



Real-time super-resolved 3D in turbid water using a fast range-gated CMOS camera

PETTER RISHOLM,^{1,*} JOSTEIN THORSTENSEN,¹ JENS T. THIELEMANN,¹ KRISTIN KASPERSEN,¹ JON TSCHUDI,¹ CHRIS YATES,² CHRIS SOFTLEY,² IGOR ABROSIMOV,² JONATHAN ALEXANDER,² AND KARL HENRIK HAUGHOLT¹

¹SINTEF Digital, Forskningsveien 1, 0373 Oslo, Norway

²Odos Imaging Limited, Scottish Microelectronics Centre, Alexander Crum Brown Road, EH9 3FF, Edinburgh, UK

*Corresponding author: petter.risholm@sintef.no

Received 19 December 2017; revised 9 April 2018; accepted 11 April 2018; posted 11 April 2018 (Doc. ID 315858); published 9 May 2018

We present a range-gated camera system designed for real-time (10 Hz) 3D estimation underwater. The system uses a fast-shutter CMOS sensor (1280 × 1024) customized to facilitate gating with 1.67 ns (18.8 cm in water) delay steps relative to the triggering of a solid-state actively Q-switched 532 nm laser. A depth estimation algorithm has been carefully designed to handle the effects of light scattering in water, i.e., forward and backward scattering. The raw range-gated signal is carefully filtered to reduce noise while preserving the signal even in the presence of unwanted backscatter. The resulting signal is proportional to the number of photons that are reflected during a small time unit (range), and objects will show up as peaks in the filtered signal. We present a peak-finding algorithm that is robust to unwanted forward scatter peaks and at the same time can pick out distant peaks that are barely higher than peaks caused by sensor and intensity noise. Super-resolution is achieved by fitting a parabola around the peak, which we show can provide depth precision below 1 cm at high signal levels. We show depth estimation results when scanning a range of 8 m (typically 1–9 m) at 10 Hz. The results are dependent on the water quality. We are capable of estimating depth at distances of over 4.5 attenuation lengths when imaging high albedo targets at low attenuation lengths, and we achieve a depth resolution (σ) ranging from 0.8 to 9 cm, depending on signal level. © 2018 Optical Society of America

OCIS codes: (150.6910) Three-dimensional sensing; (110.0113) Imaging through turbid media; (110.6880) Three-dimensional image acquisition; (150.5670) Range finding.

<https://doi.org/10.1364/AO.57.003927>

Provided under the terms of the [OSA Open Access Publishing Agreement](#)

1. INTRODUCTION

The oceans regulate weather, produce vast amounts of biomass, and are a vital part of global transport and commerce. However, despite oceans covering more than 70% of the planet's surface and ultimately supporting all living organisms, only 5% is explored by humans, and vast resources are untapped. Consequently, there is a great need for technology that can increase our knowledge of the oceans through surveillance and monitoring. Monitoring marine habitats for biodiversity, biomass, and healthiness requires sensors that provide high-quality texture and 3D data at high frame rates.

Many technologies have been proposed for underwater imaging and ranging such as sonars [1], structured light techniques [2–4], and lidars [5–7]. However, none of these technologies have been shown to provide cost-effective technology with a small footprint that makes them easily integratable with underwater vehicles, and at the same time provide high-resolution 3D data at real-time speeds that facilitate monitoring

marine life. In this paper we present a compact range-gated system based on a fast CMOS camera chip that provides a performance compromise between the long range of sonars and the high resolution of scanning lidars.

Range-gated imaging has been shown to provide accurate time-of-flight (TOF) measurements underwater by using the travel time between a laser pulse is emitted and the reflected laser pulse is detected to determine distances [8,9]. Another use of range-gated systems is to effectively reduce the impact of backscatter on image contrast by gating near objects of interest [10–12].

Current state-of-the-art range-gated systems use gated image intensifiers, coupled to CCD image sensors. The image intensifier performs two functions: short time gating, which allows for ultrashort exposure times, and optical signal level amplification. The combination of CCD and intensifiers allows for effective range-gating, but it has a number of drawbacks, primarily related to speed, spatial resolution, cost, and complexity

of the instruments. Most range-gated systems are made for use in air, and only few examples of commercial systems have been available underwater, e.g., LUCIE [13] and Aqua Lynx [14].

The main advantages of using these systems underwater is that they can effectively suppress backscatter [10], work at long ranges because they are very light sensitive [9], and can produce high depth precision because of picosecond gating resolution. In [8], the authors describe a method for estimating 3D using an intensified range-gated CCD camera. They achieve better than 1 mm range accuracy for 0.5 Mpixels because they use very short laser pulses (200 ps) in combination with gate times of 200 ps. However, construction of a depth image took 1 s, and the scanning range was limited at this update rate. Dalglish *et al.* [6] demonstrated that a pulse-gated laser line scanner was able to detect a target at up to seven attenuation lengths. A range-gated spot scanner is presented in [5] that acquires 40 k points per second with high precision and constructs a 3D point cloud on the fly.

Methods that are designed for estimating distance based on range-gated signals must be robust to the effects of attenuation and forward and backward scatter. A number of methods have been proposed to exploit the reflected signal for range estimation. In leading edge detection, the leading edge is detected as the signal crosses a certain threshold, but to find a suitable threshold can be difficult when dealing with noisy and interfering signals [15]. With varying noise and background signals, a dynamic threshold may be suitable [16]. Constant fraction detection is insensitive to pulse amplitude, but it depends on the pulse waveform (should be close to symmetric) and width [17]. Peak detection determines the range by finding the maximum of the returned signal [15]. Different weighted averaging methods have also been investigated [9,18], as well as using least squares fit of an idealized curve to the response curve [19]. In [20], they propose to use the full laser waveform to extract information beyond the first reflection to get richer information in urban settings. The returned intensity has an inverse range-squared dependency, which they model in [21] to improve the range estimates compared to a weighted average method. Some authors have also investigated how to improve the depth resolution below the gating delay step by using super-resolution methods. Super-resolution has been mostly investigated in long-range applications where long laser pulses are used [22,23]. One approach when using shorter pulses is the weighted average approach [18], but it is sensitive to curve shape and interfering scattered signals.

In this paper, we first introduce the effect scattering has on the ranged gated signals. Next, we propose an underwater range-gated system that uses a fast CMOS camera chip together with a 532 nm solid-state laser integrated in a 7-liter housing. The use of a CMOS chip facilitates high frame rates while at the same time achieves high spatial and depth resolution at a potentially lower cost and system complexity compared to a system using a gated intensifier and a CCD camera chip. We present a peak determination algorithm that is robust to scattering, and a method to improve the depth resolution 18 times beyond the resolution of the range slicing. Lastly, we validate the range estimation model and discuss the results.

2. EFFECT OF SCATTERING ON RANGE-GATED SIGNALS

In a nonturbid environment that does not exhibit any light scattering, nor light attenuation other than the r^2 falloff with distance, the range from which the most number of photons (i.e., the highest peak of the signal) is reflected will be the best depth estimate. However, turbid environments may introduce other signal peaks, due to forward and backward scatter. An example is shown in Fig. 1, where we show a response trace for a pixel using the proposed range-gated system. Seven identical targets are placed at different distances from the camera. The albedo properties of five regions of the target (top to bottom) are approximately 10%, 75%, 50%, 25%, and 90%. The blue trace shows the intensity recorded at the location of the turquoise square for different gating distances. The object at the location of the turquoise square is at a range of 7 m from the camera, while the object to the right of the turquoise square is found at a range of 3.2 m. In the proposed range-gated system, an image gated at 2 m contains all photons that are reflected off objects in the range [2 m, ∞], hence the cumulative form of the blue trace. This is formalized in Eq. (1). The (negative) derivative of this cumulative blue trace (shown in dashed red) is proportional to the number of photons that were collected on the chip from a specific distance. The derivative trace exhibits a peak in signal at 7 m due to the target, at 3.2 m due to forward scatter from the target on the right of the marker, and a continuous rise in signal from 2 m to 0 m due to backscatter from particles close to the camera. Furthermore, notice that the forward scatter peak at 3.2 m of the derivative signal is higher than the peak caused by the target at 7 m due to the attenuation of the signal with distance.

