensional profiles and intensity reconstructions of the targets were obtained in several underwater scattering conditions, at a target distance of approximately 3 meters in water. Relative distance and intensity information were retrieved up to 7.5 AL by using three different target reconstruction methods: cross-correlation, RT3D [29], and the recently developed algorithm based on joint surface detection and distance estimation [31]. The CMOS SPAD detector array used had 192×128 pixels, with TCSPC electronics integrated in each pixel [32]. This allowed us to measure the timing information simultaneously for the full field of view. Therefore, we directly interfaced the CMOS SPAD detector array to a GPU and implemented the three chosen reconstruction algorithms for single-photon LiDAR imaging with low latency. We demonstrated processing times below the data acquisition time needed to acquire histograms for an individual frame (i.e. in the order of 100 ms), and achieved low latency of < 40 ms. The investigations reported in this paper show that three-dimensional single-photon videos of moving targets can achieve a rate of 10 frames per second for moving targets in turbid conditions.

Funding. CENSIS and Leonardo UK by the, Royal Academy of Engineering under the Research Fellowship scheme (RF201617/16/31, RF\201920\19\190); Engineering and Physical Sciences Research Council (EP/T00097X/1, EP/S000631/1, EP/S026428/1); Ministry of Defence University Defence Research Collaboration (UDRC) in Signal Processing.

Acknowledgments. The authors would like to thank Prof Yvan Petillot and Leonard McLean of the Ocean System Laboratory at Heriot-Watt University for the facilities and their support during the experiments.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are available in Ref. [44].

References

1. P. F. McManamon, "Review of lidar: a historic, yet emerging, sensor technology with rich phenomenology," *Opt. Eng.* **51**(6), 060901 (2012).
2. M. E. Warren, "Automotive LIDAR Technology," in IEEE S. VLSI Circ. 2019, Kyoto, Japan.
3. A. M. Wallace, A. Halimi, and G. S. Buller, "Full Waveform LiDAR for Adverse Weather Conditions," *IEEE T. Veh. Technol.* **69**(7), 7064–7077 (2020).
4. D. McLeod, J. Jacobson, M. Hardy, and C. Embry, "Autonomous Inspection using an Underwater 3D LiDAR," in OCEANS 2013, San Diego, USA.
5. G. S. Buller and A.M. Wallace, "Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition," *IEEE J. of Sel. Top. Quant.* **13**(4), 1006–1015 (2007).
6. S. Pellegrini, G. S. Buller, J. M. Smith, A.M. Wallace, and S. Cova, "Laser-based distance measurements using picosecond resolution time-correlated single-photon counting," *Meas. Sci. Technol.* **11**(6), 712–716 (2000).
7. A. McCarthy, X. Ren, A. Della Frera, N. R. Gemmill, N. J. Krichel, C. Scarcella, A. Ruggeri, A. Tosi, and G. S. Buller, "Kilometer-range depth imaging at 1550 nm wavelength using an InGaAs/InP single-photon avalanche diode detector," *Opt. Express* **21**(19), 22098–22114 (2013).
8. Z.-P. Li, J.-T. Ye, X. Huang, P.-Y. Jiang, Y. Cao, Y. Hong, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng, F. Xu, and J.-W. Pan, "Single-photon imaging over 200 km," *Optica* **8**(3), 344–349 (2021).
9. Z.-P. Li, X. Huang, Y. Cao, B. Wang, Y.-H. Li, W. Jin, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng, F. Xu, and J.-W. Pan, "Single-photon computational 3D imaging at 45 km," *Photon. Res.* **8**(9), 1532–1540 (2020).
10. Z.-P. Li, X. Huang, P.-Y. Jiang, Y. Hong, C. Yu, Y. Cao, J. Zhang, F. Xu, and J.-W. Pan, "Super-resolution single-photon imaging at 8.2 kilometers," *Opt. Express* **28**(3), 4076–4087 (2020).
11. R. Tobin, A. Halimi, A. McCarthy, X. Ren, K. J. McEwan, S. McLaughlin, and G. S. Buller, "Long-range depth profiling of camouflaged targets using single-photon detection," *Opt. Eng.* **57**(3), 031303 (2017).
12. D. B. Lindell and G. Wetzstein, "Three-dimensional imaging through scattering media based on confocal diffuse tomography," *Nat. Commun.* **11**(1), 4517 (2020).
13. R. Tobin, A. Halimi, A. McCarthy, M. Laurenzis, F. Christnacher, and G. S. Buller, "Three-dimensional single-photon imaging through obscurants," *Opt. Express* **27**(4), 4590–4611 (2019).

