R27

Analysis of a modular SPAD-based direct time-of-flight depth sensor architecture for wide dynamic range scenes in a LiDAR system

Preethi Padmanabhan[1], Chao Zhang[2] and Edoardo Charbon[1,2]

1. Advanced Quantum Architecture Laboratory (AQUA), EPFL, Neuchâtel, Switzerland
2. Applied Quantum Architecture Laboratory (AQUA), Delft University of Technology, The Netherlands

Email: preethi.padmanabhan@epfl.ch, c.zhang-10@tudelft.nl, edoardo.charbon@epfl.ch

Phone: +41216954413, +31626585961, +41216954409

I. Introduction

The growing need for depth sensing has significantly driven light-detection and ranging (LiDAR) systems in a wide range of applications. Direct time-of-flight (DTOF) image sensors based on time-correlated single-photon counting (TCSPC) are being actively explored for their high speed and accurate ranging capabilities over long distances and robustness to high background noise. In a DTOF sensor, depth sensing is achieved by transmitting a periodic light source (typically a pulsed laser, also considered in this work) to a target and detecting the time-of-arrival of the reflected photons by a high performance photodetector such as avalanche photodiodes (APDs), single-photon avalanche diodes (SPADs) or silicon photomultipliers (SiPMs) and supporting electronic circuitry to measure the time-of-flight. Time-to-digital converters (TDCs) are typically used for this purpose [1-3, 6-9].

There are various challenges in acquiring a quality LiDAR measurement among them being robustness to ambient light, an eye-safe illuminator, wide dynamic range targets, adverse weather conditions, interference in a multi-LiDAR scenario, etc. Ambient light suppression has been addressed by coincidence detection on chip [2, 6, 7], a well-known technique utilizing spatio-temporal closeness of photons within a laser pulse to filter out background noise photons. While TCSPC with coincidence detection have shown effective noise-filtering properties, imaging in a wide dynamic range scenario is an ongoing challenge. The work in [7, 9] addresses this by implementing variable coincidence thresholds. In this paper, we address the issues of a wide-dynamic range scene by proposing a modular SPAD-based DTOF sensor in a shared architecture based on coincidence and pixel clustering. While keeping modularity and sharing at the core of the sensor architecture [3, 4], this paper extends it towards a more robust solution to DTOF-based sensors by proposing:

1. tunable coincidence detection based on photon activity including the address/ID map of the contributing pixels within a coincidence event,
2. multiple timestamping within a shared architecture and enhancing data throughput,
3. modular gating mechanism to configure pixels around the range of interest, all of which can be performed simultaneously.

The proposed DTOF scheme is studied in a flash LiDAR setup by developing an analytical model on MATLAB supported by simulation results explaining the advocated concepts.

II. Flash LiDAR System

In a flash LiDAR the target scene ($target\_area$) is uniformly illuminated with a wide angle laser beam, shown in Figure 1 with horizontal and vertical fields of view (FOV) denoted as 'θ_H' and 'θ_V' respectively. The distance to a target plane is denoted as 'd'. The back-reflected photons from the target are then detected by the SPAD sensor collected through a

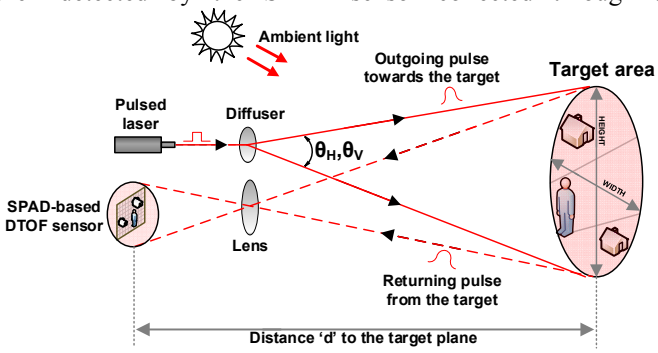

Figure 1: Conceptual Flash LiDAR system diagram.