3. MATERIAL AND METHODS

In this section, we first present the hardware and the sequencer, which constitutes the range-gated system. Next, we describe our approach to range-gated depth estimation, an approach to achieve super-resolution, and the FPGA implementation of the algorithms to facilitate real-time depth calculations. The performance of an underwater 3D camera is dependent on the water quality/attenuation. In the Appendix, we describe the system we developed for measuring water attenuation.

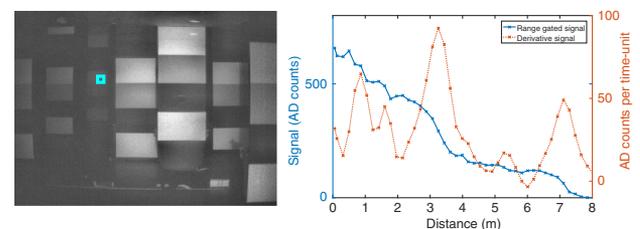


Fig. 1. Effect of scattering and attenuation on delay sweeps acquired in water with attenuation length of 2.7 m. Left, histogram-equalized intensity image gated at 2 m from the camera. Right, range-gated sweep in blue from the location of the turquoise square. Markers indicate the gating distances. The (negative) derivative of the range gated sweep is included in dashed red.

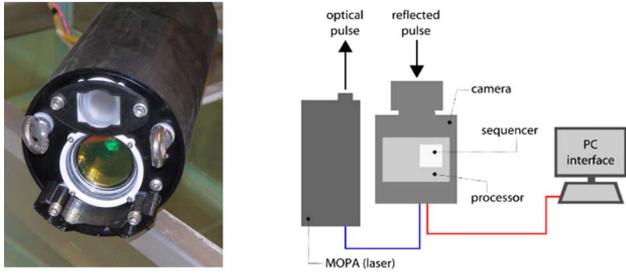


Fig. 2. Total weight of the 3D imaging system is approximately 9 kg, and the volume is 7 liters. Left, the front of the housing is shown with openings for the (lower) camera and (upper) laser. Right, the internals of the camera consist of a solid-state laser, a CMOS camera, which includes a sequencer and 3D processor, and a 70 m Ethernet connection to a PC that visualizes the acquired depth and intensity data.

A. System Overview

The range-gated system consists of a camera from ODOS imaging and a laser from Bright Solutions. Figure 2 shows images of the housing and camera internals. The housing has a diameter of 155 mm and a length of 370 mm, which constitutes approximately 7 liters.

The camera has a fast black-and-white CMOS chip with a minimum shutter duration of 10 microseconds at a resolution of up to 1280 × 1024 pixels. The number of images that can be acquired per second is correlated with the region of interest that is used. At 0.5 Mpixels, the camera delivers a frame rate of 1 kHz, but if the full frame is used it delivers 400 Hz. An on-board sequencer is integrated in the camera firmware, which allows for fine-tuned control of the opening of the camera shutter in relation to triggering of the laser pulse at steps of 1.67 ns, which facilitates range gating. A Gigabit Ethernet connection over a 70 m cable is used to control and transfer images from the camera.

The laser is a 532 nm solid-state laser with active Q-switch and a pulse width of 1 ns, repetition rate of 1 kHz, and a pulse energy of 3.5 mJ.

A software API provides the user with full freedom in customizing acquisition sequences, i.e., how to best utilize the 1000 exposures per second. The user can control the number of distances to gate in a sequence, the spatial/temporal step size between consecutive ranges, as well as how many exposures to average at each range.

B. Range Gating/Sequencer

For 3D ranging purposes, a shutter duration of 10 microseconds means that the shutter closes after the return of the entire signal, and an image $I(\mathbf{x}, z): R^3 \rightarrow R$, where $\mathbf{x} \in \Omega_I \subset R^2$ is a point in the 2D image domain, gated at a distance z can be viewed as the integration of photons returned from distances from z and outwards,

$$I(\mathbf{x}, z) = \int_z^\infty I'(\mathbf{x}, z) dz, \tag{1}$$

where $I' \equiv \frac{\partial I}{\partial z}$ is a measure for the number of photons that were detected from an infinitesimal range.

The system facilitates range gating with a temporal sample increment of $\Delta_t = 1.67$ ns. The speed of light underwater is $c_w \approx 22.5$ cm/ns, and because the light has to travel back and forth, the minimum spatial sample increment is $\Delta_z = \frac{1}{2} \Delta_t c_w = 18.8$ cm. The zero point in time/space is defined by the emission of the laser pulse, and the gating is defined as delays of increments Δ_z relative to the laser pulse trigger. The sequencer facilitates acquisition of delay sweeps, i.e., acquisition of a set of images $\{\hat{I}(\mathbf{x}, z_{\min} + \Delta_z i)\}_{i=1, \dots, N_r}$, where $\hat{I}(\mathbf{x}, z_{\min} + \Delta_z i) = \sum_{j=1}^{N_a} I(\mathbf{x}, z_{\min} + \Delta_z i)$ are N_a averaged images from the same distance, gated at regular distances from the camera. The sequencer also facilitates binning N_b^2 pixels, where $N_b \in \{1, 2, 4, 8\}$. In Fig. 1, we show a delay sweep curve for a pixel, and notice that the response of the delay sweep curve is a result of the temporal convolution between the returned (1 ns) laser pulse and the temporal response curve of the camera. The opening of the shutter takes approximately 15 ns. The range gating plots can be viewed as a cumulative plot of the number of photons that are reflected from a certain distance from the camera and outwards. The derivative $I'(\mathbf{x}, z_i) = I(\mathbf{x}, z_{i+n}) - I(\mathbf{x}, z_{i-n})$ is a measure of the number of photons that are collected in a pixel from the range $[z_{i-n}, z_{i+n}]$. Example images gated at two different distances are shown in Fig. 3. The albedo for the five regions (top to bottom) of the targets are 10%, 75%, 50%, 25%, and 90%. The intensity axis is set to the 1st and 98th percentile of the image intensities. Notice the backscatter halo in the image gated at 0.1 m, which is gated away in the two images gated further from the camera (2.2 m and 3.4 m).

C. Depth Estimation

The main assumption, which lays the foundation for the design of the depth estimation algorithm, is that the point on the delay sweep curve where a pixel (photosensitive area) detects the most photons per unit time represents the distance to the target. Consequently, the proposed algorithm aims to find the peak of a differentiated delay sweep curve.

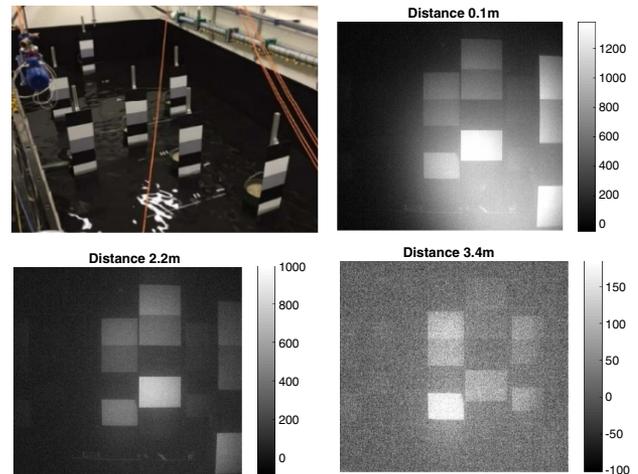


Fig. 3. Range-gated images acquired in water with attenuation length 1.2 m. Top left, 1 m × 0.3 m multi-albedo targets placed in a 4 m × 8 m × 1 m pool. Top right, image gated at 0.1 m from camera. Bottom left, image is gated at 2.2 m. Bottom right, image is gated at 3.4 m from the camera.