14. R. Tobin, A. Halimi, A. McCarthy, P. J. Soan, and G. S. Buller, "Robust real-time 3D imaging of moving scenes through atmospheric obscurant using single-photon LiDAR," *Sci. Rep.* **11**(1), 11236 (2021).
15. J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal, "High-flux single-photon lidar," *Optica* **8**(1), 30–39 (2021).
16. J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal, "Dead Time Compensation for High-Flux Ranging," *IEEE Trans. Signal Process.* **67**(13), 3471–3486 (2019).
17. A. Maccarone, A. McCarthy, X. Ren, R. E. Warburton, A. M. Wallace, J. Moffat, Y. Petillot, and G. S. Buller, "Underwater depth imaging using time-correlated single-photon counting," *Opt. Express* **23**(26), 33911–33926 (2015).
18. F. M. Caimi and F.R. Dalgleish, "Performance considerations for continuous-wave and pulsed laser line scan (LLS) imaging systems," *J. Eur. Opt. Soc.* **5**, 10020S (2010).
19. A. Maccarone, A. McCarthy, A. Halimi, R. Tobin, A. M. Wallace, Y. Petillot, S. McLaughlin, and G. S. Buller, "Depth imaging in highly scattering underwater environments using time-correlated single-photon counting," *Proc. SPIE* **9992**, 99920R (2016).
20. A. Maccarone, G. Acconcia, U. Steinlehner, I. Labanca, D. Newborough, I. Rech, and G. S. Buller, "Custom-Technology Single-Photon Avalanche Diode Linear Detector Array for Underwater Depth Imaging," *Sensors* **21**(14), 4850 (2021).
21. P. Peronio, I. Labanca, G. Acconcia, A. Ruggeri, A. A. Lavdas, A. A. Hicks, P. P. Pramstaller, M. Ghioni, and I. Rech, "32-channel time-correlated-single-photon-counting system for high-throughput lifetime imaging," *Rev. Sci. Instrum.* **88**(8), 083704 (2017).
22. A. Maccarone, F. Mattioli Della Rocca, A. McCarthy, R. Henderson, and G. S. Buller, "Three-dimensional imaging of stationary and moving targets in turbid underwater environments using a single-photon detector array," *Opt. Express* **27**(20), 28437–28456 (2019).
23. M. K. Moghimi and F. Mohanna, "Real-time underwater image enhancement: a systematic review," *J. real-time Image Pr.* **18**(5), 1509–1525 (2021).
24. P. Risholm, J. Thorstensen, J. T. Thielemann, K. Kaspersen, J. Tschudi, C. Yates, C. Softley, I. Abrosimov, J. Alexander, and K. H. Haugholt, "Real-time super-resolved 3D in turbid water using a fast range-gated CMOS camera," *Appl. Optics* **57**(14), 3927–3937 (2018).
25. M. Massot-Campos and G. Oliver-Codina, "One-shot underwater 3D reconstruction," in *IEEE Emerging Technology and Factory Automation (ETFA) 2014*, Barcelona, Spain.
26. M. Massot-Campos, G. Oliver-Codina, H. Kemal, Y. Petillot, and F. Bonin-Font, "Structured light and stereo vision for underwater 3D reconstruction," in *OCEANS 2015*, Genova, Italy.
27. L. Parmesan, N. Dutton, N. Krstajic, N. Calder, A. Holmes, L. A. Grant, and R. K. Henderson, "A 256 × 256 SPAD array with in-pixel Time to Amplitude Conversion for Fluorescence Lifetime Imaging Microscopy," in *International Image Sensor Workshop 2015*, Vaals, Netherlands.
28. N. Krstajic, S. Poland, J. Levitt, R. Walker, A. Erdogan, S. Ameer-Beg, and R. K. Henderson, "0.5 billion events per second time correlated single photon counting using CMOS SPAD arrays," *Opt. Lett.* **40**(18), 4305–4308 (2015).
29. J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tournet, and S. McLaughlin, "Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nat. Commun.* **10**(1), 4984 (2019).
30. S. Plosz, A. Maccarone, S. McLaughlin, G. S. Buller, and A. Halimi, "Real-time reconstruction of 3D videos from single-photon LiDAR data in the presence of obscurants," accepted for publication in *IEEE Trans. Comput. Imaging* (2023).
31. K. Drummond, S. McLaughlin, Y. Altmann, A. Pawlikowska, and R. Lamb, "Joint surface detection and depth estimation from single-photon Lidar data using ensemble estimators," in *2021 Sensor Signal Processing for Defence Conference (SSPD)*, Edinburgh, UK.
32. R. K. Henderson, N. Johnston, F. Mattioli Della Rocca, H. Chen, D. Day-Uei Li, G. Hungerford, R. Hirsch, D. Mcloskey, P. Yip, and D. J. S. Birch, "A 192 × 128 Time Correlated SPAD Image Sensor in 40-nm CMOS Technology," *IEEE J. Solid-St. Circ.* **54**(7), 1907–1916 (2019).
33. S. Pellegrini, B. Rae, A. Pingault, D. Golanski, S. Jouan, C. Lapeyre, and B. Mamdy, "Industrialised SPAD in 40 nm Technology," in *IEDM Tech. Dig. 2017*, San Francisco, USA.
34. K. Morimoto and E. Charbon, "High fill-factor miniaturized SPAD arrays with a guard-ring-sharing technique," *Opt. Express* **28**(9), 13068–13080 (2020).
35. I. Gyongy, A. T. Erdogan, N. A. W. Dutton, G. Mora Martín, A. Gorman, H. Mai, F. Mattioli Della Rocca, and R. K. Henderson, "A Direct Time-of-flight Image Sensor with in-pixel Surface Detection and Dynamic Vision," *IEEE J. Sel. Top. Quant.*, 10.1109/JSTQE.2023.3238520 (2023).
36. A. Laux, R. Billmers, L. Mullen, B. Concannon, J. Davis, J. Prentice, and V. Contarino, "The a, b, c, s of oceanographic lidar predictions: a significant step toward closing the loop between theory and experiment," *J. Mod. Opt.* **49**(3–4), 439–451 (2002).
37. S. Q. Duntley, "Light in the sea," *J. Opt. Soc. Am.* **53**(2), 214–234 (1963).
38. P. W. R. Connolly, X. Ren, R. K. Henderson, and G. S. Buller, "Hot pixel classification of single-photon avalanche diode detector arrays using a log-normal statistical distribution," *Electron. Lett.* **55**(18), 1004–1006 (2019).