receiving lens with a diameter, D = 21mm, f-number of 1.4 and aperture area, $A_{aperture}$. Due to the fact that the entire scene is illuminated at once, a sensor with an array of pixels is preferred in order to reconstruct the target scene with a given spatial resolution [8, 9]. Figure 2 shows the simulation results on a SPAD array with a spatial resolution of 32×32 pixels used to detect photons reflected from a 40% reflectivity (r) flat target (Lambertian) at distances varying over 1-50m and θ_H and θ_V being 45° each. The average power, P_{avg}, of the laser is 10mW at a wavelength (λ) of 905nm (preferred choice for LiDAR) and repetition rate (f_{laser}) of 2MHz is assumed. The ambient light condition is assumed to be 50klux. A bandpass filter with a full-width-at-half-maximum (FWHM) of 15nm centered around 905nm has been considered. A sub-micron SPAD device with a photon detection probability (PDP) of 5% at 905nm, a fill factor (FF) of 50%, a temporal jitter of 100ps FWHM and a dead time (t_d) of 8ns has been considered for simulations. The effect of dark count rate (DCR) of the SPADs has not been considered for simulations here, given that it is much lower than the ambient light activity which is a dominant source of noise. The number of events being detected per pixel per second are estimated for signal as well as the noise photons using standard equations as shown below [10-12]. For the returning signal photons,

$$\#signal\_rate = \frac{P_{avg}}{E_{photon}} \times r \times tf \times \frac{1}{2\pi d^2} \times \frac{1}{\#pixels} \times PDE \times A_{aperture} \,. \quad (1)$$

Here, E_{photon} is the energy of a photon (hc/λ) at $\lambda = 905nm$ and tf, is the transmission factor representing the lens efficiency (80%), PDE is the photon detection efficiency

(PDP×FF). Similarly, the number of events detected per pixel per second from the background noise photons are given by,

$$\#noise\_rate = BG\_flux \times r \times tf \times \frac{1}{2\pi d^2} \times target\_area \times \frac{1}{\#pixels} \times PDE \times A_{aperture}. \quad (2)$$

BG_flux is the background solar power density on the target area, $4d^2\tan(\theta_H/2)\,(\theta_V/2)$, and the solar spectral irradiance is used to calculate the effective power density for a 50klux ambient light condition [10, 13]. The resulting count rates indicated as events detected per second per pixel are shown in Figure 2; no particular noise filtering mechanism has been modelled for this simulation. As can be seen, from 4m, the system starts approaching a negative SBR regime, exemplifying the requirement to have noise filtering circuits on the sensor level.

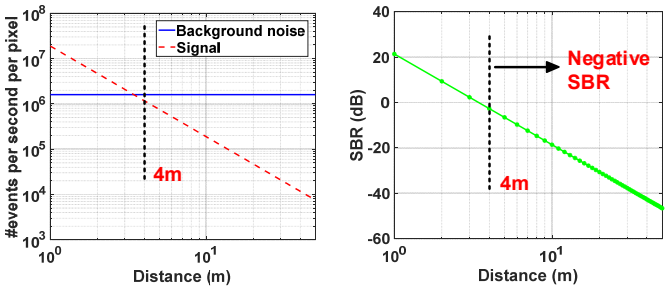

Figure 2: Simulation results of the number of events per pixel per second (assuming 32×32 array) and the SBR for 1-50m target distances; signal-to-background noise ratio (SBR); background noise is modelled with Planck's law of blackbody radiation and Poisson statistics.

III. ANALYTICAL MODEL OF THE PROPOSED SENSOR ARCHITECTURE

A. Proposed DTOF sensor scheme

Figure 3 shows the generic block diagram of a module of the proposed DTOF scheme with an example case of a sensor with 32×32 SPAD pixels. The module is visualized as a subgroup comprising of 8x8 SPADs. The subgroup size governs the effective activity rate for a given incoming photon flux from the target and the ambient light and thus, has been chosen accordingly [4]. Two subgroups share a single always-on time-to-digital converter (TDC) with independent lines for timestamping in the respective subgroups. Within the subgroup is a combination tree, which we call as the "*coincID*" tree. This tree is modelled to propagate events and perform "*coincidence*" along with generating the "*address or ID*" of the SPADs which contributed to an event. The propagated event (DTOF_sample in Figure 3) through the subgroup is passed to a shared TDC to sample the internal states generating a full timestamp of the first incoming photon within a coincidence event. The subgroup is further clustered into $\#N$ minigroups comprising of $\#M$ pixels each. Every minigroup also provides a coarse timestamp along with generating the spatial address/ID of the contributing events at the minigroup level. In addition to performing coincidence, every subgroup is modelled to optionally operate under time-gating set electrically around the desired target range. As mentioned in the figure, there are four different modes of operation with coincidence and gating. Thus, based on the chosen mode, a valid signal is generated when the mode-dependent condition is satisfied in order to write the data into the FIFO register. After every detection event, the *coincID* tree resets itself, making it available for the successive detection. The local digital processing and communication units (DPCUs) in the subgroup keep track of the number of photons detected within a coincidence window, allowing us to set different thresholds in every subgroup.