D. Derivative Filter

We observe from Fig. 1 that the steepest part of the delay sweep curve, which represents the position of the target that we are interested in detecting, is approximately $2\Delta_z = 37.6$ cm long. Hence, the derivative kernel we use is $D = [-1 \ 0 \ 1]$ to be sensitive to the signal increase caused by objects while limiting the influence of the signal increase due to backscatter, which has a longer rise time.

Signals from objects far away (or signals from objects with a low albedo) will quickly drown in noise due to low signal levels. Several sources contribute with noise to the delay sweep signal, e.g., readout/sensor noise, shot noise, and intensity noise from the laser. Some of this noise can be reduced by averaging and/or binning images. However, because of different constraints (1 kHz image acquisition, and the wish for high observed frame rate), it is only practical to average a few frames (1–8) at each delay step, which only reduces the noise by a relatively small factor (square root of the number of averages). As mentioned in the previous section, the depth estimation is based on detecting peaks in the derivative signal. Any noise in the delay sweep curve is increased during differentiation by a factor of $\sqrt{2}$. Consequently, to improve the signal to noise before peak finding we apply a Gaussian low-pass filter in the z direction. We have found through Monte Carlo simulations that a smoothing filter G of length $4\Delta_z$ and with a standard deviation of $\sigma = \Delta_z$ provides a good compromise between smoothing out noise, while retaining pertinent information. The two filters are convolved such that a combined derivative/smoothing $F = G * D$ filter of length $6\Delta_z$ is convolved with the delay sweep curve. However, to avoid cropping the resulting derivative delay sweep signal by the length of the filter (i.e., reducing the range), we rather perform a linear extrapolation of the delay sweep curves by three samples on each end. A linear extrapolation also limits the introduction of peaks at the boundaries, especially close to the camera in backscatter.

E. First Peak Finding

As can be observed from the delay sweep in Fig. 1, the derivative of a sweep may have several peaks—there may be many small peaks caused by noise, there may be peaks caused by forward scatter from nearby objects, a peak from backscatter, and a peak caused by the actual object. The simplest approach of searching for $Z(\mathbf{x}) = \operatorname{argmax}_{z_i} \{I'(\mathbf{x}, z_i)\}_{i=1, \dots, N}$ will in many cases detect

backscatter, and may also pick out the peak caused by forward scatter of a nearby bright object, which can often be stronger than the peak caused by an object further away (see Fig. 1 for example). The design of our algorithm is based on the insight that the most distant peak, which is higher than a noise floor T_n , is the most probable object peak. By always selecting the most distant peak, we avoid selecting forward scatter peaks caused by bright objects that are closer to the camera, and peaks caused by backscatter,

$$Z(\mathbf{x}) = \max_{z_i} \{z_1, \dots, z_N\},$$

$$\text{where } \begin{cases} I'(\mathbf{x}, z_{i-1}) < I'(\mathbf{x}, z_i) < I'(\mathbf{x}, z_{i+1}) \\ I'(\mathbf{x}, z_i) > T_n \end{cases} \quad (2)$$

In areas where there is no object, but where the pixel represents a ray that curves space close to a bright object, we may

observe a peak caused by forward scatter. We have not found an effective approach to filter out these peaks, but this will be addressed in further work.

F. Max Peak Finding

The sensor noise is approximately $\sigma_{\text{sensor}} = 72$ DN. The noise of the derivative signal that we perform peak finding on is $\sqrt{\sigma_{\text{sensor}}^2 + \sigma_{\text{sensor}}^2} = \sqrt{2}\sigma_{\text{sensor}}$. Through simulations, we have found that a noise threshold based on $3\sqrt{2}\sigma_{\text{sensor}}$ provides a good compromise between picking up distant peaks, while avoiding to pick up peaks caused by sensor noise. According to statistical theory, only 0.3% of samples from a Gaussian distribution will fall outside of the range $[-3\sigma, 3\sigma]$. Consequently, in theory, the noise threshold $T_n = \sqrt{2} \frac{3\sigma_{\text{sensor}}}{\sqrt{(N_a N_b^2)}}$, where N_a and N_b^2 are the number of pixels that are averaged and binned respectively, should only provide a 0.3% chance of picking up a noise peak.

However, for some pixels, there may not be any peaks higher than T_n even when there is an object along the pixel ray, because the signal has been strongly attenuated. Consequently, for the pixels where no peak was found using the first-peak finding algorithm, we do a second pass and report the maximum peak: $Z(\mathbf{x}) = \operatorname{argmax}_{z_i} \{I'(\mathbf{x}, z_i)\}_{i=1, \dots, N_r}$.

The peak heights $I'(\mathbf{x}, Z(\mathbf{x}))$ can be viewed as a confidence measure of the detected peaks and, depending on the use-cases, can be used to filter out unlikely peaks in postprocessing.

G. Super-Resolution

The maximum depth resolution we can achieve based on the previous first- and max-peak procedures is $\Delta_z = 18.8$ cm since we are searching through discrete samples. However, the underlying signal is strong enough to support a significant improvement in depth precision by carefully designing an interpolation (super-resolution) algorithm. The discrete samples $I'(z + i\Delta_z)$, $i = 1, \dots, N_r$ are samples from a continuous underlying function. We measured this function with high resolution (after performing derivative and Gaussian filtering as noted above), through repeated measurements with small shifts in the sample points, in a scatter-free environment. The resulting curve I' is shown in Fig. 4 (left), and without super-resolution, the algorithm would report the position of one of the data points (red crosses) which are distributed with fixed spacing of $\Delta_z = 18.8$ cm. In principle, calculating a weighted average on the full curve will allow accurate positioning of the

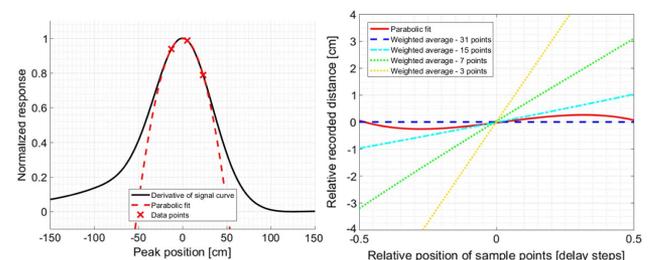


Fig. 4. Left, (black) finely resolved signal curve; (red) parabolic fit to three data points. Right, positioning bias with respect to shifts in position of the sample points relative to the signal peak, using weighted average or parabolic peak fit. Parabolic fit provides a robust method, with the use of only three data points.

center of the curve. However, in our case, the response curve will be influenced by backscatter and forward scatter, ambient light, etc. This prevents us from using the entire curve for center positioning. If using only a central part of the curve [e.g., N points around $Z(\mathbf{x})$], the calculated center position will be biased with regards to the position of the data points relative to the actual peak of the response curve. This is shown in Fig. 4 (right). In a real situation, we are often limited to using only a few data points around the signal peak, e.g., 3–7 points in total. With such a limited number of data points available, we observe a significant bias, which in practice will limit our measurement accuracy.

An alternative method is based on the observation that the peak of the curve, where we find the distance to our object, closely resembles a parabola. Therefore, we choose a parabolic fit for our interpolation, using only the three central data points [Fig. 4 (top)]. In Fig. 4 (bottom), we see that the parabolic fit has below 3 mm bias with regards to the position of the data points, due to the good resemblance between the signal curve and a parabola close to the peak of the signal curve. The bias can be further reduced by applying a wider Gaussian filter; however, this will increase influence from other close-lying signal peaks. In the results section (Fig. 8), we show that a depth resolution down to 0.8 cm can be achieved using the super-resolution algorithm. This is more than 20 times higher resolution than the sampling interval of 18.8 cm, i.e., the super-resolution method is highly efficient in increasing depth precision.