39. Q. Legros, J. Tachella, R. Tobin, A. McCarthy, S. Meignen, G. S. Buller, Y. Altmann, S. McLaughlin, and M. E. Davies, "Robust 3D reconstruction of dynamic scenes from singlephoton Lidar using beta-divergences," *IEEE Trans. Image Process.* **30**, 1716–1727 (2021).
40. Q. Legros, S. McLaughlin, Y. Altmann, S. Meignen, and M. E. Davies, "Robust depth imaging in adverse scenarios using single-photon Lidar and beta-divergences," in *2020 Sensor Signal Processing for Defence Conference (SSPD), Edinburgh, UK*.
41. G. Guennebaud and M. Gross, "Algebraic point set surfaces," *ACM Trans. Graph.* **26**(3), 23–31 (2007).
42. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Trans. Image Processing* **13**(4), 600–612 (2004).
43. Z. Wang, E. P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers* (2003).
44. A. Maccarone, K. Drummond, A. McCarthy, U. Steinlehner, J. Tachella, D. Aguirre Garcia, A. Pawlikowska, R.A. Lamb, R.K. Henderson, S. McLaughlin, Y. Altmann, and G.S. Buller, "Submerged single-photon LiDAR imaging sensor used for real time 3D scene reconstruction in scattering underwater environments," Heriot-Watt University (Apr 2023). Fig7_50frames.xlsx, Fig7_1000frames.xlsx, Fig7_200frames.xlsx, Fig7_5000bins_vs_AL.xlsx, Fig10_100bins_vs_AL.xlsx, Fig10_50bins_vs_AL.xlsx, Fig10_200bins_vs_AL.xlsx, Fig6_Ensemble_vs_AL.xlsx, Fig10_400bins_vs_AL.xlsx, Fig6_RT3D_vs_AL.xlsx, Fig6_X_Corr_vs_AL.xlsx. 10.17861/15cf027a-d54a-4c07-b9c6-c6a5374d54f8.