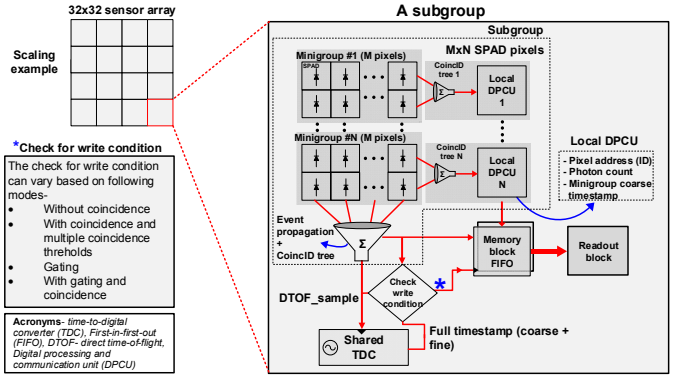

Figure 3: Generic block diagram of the proposed DTOF scheme with an example case of 32x32 sensor array.

B. Analysis of signal detection in the proposed scheme

For the proposed subgroup in a sensor array of 32×32 pixels, effective count rates for noise and signal photons, eff_noise_pixel and eff_signal_pixel respectively, are evaluated per pixel considering a non-paralyzable model [14]. Let eff_noise_sg and eff_signal_sg indicate the cumulative noise and signal count rates within a subgroup, sg, and similarly, let eff_noise_mg and eff_signal_mg indicate the cumulative rates, within a minigroup, mg. The effective number of photons within one laser pulse can then be estimated for a pixel, minigroup and subgroup. For the noise photons per pixel which are distributed uniformly in time,

$$\#noise\_pixel = eff\_noise\_pixel \times \left(t_{window} - th \times {}^{t_d}/_{64}\right). \quad (3)$$

where t_{window} is the length of the coincidence window, th is the coincidence threshold and t_d is divided by a factor of the number of pixels in the subgroup = 64, to account for the effective dead time in the subgroup. For signal photons per pixel which are concentrated within the pulse width of the laser,

$$\#signal\_pixel = eff\_signal\_pixel \times \left({}^{1}/_{f_{laser}} - th \times {}^{t_d}/_{64}\right). \quad (4)$$

The cumulative number of photons per pulse can similarly be estimated for minigroup and subgroup, represented as $\#noise\_mg$, $\#signal\_mg$, $\#noise\_sg$ and $\#signal\_sg$. For every minigroup, mg, with $\#M$ pixels, the cumulative number of noise photons per laser period is

$$\#noise\_mg = \sum_1^M \#noise\_pixel. \quad (5)$$

For every subgroup, sg, with $\#N$ minigroups, the cumulative number of noise photons per laser period is

$$\#noise\_sg = \sum_1^N \#noise\_mg. \quad (6)$$

The cumulative number of signal photons for minigroup and subgroup are similarly calculated. The probabilities of photon detection can be modelled as a Poisson arrival process with probability of detecting k photons being,

$$\Pr(k) = \frac{(rate\times t)^k \exp(-rate\times t)}{k!}. \quad (7)$$

Here, $rate$ is the average number of photons per second and t is the desired time interval [15]. In the coincidence mode, the probability of detecting noise photons satisfying coincidence within the subgroup can be estimated as,

$$Pr_{noise=th} = Pr_{noise,pixel} \times Pr_{noise,sg-pixel}, \quad (8)$$

where $Pr_{noise,pixel}$ is probability of detecting a noise photon in a given pixel,

$$Pr_{noise,pixel} = \#noise\_pixel \times \exp(-\#noise\_pixel). \quad (9)$$

$Pr_{noise,sg-pixel}$ is the probability of detecting $\#(th-1)$ photons in the rest of the subgroup, which is calculated using $\#noise\_mg$ and $\#noise\_sg$. For a given pixel belonging to a subgroup, sg, the probability of detecting $\#(th-1)$ photons in the minigroup, mg, to which a pixel belongs is

$$Pr_{noise,mg} = \frac{\#noise\_mg^{(th-1)}\times(\exp(-\#noise\_mg)}{(th-1)!}. \quad (10)$$