These results show that an interpolating parabola provides a robust and accurate fit of the peak position around the discrete maximum $I'(\mathbf{x}, z_i)$ and can be used to detect the underlying peak with minimum bias with regards to small shifts in the sample points. By fitting a parabola to the three points near the peak [$I'(\mathbf{x}, z_{i-1})$, $I'(\mathbf{x}, z_i)$, and $I'(\mathbf{x}, z_{i+1})$], differentiating, and setting to zero, the super-resolved peak is computed as

$$Z_*(\mathbf{x}) = Z(\mathbf{x}) + \frac{I'(\mathbf{x}, Z(\mathbf{x}) - \Delta_z) - I'(\mathbf{x}, Z(\mathbf{x}) + \Delta_z)}{2(I'(\mathbf{x}, Z(\mathbf{x}) - \Delta_z) - 2I'(\mathbf{x}, Z(\mathbf{x})) + I'(\mathbf{x}, Z(\mathbf{x}) + \Delta_z))} \quad (3)$$

H. FPGA Implementation

A Gigabit Ethernet connection is used to interface with the camera, which constrains the effective transfer rate to approximately 600 Mb/s. This means that it is not feasible to transfer the 1000 images (with a resolution of 960×512 and 16-bit pixels, this would be about 1 GB/s) that are acquired every second to a PC for processing and visualization. Consequently, the depth estimation algorithm has been implemented in the camera FPGA to facilitate real-time streaming of image data for visualization on the PC side. The following data can be streamed to the PC side:

- Averaged intensity image, where the user can specify j and k :

$$\hat{I}_j^k = \sum_{i=j}^k \hat{I}(\mathbf{x}, z_{\min} + i\Delta_z) / (k - j + 1), \quad (4)$$

- Depth related data:

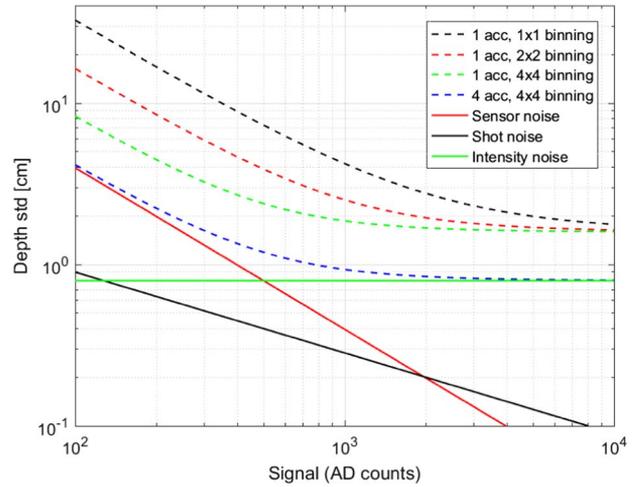


Fig. 5. Dashed lines, theoretical limit to depth precision due to sensor noise, shot noise, and laser intensity noise for varying number of accumulations and binning. Solid lines, individual noise contributions for $N_a = 4$ and $N_b = 4$.

- Depth image: $Z_*(\mathbf{x})$
- Peak height image (confidence): $I'(\mathbf{x}, Z(\mathbf{x}))$
- Full set of sweep data: $\{\hat{I}(\mathbf{x}, z + i\Delta_z)\}_{i=1, \dots, N_r}$.

Individual pixel binning with factors of 1, 2, 4, or 8 can be performed on the three data streams. Binning of the data in the FPGA is generally performed either to increase the signal-to-noise ratio, or to reduce the required data bandwidth. The ability to transfer the full set of sweep data has been included here so that we can use it in the future to estimate backscatter profiles and subtract the backscatter from the images to improve the contrast and visual appearance.

The two main constraints that affect the PC side frame rate are (1) that the camera can only acquire images at a 1000 Hz and (2) that the transfer bandwidth is limited to approximately 600 Mb/s. Assuming we would like to keep a PC-side frame rate of 10 Hz, that means we have 100 exposures to construct a depth measurement. The more exposures that are averaged at each range, the higher depth precision is feasible (this is shown in Fig. 5), but at the cost of being able to cover a smaller range. A good compromise between range and depth precision is to use $N_a = 4$, which allow us to sample 25 ranges. With Δ_z distance between samples, we cover a range of 4 m. It is also possible to increase the step between samples, e.g., to $2\Delta_z$ to cover 8 m, but this comes at a small cost of being a little less sensitive to peaks in the backscatter region.

4. EXPERIMENTS AND RESULTS

In this section, we first present predictions of the theoretical depth precision of the system derived from measurable characteristics of the signal/sensor noise and system response. We validate the system in terms of depth precision, at what distances we can detect objects, and how robust the system is to scattering both in a controlled indoor pool environment where the attenuation length of the water is varied by adding clay to the water, and in the wild. We show that the empirical results correspond to the theoretical predictions. Finally, we show some

qualitative results from imaging fish in a fish farm. A robust attenuation measurement tool was developed to provide an easy reference for the results. The tool is presented in Appendix A.

A. Theoretical Depth Resolution

In a previous work [24], we have developed theory to predict the precision obtainable from a TOF system.

$$\sigma_{\text{TOF}} = \frac{c\tau_{\text{response}} \sigma_{\text{tot}}}{2\sqrt{2m} S}, \quad (5)$$

where m is the number of temporal data points used, c is the speed of light, τ_{response} is the total system response time, and S and σ_{tot} are the signal and the total noise of a single datapoint (both in units of photoelectrons). The effective signal $S = S_1 N_a N_b^2$ where S_1 is the signal from a single acquisition, N_a and N_b^2 are the number sample points that are averaged and binned respectively. Total noise is

$$\sigma_{\text{tot}} = \sqrt{S + N_a N_b^2 \sigma_{\text{sensor}}^2 + (\sigma_{\text{int}} S / \sqrt{N_a})^2}, \quad (6)$$

where the contributions are from shot noise, readout noise, and laser intensity noise. Considering the ratio σ_{tot}/S , which ultimately determines the measurement precision σ_{TOF} , we get

$$\frac{\sigma_{\text{tot}}}{S} = \sqrt{\frac{1}{N_a N_b^2} \left(\frac{1}{S_1} + \frac{\sigma_{\text{sensor}}^2}{S_1^2} \right) + \frac{1}{N_a} \sigma_{\text{int}}^2}. \quad (7)$$

We see that shot noise and sensor noise are both reduced with increasing N_a and N_b^2 , as each measurement and pixel are statistically independent, while intensity noise is reduced only with increasing N_a , as intensity noise is common mode for all binned pixels. We also see that the three contributions have a different dependence on signal level. Intensity noise is independent on S_1 , while shot noise and dark noise contributions decrease with $1/\sqrt{S_1}$ and $1/S_1$, respectively. These trends are shown in Fig. 5, where we also see that increased binning does not improve precision at high signal levels where intensity noise dominates. In our system, the Gaussian/derivative filter has a length of $m = 7$. We have measured $\tau_{\text{response}} \in [15 \text{ ns}, 17 \text{ ns}]$, $\sigma_{\text{sensor}} \in [70, 85]$ AD counts, and $\sigma_{\text{int}} \in [3.0\%, 4.5\%]$.

B. Pool Results: General Setup

A number of studies were performed in an $8 \text{ m} \times 4 \text{ m}$ pool with a depth of 1 m, which can be seen in Fig. 3. The walls and floor of the pool were painted a matted black to avoid reflections off the pool surfaces. Brown clay was used to increase the turbidity (lower the attenuation length) of the water. Appendix A describes the tool we built to measure the attenuation length of the water.