The cumulative number of photons in the rest of the subgroup, except the given mg, can be calculated as,

$$\#noise,(sg-mg) = \#(noise\_sg - noise\_mg) \quad . \quad (11)$$

Now, using this, the probability of detecting $\#(th-1)$ photons can be deduced as,

$$Pr_{noise,(sg-mg)} = \frac{\#noise,(sg-mg)^{(th-1)}\times\exp(-\#noise,(sg-mg))}{(th-1)!} \quad . \quad (12)$$

The probability $Pr_{noise,sg-pixel}$ is then a union of (10) and (12) given that the detection of photons in the minigroup and the rest of the subgroup is independent of each other since every minigroup is modelled to have its own timestamping circuitry.

$$Pr_{noise,sg-pixel} = Pr_{noise,mg} \cup Pr_{noise,(sg-mg)}. \quad (13)$$

Similarly, the probability of detecting only signal photons satisfying coincidence within the subgroup can be estimated as,

$$Pr_{signal=th} = Pr_{signal,pixel} \times Pr_{signal,sg-pixel}. \quad (14)$$

The final conditional probability of detecting a signal photon per pixel in the presence of noise photons can then be calculated as follows,

$$Pr_{signal|noise} = (1 - Pr_{noise,pixel}) \times (1 - Pr_{noise=th}) \times (Pr_{signal=th}). \quad (15)$$

In the no-coincidence mode, this conditional probability is estimated only for a single photon. We can determine the detected noise and signal events for a given number of laser frames, $\#lasers_{frame}$ as,

$$\#noise_{th} = Pr_{noise=th} \times \#lasers_{frame}. \quad (16a)$$

$$\#signal_{th} = Pr_{signal|noise} \times \#lasers_{frame}. \quad (16b)$$

From (16a, 16b), we can now compute the histogram around the target TOF, related to the target distance, d ($d = c \times TOF/_2$). The system temporal jitter, $jitter_{total}$ is the total contribution from the laser jitter, SPAD temporal jitter and the jitter from electronic circuitry such as TDC (resolution T_{LSB}), combination tree and alike. The mean value of noise can be modelled given its uniform distribution in time over the interval, $1/f_{laser}$,

$$noise(t) = \frac{\#noise_{th}}{1/f_{laser}} \times jitter_{total}. \quad (17)$$

The mean value of signal events around the target can be modelled considering the Gaussian shape of the laser pulse, from the error function of a Gaussian distribution as,

$$signal(t) = \frac{\#signal_{th}}{\sqrt{2\pi\times jitter_{total}}} T_{LSB} \times \exp\left(\frac{-(t-TOF)^2}{2\times jitter_{total}{}^2}\right). \quad (18)$$

The final histogram can then be modelled as a combination of (17) and (18) [13].

IV. Simulation Results

The proposed DTOF sensor scheme is modelled analytically on MATLAB and various probabilities of detection as discussed in previous section are computed by extending to an array of 32×32 pixels. Figure 4a, 4b show a 32×32 resolved image measured in a scanning LiDAR setup from our previous work [3]. This measured image is used as an input target scene to evaluate the feasibility of the proposed DTOF sensor scheme.

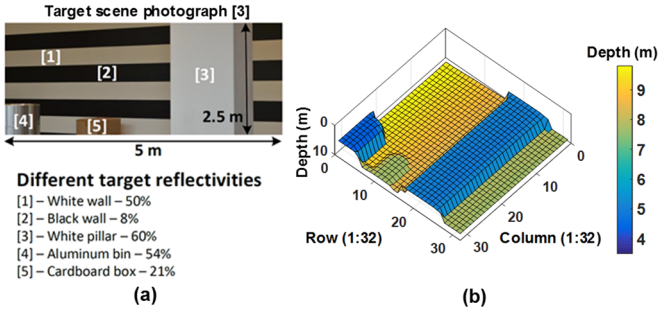

Figure 4: (a) Photograph of the target scene (b) shows measured result of the target in a scanning LiDAR setup [3].

There are two schemes modelled- scheme 1 is a sensor model which does not have any particular noise filtering mechanism and uses a "winner-take-all" approach to propagate events and generate timestamps; scheme 2 is a sensor model of the proposed DTOF scheme shown in Figure 3. The targets in Figure 4a, 4b range in between 4m and 9m. Target reflectivities, r, and target distances, d, are the extracted quantities from this figure used for simulation results shown through Figures 5-8.

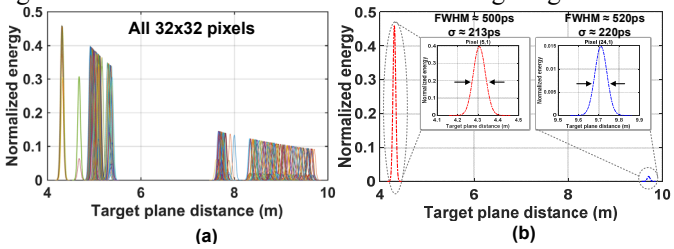

Figure 5: (a) Normalized energy of the laser pulse over varying target distances of all pixels; (b) Maximum and minimum energies.

By modelling the illuminated laser pulse as a Gaussian distribution, the normalized energy of the laser pulse (Figure 5) over varying target distances is plotted for the target scene introduced in Figure 4. Figure 5b shows the simulated maximum and the minimum energy observed at the nearest (≈ 4.3m) and the farthest (≈ 9.7m) target distance plane respectively. The FWHM value of the jitter (2.355σ) observed is between 500ps to 520ps at 4.3m and 9.7m respectively. Scheme 1 is first simulated analytically in the presence of a 5klux ambient light condition considering 100 laser pulses in a frame (Figure 6a).

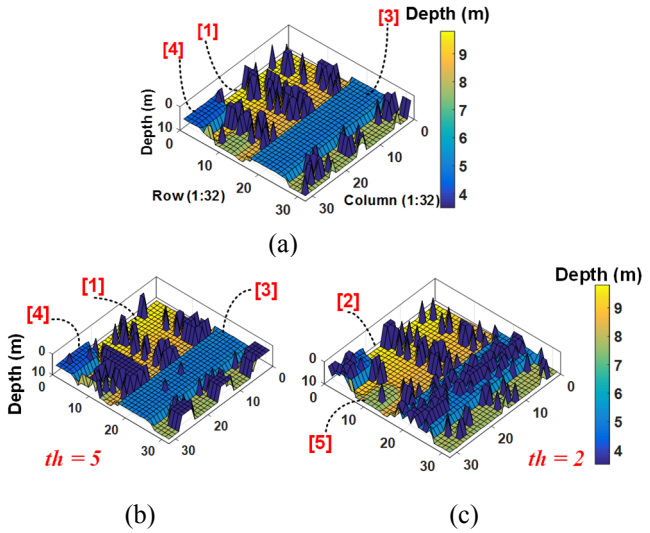

Figure 6: Simulation results in a flash LiDAR model (a) the target scene in scheme 1 (no coincidence); (b), (c) illustrate reconstruction using single coincidence thresholds.

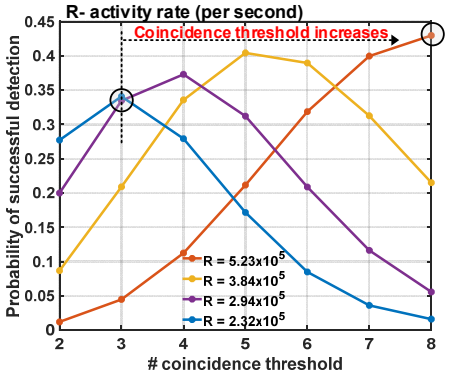

Figure 7: The relationship between coincidence threshold and the incoming activity rate received per second.

The histogram is computed for every pixel and the combined 32×32 image is then reconstructed by taking the peak values on every computed histogram. Since, scheme 1 does not include any noise filtering mechanism, one can observe the expected degradation in the reconstructed (simulated) image in Figure 6a compared to the measured image in Figure 4b, taken under minimal noise in a scanning setup. We also observe that the white wall (object [1]), aluminium bin (object [4]) and the white pillar (object [3]) are readily reconstructed in comparison to the rest of the scene where there is up to 11.9% incorrect sampling.