The same acquisition parameters were used across all pool experiments. We acquired sweeps covering the whole range of the pool [0–8 m] with $\Delta_z = 18.8 \text{ cm}$, number of averages per range was $N_a = 4$, binning before depth estimation $N_b = 4$, and a noise threshold of $T_n = 36$.

C. Pool Results: Depth Precision

To study how far and at what precision we were able to detect an object under different conditions, we imaged a flat 1 m \times 1 m multi-albedo (70%, 30%, 50%, 10%, 90%) target at

different distances ranging from 2.5 to 7 m from the camera and in water qualities ranging from 0.7 m attenuation length up to 2.6 m. With one target at one distance in the pool, forward scattering will not be an issue. Figure 6 shows an example intensity image from the acquired dataset gated at 1.9 m with the target at 4 m as well as the corresponding depth map.

In Fig. 7 we summarize the findings, where we plot for each distance, the mean and standard deviation of the depth estimates over an 8×8 neighborhood for each of the attenuation lengths. The plot shows that at 0.7 m attenuation length, we are able to get reliable depth estimates (standard deviation of less than 10 cm) up to a distance of 4.5 m, while at 1.6 m attenuation length we get reliable depth estimates up to 6.5–7.0 m. At long attenuation lengths, we are limited by the r^2 effect, but at shorter attenuation lengths, we have empiric evidence that we are able to see at least 4.5 attenuation lengths.

To validate the theoretical depth precision we presented in Fig. 5, we also present the depth precision as function of signal level in different water qualities based on the pool data in Fig. 8. Note that when using only a single 3D frame for estimating depth precision, the noise contribution due to laser intensity noise is not taken into account. Intensity noise will cause a

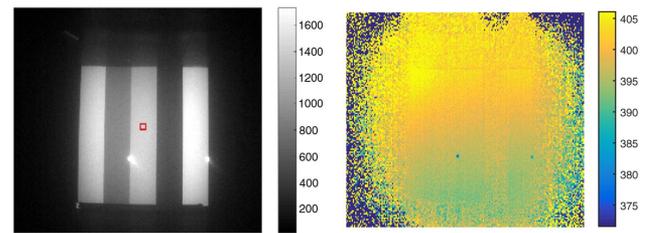


Fig. 6. Flat target with five albedos (left to right: 70%, 30%, 50%, 10%, 90%) acquired at 4 m in water with attenuation length of 1.6 m. Left, an intensity image of the flat target with varying optical reflective properties. Image is gated at 1.9 m. Colorbar shows AD counts. The square indicates the ROI used when calculating the depth precision as reported in Fig. 7. Right, estimated depth map (cm).

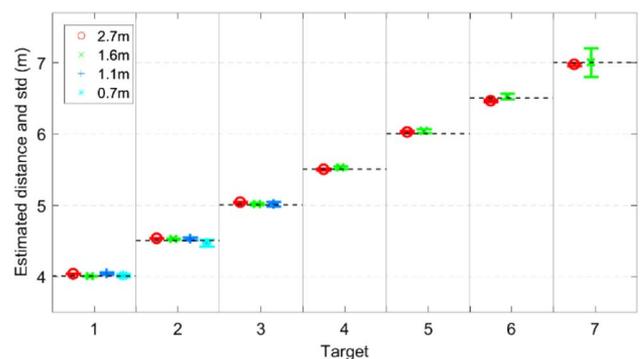


Fig. 7. Depth estimation results of single flat target. The x axis shows the target placed at seven different distances from 4 to 7 m at 0.5 m intervals. The circle denotes the mean over an 8×8 spatial neighborhood, while the length of the bar indicates 2σ . The various colors represent different attenuation lengths.

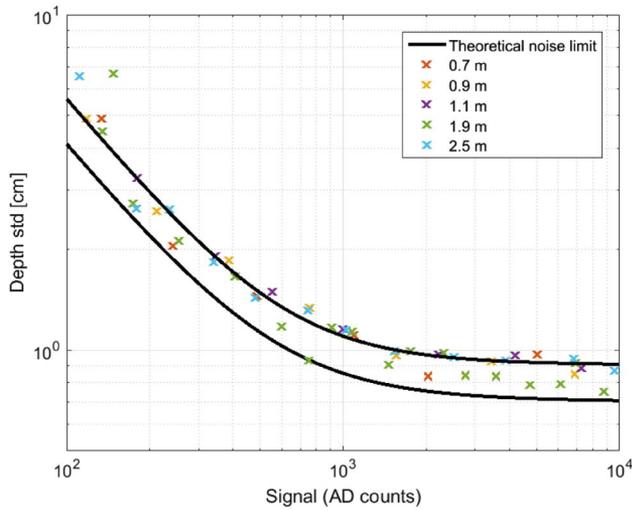


Fig. 8. Depth precision plotted against signal at different attenuation lengths (crosses). Notice that the results are independent of attenuation lengths. The results are in agreement with the theoretical results, using $\sigma_{\text{dark}} = 70$, $\tau_{\text{response}} = 15$ ns and $\sigma_{\text{fixed}} = 0.7$ cm (lower curves); and $\sigma_{\text{dark}} = 85$, $\tau_{\text{response}} = 17$ ns and $\sigma_{\text{fixed}} = 0.9$ cm (upper curves).

common mode noise in the entire depth image, which will cause the measured range to vary between 3D frames.

The results show that depth precision is independent of water quality, and it depends only on signal intensity. Furthermore, it is clear that the experimentally obtained depth precision is consistent with system parameters at low signal levels only, using (lower curves) $\sigma_{\text{dark}} = 70$, $\tau_{\text{response}} = 15$ ns, and (upper curve) $\sigma_{\text{dark}} = 85$, $\tau_{\text{response}} = 17$ ns. At high intensities, the depth precision is limited to around 0.8 ± 0.1 cm. As described above, we estimate depth precision by calculating the standard deviation over a small region of pixels. The remaining standard deviation of 0.8 ± 0.1 cm arises from fixed pixel-to-pixel variations in absolute distance. It is clear that for signals above ~ 1000 counts, this variation constitutes the limit of our depth precision.

D. Pool Results: Effect of Scattering

As shown in Fig. 1, scattering can cause peaks in the derivative and can therefore cause faulty depth detections. The depth estimation algorithm has incorporated two measures to reduce detection of unwanted peaks caused by scattering: linear extrapolation (instead of a constant extrapolation, which would cause a peak in the derivative, especially in the backscatter region) of signal before convolving with the derivative filter to avoid detecting backscatter peaks, and a first peak search before defaulting to a max peak search to detect objects further out than peaks caused by forward scatter. To study how well the algorithm handles scattering, we imaged a “forest” of multi-albedo $100 \text{ cm} \times 30 \text{ cm}$ targets placed at different distances and positions relative to the camera, as shown in Fig. 3. Sweeps were acquired at different attenuation lengths, and the estimated distance and depth precision were calculated from the resulting depth maps. In Fig. 9, we summarize the results. The targets are numbered 1–7 as going from left to

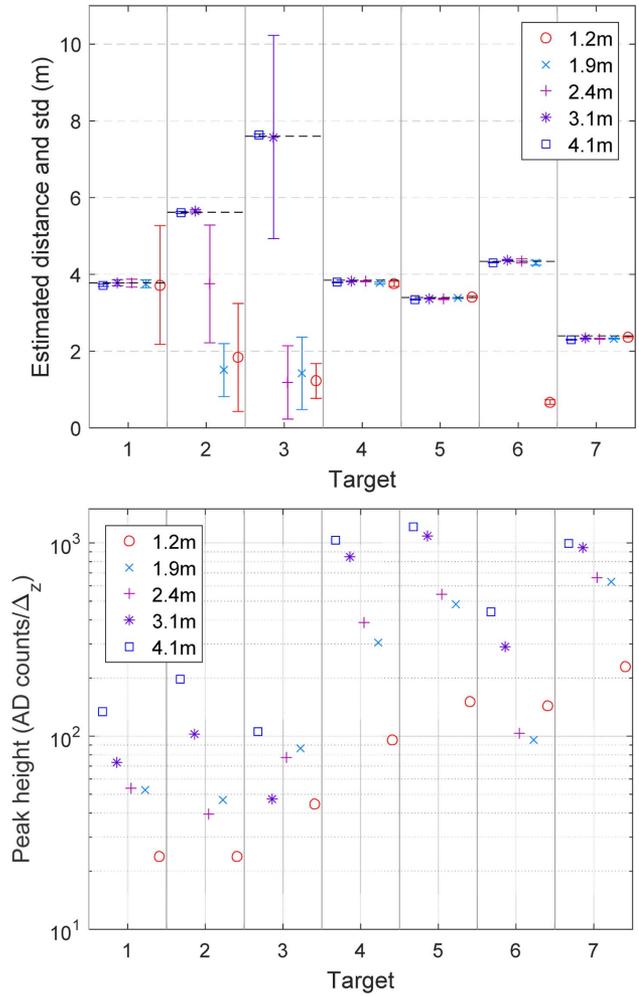


Fig. 9. Top, depth estimation results from dataset with multiple targets at fixed distances. The flat stapled lines show the ground truth distances. Bottom, corresponding plot where the y axis (log scale) shows the peak height (AD counts/ Δ_z), which is proportional to the number of photons that are reflected and collected from the object.

right in the image in Fig. 3. In the top plot of Fig. 9, we plot the mean and standard deviation of the depth estimates for the different attenuation lengths. The plots for each target are ordered from left to right with decreasing attenuation length. The results show that targets 4, 5, and 7 are detected with high precision across all turbidities. However, target 3 at 7.7 m distance is only detected in water with a high attenuation length (4.1 m and 3.1 m). For lower attenuation lengths, the forward scatter from target 4 is detected because the signal from target 3 is so attenuated that the peak caused by the target is lower than T_m , and the peak caused by the forward scatter is higher. We also observe that forward scatter affects target 6 for the most turbid water (attenuation length of 1.2 m). The estimated results from target 2 and 3 at high turbidities are drawn even closer to the camera than the targets causing forward scattering at 3.8 m, which means that there are some backscatter peaks that are not suppressed. This is a consequence of the real signal being too attenuated to be reliably detected. The corresponding height of the detected peak is shown in the bottom plot of Fig. 9.

E. Sea Trials: Signal-to-Noise Ratio Versus Depth Resolution

The 3D precision that is possible to attain is highly correlated with the signal-to-noise ratio (SNR). We define $SNR \equiv S_{\text{white}}/\sigma_{\text{dark}}$. A white or reflective target far away should give the same accuracy and standard deviation as a dark and close target (they will exhibit the same SNR). The signal level is dependent on the distance to the target, the albedo of the target, as well as the attenuation length of the water. The noise is reduced with increasing binning and accumulations. Figure 10 shows a multi-albedo target (1 m \times 0.3 m with albedos right to left: 10%, 75%, 50%, 25%, 90%, for the five regions) acquired at approximately 10 m during sea trials where we estimated the attenuation length to be 3.5 m. Notice that there is a weak ghost of the rope visible in the image gated at 8 m in Fig. 10. The reason for that is the exponential decay of photoelectrons in the shutter. The light reflected from the rope close to the camera generates a high number of photoelectrons. In a first-order description, these photoelectrons will be drained exponentially with a short time constant. However, there are some residual photoelectrons left when the shutter opens at 8 m.

We imaged the target at 10 different distances and varying positions in the image (illumination is lower on the edges of the image compared to the center). In Fig. 11, we summarize all the sample points from the sea trials in the way of SNR versus depth precision from experiments imaging the target at different distances. Notice the trend, which shows that with increasing SNR, the depth precision increases. We got an SNR of 3 at 14 m range and a standard deviation of 10 cm. With good signal levels at shorter ranges, we approach a depth precision of 1 cm. Also in sea trials, the achieved depth precision is in

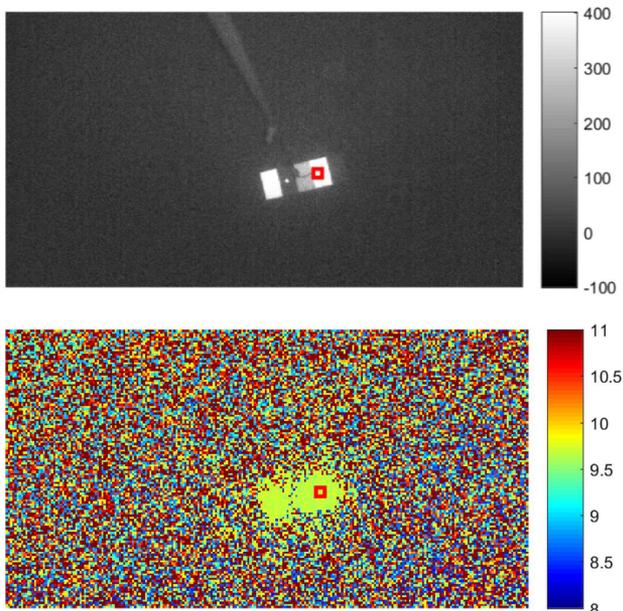


Fig. 10. Corresponding intensity image (gated at 8 m) and depth map (in meters) from sea trial. The signal of the brightest part of the target was used to compute the SNR. A 3×3 region over the target in the depth map was used to compute the depth precision and distance. The target was estimated to be at 9.7 m, and we obtained an SNR of 17 and a depth precision of 2.3 cm.

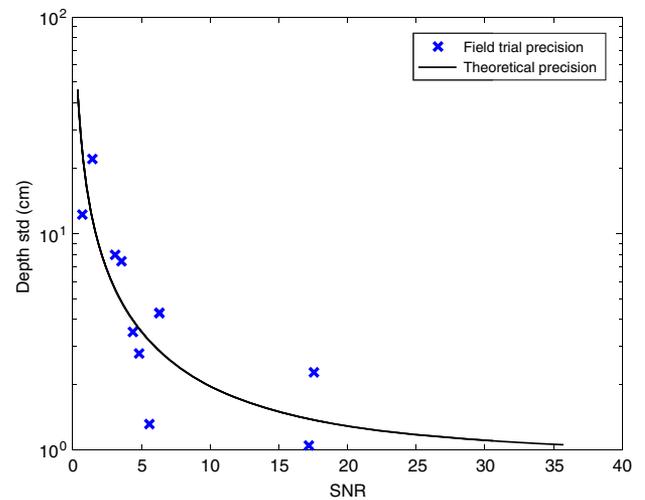


Fig. 11. SNR versus depth precision during field trials. At high signal levels, we approach a depth precision of 1 cm. These measurements correlate with the theoretical precision presented in Fig. 8.

agreement with theory. We found that when we were further out than approximately 4 times the attenuation length, we could not detect the target reliably anymore. The variation around the theoretical precision can be explained by the use of a relatively small target (in pixels). Imaging it from 10 m+ resulted in very small areas that we could extract meaningful information from. Hence, the neighborhoods that were used to extract SNR and depth precision were small (3×3), which made the standard deviation calculations sensitive to outliers.

F. Contrast Enhancement Through Range Gating

In turbid waters, backscatter is known to reduce the contrast/SNR in the images because it adds a slowly spatially varying veil (a DC component) of intensity to the image. This effectively increases both the intensity and the noise of areas representing black objects and consequently reduces the apparent contrast. In Fig. 12, we show the SNR as a function of gating distance (0 to 6 m) when an object is placed 3 m from the camera. The target is a multi-albedo target (same as in Fig. 1), and the SNR is computed as the difference between a white-and-black region divided by the standard deviation of the signal in the black region. The plot shows that the SNR is highest when gating approximately 0.5 m in front of the target.