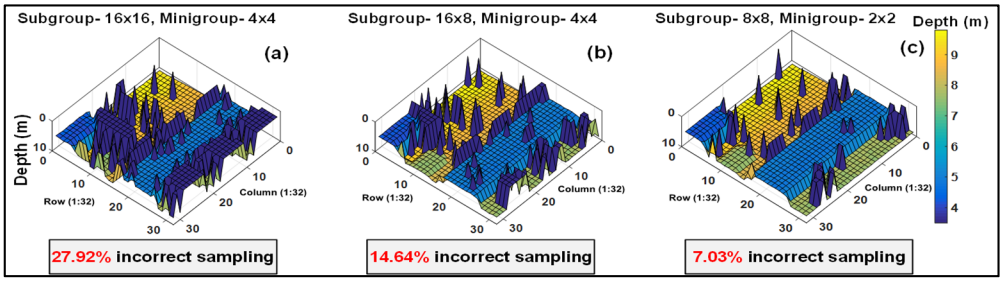

Figure 8: The proposed grouping scheme is illustrated for different subgroup and minigroup sizes.

Figure 6b, 6c show the simulation results of scheme 2 under 50klux background noise and illustrate the problem with single coincidence thresholds. In Figure 6b, with $th = 5$, we see objects [1], [3] and [4] being more accurately reconstructed compared to rest of the scene owing to their higher relative reflectivities. Similarly, in Figure 6c, with $th = 2$, objects [2] and [5] take preference over the rest of the scene. Figure 7 in fact shows how coincidence threshold, th, increases for increasing photon activity, R, to provide a successful detection (temporal error (σ) < 230ps), implying that a single threshold cannot yield an accurate reconstruction of a scene and a higher (lower) reflective target will imply higher (lower) photon activity requiring higher (lower) coincidence thresholds. In the proposed DTOF scheme, every subgroup can have a unique threshold. The proposed grouping scheme is illustrated for different subgroup and minigroup sizes in Figure 8. As can be seen through (a) to (c), a subgroup of 8x8 for coincidence and minigroup of 2×2 in (c) allows improved 3D reconstruction (only 7% incorrect sampling) by performing coincidence within a subgroup of 8 × 8 SPADs and additionally, enabling simultaneous TOF measurement within the 16 minigroups containing 2×2 pixels each, particularly useful for multi-target scenario.

V. CONCLUSIONS AND FUTURE WORK

An alternative DTOF sensor architecture was proposed in this paper with emphasis on wide dynamic range target scenes. The modularity and configurability in terms of pixel clustering within the subgroups along with coincidence will allow us to explore low-high FOVs for 3D imaging. The feature of providing multi-pixel timing, counting and ID information within the minigroups allows exploitation of this data for efficient photon-by-photon image processing within a shared architecture [5]. While analytical modelling conceptually validates the proposed concepts, next step would be to build a simulator based on Monte Carlo ray tracing models to predict the behavior realistically in a LiDAR system with the proposed sensor architecture. In parallel, another step will include a CMOS implementation of the proposed sensor to test the concepts in a real LiDAR scenario.

ACKNOWLEDGMENT

The authors would like to thank A.R. Ximenes for useful discussions on DTOF sensor architectures and trade-offs.

REFERENCES

[1] C. Veerappan et al., ISSCC 2011.
[2] C. Niclass et al., JSSC 2013.
[3] A.R.Ximenes et al., ISSCC 2018.
[4] A.R.Ximenes et al., Sensors 2018.
[5] R. Joshua et al., TCI 2017.
[6] D. Portaluppi et al., JSTQE 2018.
[7] M. Beer et al., Sensors 2018.
[8] S. Lindner et al., VLSI 2018.
[9] R.K. Henderson et al., ISSCC 2019.
[10] S. Gnecchi et al., IISW 2017.
[11] P McManamon et al., OE 2012.
[12] PF McManamon et al., OE 2017.
[13]C. Niclass et al., Imaging based on single photon detectors, 2007.
[14] E. Sarbazi et al., PIMRC 2015.

PATENT APPLIED- Coincidence-based direct-time-of-flight sensor architecture for wide dynamic range scenes in a LiDAR system, 2019.