The plot also shows that the SNR stays relatively constant until the backscatter response takes off at distances closer than 1 m to the camera (see the intensity plot). From this we can interpret that we can in general average intensity images from 0.5 m in front of a target and towards the camera until we approach 1 m from the camera to increase the SNR. The camera interface is designed to be able to do this. Based on the previous depth frame, we can estimate at what distances we have objects and adjust the j and k parameters of \hat{I}_j^k accordingly to extract high-contrast intensity images.

G. Sea Trials: Schools of Fish

One important potential application of the proposed system is surveillance and monitoring of fish for sustainable farming and

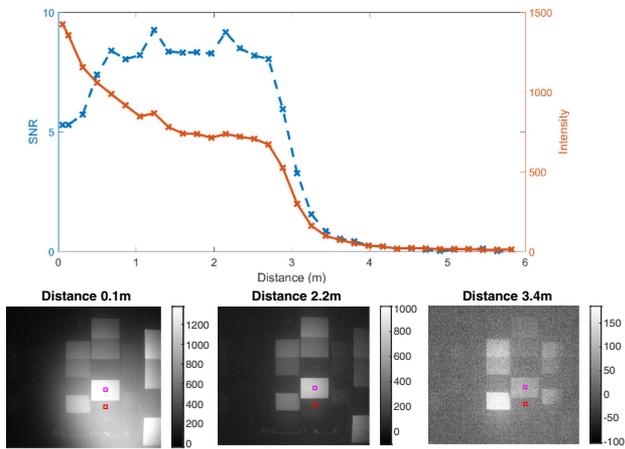


Fig. 12. SNR variation with regards to the gating distance. Top, blue curve shows the SNR with respect to gating delay. The orange curve shows the delay sweep curve of the signal (magenta square). Bottom, images from the sweep stack at 0.1 m with an SNR of 9, 2.2 m with an SNR of 9, and 3.4 m with an SNR of 1. The squares in the images illustrate where the pixel values were taken to calculate the SNR.

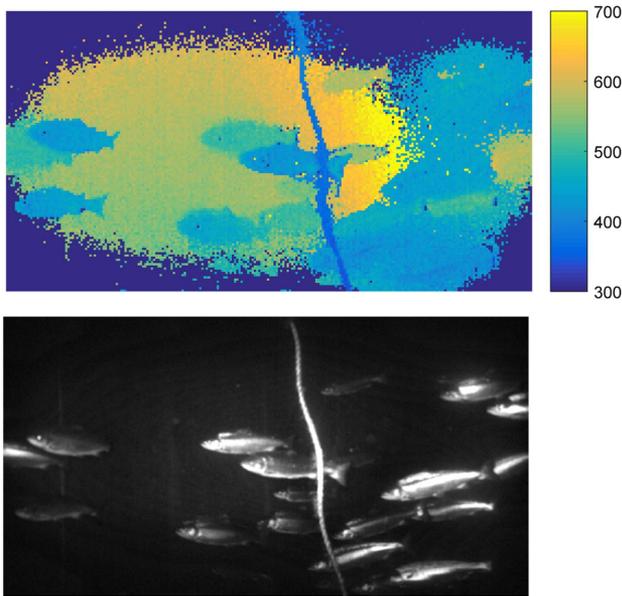


Fig. 13. Corresponding depth map (in cm) and intensity image (gated at 2 m) of a school of fish. All distances that were detected with peak heights lower than 40 were filtered out (represented as dark blue). Notice that the fish cage net is detected at 6 m, which cannot be easily seen in the histogram-equalized intensity image.

harvesting. We show here some qualitative results from imaging fish in a fish farm. The fish were salmon, approximately 50 cm long and with a mass of ~ 2 kg. We used $N_a = 4$ per range and used a step between ranges of $2\Delta_z$ that resulted in a range of (1–9 m) and a frame rate of 10 Hz. Figure 13 shows some qualitative results of a school of fish swimming.

5. DISCUSSION

We have presented an underwater range-gated system built around a customized fast-shutter CMOS camera and a solid-state actively Q-switched 532 nm laser. The system provides an effective solution for underwater imaging with a great compromise between speed of acquisition, imaging range, and resolution compared to other available underwater imaging technologies. The use of a CMOS sensor instead of a combined CCD chip and an intensifier may potentially lead to lower cost and complexity of the range-gated system.

We present an algorithm for peak detection in the range-gated signal trace that is designed to be sensitive to peaks caused by objects and suppress forward scatter and backscatter peaks. The system allows for range slicing with steps of 18 cm. To achieve super-resolution depth precision below the slicing step size, we fit a parabola to the sample points around a peak and find the analytical maximum peak position.

There is always a trade-off between frame rate, range, and depth precision. We have found a good compromise is to image 25 ranges and use four exposures per range for a total of 100 images per frame. This facilitates real-time (10 Hz) depth estimates over a range of up to 8 m, and depending on the SNR, down to a depth resolution of 1 cm. The empirical performance results agree with theoretical performance predictions that we present. We also show that we are able to estimate the distance of objects with high albedos at distances of at least 4.5 times the attenuation length of the water in low attenuation length situations.

Forward scatter may cause signal peaks in pixels where the corresponding pixel ray does not intersect any objects. An unresolved issue is how to handle such forward scatter peaks, but in future works we will investigate whether the peak width can help discriminate between forward scatter and object peaks. We will also investigate whether generative Bayesian models can discriminate between these peaks in postprocessing.

Range gating is also an important technique to increase contrast in underwater imaging. We show that backscatter degrades the contrast in images, but with selective gating, e.g., based on the depth estimates, high-contrast images can be acquired. The ideal gating distance in terms of the image SNR is shown to be approximately 1 m in front of the target of interest. However, even when gating past the worst of the backscatter (1 m and outwards), there may still be backscatter present in the image, which can be detrimental to contrast when we are viewing object signals that are barely above the noise level. In future work, we will aim to estimate the backscatter profile based on 8×8 binned full delay sweeps, and subtract the backscatter from the gated image to enhance the contrast.

Even though the current system is designed for underwater use, the performance in terms of depth precision versus SNR should be comparable in air (scaled by the difference between speed of light in water/air). In a nonturbid air environment, the attenuation length would be a negligible factor and the range of the system would be limited by the r^2 falloff in illumination with distance. In turbid air environments caused by, e.g., rain, fog, snow, or smoke, the system should be just as effective as underwater to handle the scattering effects.

We believe that the presented range-gated system is suitable for a wide range of underwater surveillance and monitoring applications. Specifically, we believe that it is ideal for monitoring marine habitats and estimating biomass and generally for underwater surveillance. We have shown qualitative results from imaging schools of fish, and future work will involve using the system to extract measurements (length, estimates of weight, swimming speed) of fish in both fish farms and the wild.

APPENDIX A: ATTENUATION MEASUREMENT TOOL

In this section, we provide a description of the instrument we developed to measure the optical attenuation in different waters. The instrument can be seen in Fig. 14. The requirements were that we needed to have a simple instrument that could quickly provide us with a measurement of water quality. As the water quality may change along the water column, the instrument needed to be submersible to 70 m (the length of the Ethernet cord of the range-gated camera) and provide rapid updates of the measured attenuation length.

Optical attenuation in water is characterized by absorption and scattering. Optical transmission is described by $T(l) = \exp(-l(a + b))$, where l is the length the light has traveled, a is the absorption coefficient, and b is a scattering-loss coefficient. Both a and b are functions of the optical wavelength. The sum $c = a + b$ is called the attenuation coefficient and $1/c$ the attenuation length. Some advanced instruments can determine a and b separately, while simpler instruments, like the one described, measure only c . When just c is measured, the measured value will depend on the acceptance angle of the instrument. Therefore, the acceptance angle is often given along with the measured value.

The attenuation meter consists of a 525 nm blinking LED light source (duty cycle of 5 s) and a monochrome camera with an 8 mm focal length $f/\#1.4$ lens. The camera and light source are mounted facing each other at a distance of $l = 0.95$ m on a rigid pole. The LED is placed in a white cavity resembling an integrating sphere. There are two layers of diffusing plastic foils, separated 5 mm, at the exit. This gives a



Fig. 14. Attenuation measurement instrument. The blinking green light source can be seen in the upper left corner, while the camera is located in the lower right corner. They are mounted 0.95 m from each other.

uniform light source. The light source is circular with 37 mm diameter. The camera resolution is 1280×1024 with $5.3 \mu\text{m}$ pixel pitch, which means that the angular extent of the light source, as seen from the camera, is 2.2 deg (59 pixels) diagonally. The uniformity measured by the camera is better than 1%.

The camera streams a live video to a top-side computer through a 100 m Ethernet cable. The camera is powered via the cable (POE), while the LED is battery powered with a lifetime of 48 h. We use images in a full-duty cycle to estimate the signal (average images where light source is on and subtract the average of images where light source is off). The maximum signal value is used as the measured signal level v_m . The optical transmission can be written as $T = \exp(-lc) = v_m/v_r$, where v_r is a reference signal value found through calibration. The reference value v_r is the theoretical value measured in water with 0 attenuation. This value cannot be measured directly since there is no such water. We propose two different methods to approximate v_r in air. One approach to approximate v_r in air is by moving the light source closer to the camera (0.95 m/1.32) so that the light source has the same extent as it would have at 0.95 m in water. When using the estimated v_r at 0.95 m, we measure a transmission of 89%. 9% of this is related to increased Fresnel reflection at the polycarbonate windows ($n = 1.6$) in air ($n = 1$) as compared with water ($n = 1.32$). The remaining 2% is related to the point-spread function of the camera and lens. A more convenient way to calibrate way to v_r in air is to adjust v_r to give a reading of 8.2 m, which corresponds to 89% transmission.

Generally the signal v_r will vary with LED emission intensity and camera sensitivity, which could both vary with temperature. In use, we observe a drift in v_m of less than 1%. This results in an estimated accuracy in measured attenuation length of 1% at $\frac{1}{c} = 1$ m, 5% at $\frac{1}{c} = 5$ m, and 6% at $\frac{1}{c} = 8$ m.

In this paper, we report the ability to perform 3D measurements at at least 4.5 attenuation lengths (at low attenuation lengths), where the attenuation length is measured as described in this appendix. 4.5 attenuation lengths correspond to a transmission back and forth of $1/8100$ if all the attenuation is calculated as absorption. In lab measurements, we obtain similar performance when introducing signal transmission of $1/5500$ by using attenuation filters and a smaller camera aperture. This factor, $1/5500$, corresponds to 4.3 attenuation lengths. Because forward scattered light, within the field of view, is still good for illumination, we are able to see slightly longer than what pure absorption would have allowed. The agreement between these numbers provides an indication that the attenuation measured with our tool is adequate for predicting camera system performance in different waters.

Funding. Horizon 2020 Framework Programme (H2020) (633098).

Acknowledgment. The range-gated system was developed in the Horizon 2020 project UTOFIA (<https://www.utofia.eu/>). We would like to thank all the consortium partners for their contribution and support.

REFERENCES

1. M. D. Aykin and S. Negahdaripour, "Forward-look 2-D sonar image formation and 3-D reconstruction," in *Oceans* (IEEE, 2013), pp. 1–10.
2. F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and A. V. Rationale, "Experimentation of structured light and stereo vision for underwater 3D reconstruction," *ISPRS J. Photogramm. Remote Sens.* **66**, 508–518 (2011).
3. Q. Zhang, Q. Wang, Z. Hou, Y. Liu, and X. Su, "Three-dimensional shape measurement for an underwater object based on two-dimensional grating pattern projection," *Opt. Laser Technol.* **43**, 801–805 (2011).
4. S. G. Narasimhan and S. K. Nayar, "Structured light methods for underwater imaging: light stripe scanning and photometric stereo," in *Oceans* (IEEE, 2005), pp. 2610–2617.
5. D. McLeod, J. Jacobson, M. Hardy, and C. Embry, "Autonomous inspection using an underwater 3D LiDAR," in *Oceans* (IEEE, 2013), pp. 1–8.
6. F. Dagleish, F. Caimi, W. Britton, and C. Andren, "Improved LLS imaging performance in scattering-dominant waters," *Proc. SPIE* **7317**, 73170E (2009).
7. L. K. Rumbaugh, E. M. Bollt, W. D. Jemison, and Y. Li, "A 532 nm chaotic lidar transmitter for high resolution underwater ranging and imaging," in *Oceans* (IEEE, 2013), pp. 1–6.
8. J. Busck and H. Heiselberg, "Gated viewing and high-accuracy three-dimensional laser radar," *Appl. Opt.* **43**, 4705–4710 (2004).
9. J. F. Andersen, J. Busck, and H. Heiselberg, *Submillimeter 3-D Laser Radar for Space Shuttle Tile Inspection* (Danish Defense Research Establishment, 2013).
10. C. Tan, G. Seet, A. Sluzek, and D. He, "A novel application of range-gated underwater laser imaging system (ULIS) in near-target turbid medium," *Opt. Lasers Eng.* **43**, 995–1009 (2005).
11. B. A. Swartz, "Laser range gate underwater imaging advances," in *Oceans* (IEEE, 1994), Vol. 2, p. II-722.
12. D.-M. He, "Underwater laser-illuminated range-gated imaging scaled by 22.5 cm ns⁻¹ with serial targets," *J. Ocean Univ. China* **3**, 208–219 (2004).
13. A. Weidemann, G. R. Fournier, L. Forand, and P. Mathieu, "In harbor underwater threat detection/identification using active imaging," *Proc. SPIE* **5780**, 59 (2005).
14. A. Andersson, "Range gated viewing with underwater camera," in *Linköpings universitet, Institutionen för systemteknik* (2005).
15. B. Jutzi and U. Stilla, "Laser pulse analysis for reconstruction and classification of urban objects," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **34**, 151–156 (2003).
16. R. Olsson, I. Eriksson, J. Powell, and A. F. H. Kaplan, "Advances in pulsed laser weld monitoring by the statistical analysis of reflected light," *Opt. Lasers Eng.* **49**, 1352–1359 (2011).
17. G. Kamermann, "Laser radar," in *Active Electro-Optical Systems, The Infrared & Electro-Optical Systems Handbook* (SPIE Optical Engineering, 1993).
18. S. Chua, N. Guo, C. Tan, and X. Wang, "Improved range estimation model for three-dimensional (3D) range gated reconstruction," *Sensors* **17**, 2031 (2017).
19. P. Andersson, "Long-range three-dimensional imaging using range-gated laser radar images," *Opt. Eng.* **45**, 034301 (2006).
20. B. Jutzi and U. Stilla, "Simulation and analysis of full-waveform laser data of urban objects," in *Urban Remote Sensing Joint Event* (IEEE, 2007), pp. 1–5.
21. S. Y. Chua, X. Wang, N. Guo, and C. S. Tan, "Range compensation for accurate 3D imaging system," *Appl. Opt.* **55**, 153–158 (2016).
22. W. Xinwei, L. Youfu, and Z. Yan, "Triangular-range-intensity profile spatial-correlation method for 3D super-resolution range-gated imaging," *Appl. Opt.* **52**, 7399–7406 (2013).
23. M. Laurenzis, F. Christnacher, N. Metzger, E. Bacher, and I. Zielenski, "Three-dimensional range-gated imaging at infrared wavelengths with super-resolution depth mapping," *Proc. SPIE* **7298** 729833 (2009).
24. G. Bouquet, J. Thorstensen, K. A. Hestnes Bakke, and P. Risholm, "Design tool for TOF and SL based 3D cameras," *Opt. Express* **25**, 27758–27769 (2017).