Enhanced single-photon time-of-flight 3D ranging

Rudi Lussana[1,*], **Federica Villa**[1], **Alberto Dalla Mora**[2], **Davide Contini**[2], **Alberto Tosi**[1] **and Franco Zappa**[1]

[1] *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy*
[2] *Dipartimento di Fisica, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy*
[*]*rudi.lussana@polimi.it*

Abstract: We developed a system for acquiring 3D depth-resolved maps by measuring the Time-of-Flight (TOF) of single photons. It is based on a CMOS 32×32 array of Single-Photon Avalanche Diodes (SPADs) and 350 ps resolution Time-to-Digital Converters (TDCs) into each pixel, able to provide *photon-counting* or *photon-timing* frames every 10 μs. We show how such a system can be used to scan large scenes in just hundreds of milliseconds. Moreover, we show how to exploit TDC unwarping and refolding for improving signal-to-noise ratio and extending the full-scale depth range. Additionally, we merged 2D and 3D information in a single image, for easing object recognition and tracking.

References and links

1. X. Wei, S. L. Phung, and A. Bouzerdoum, "Pedestrian sensing using time-of-flight range camera," Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society, 43–48 (2011).
2. E. Kollorz, J. Penne, J. Hornegger, and A. Barke, "Gesture recognition with a time-of-flight camera," Int. Journal of Intelligent Systems Technologies and Applications, **5**(3), 334–343 (2008).
3. Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt, "3D shape scanning with a time-of-flight camera," Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), 1173–1180 (2010).
4. W. Becker, A. Bergmann, M. A. Hink, K. König, K. Benndorf, and C. Biskup, "Fluorescence lifetime imaging by time-correlated single-photon counting," Microscopy Research and Technique, **63**(1), 58–66 (2004).
5. X. Michalet, R. A. Colyer, G. Scalia, A. Ingargiola, R. Lin, J. E. Millaud, S. Weiss, Oswald H. W. Siegmund, Anton S. Tremsin, John V. Vallerga, A. Cheng, M. Levi, D. Aharoni, K. Arisaka, F. Villa, F. Guerrieri, F. Panzeri, I. Rech, A. Gulinatti, F. Zappa, M. Ghioni, and S. Cova, "Development of new photon-counting detectors for single-molecule fluorescence microscopy," Phil. Trans. R. Soc. B, **368**(1611):20120035 (2012).
6. D. O'Connor and D. Phillips, "Time-correlated single photon counting", Academic Press, London (1984).
7. M. M. Ter-Pogossian, A. Nizar, D. C. Ficke, J. Markham, and D. L. Snyder, "Photon time-of-flight-assisted positron emission tomography," J. Computer Assisted Tomography, **5**(2), 227–239 (1981).
8. S. Cova, M. Ghioni, A. Lacaita, and C. Samori, "Avalanche photodiodes and quenching circuits for single-photon detection," Applied Optics, **35**(12), 1956–1976 (1996).
9. F. Villa, R. Lussana, D. Bronzi, S. Tisa, A. Tosi, F. Zappa, A. Dalla Mora, D. Contini, D. Durini, S. Weyers, and W. Brockherde, "CMOS imager with 1024 SPADs and TDCs for single-photon timing and 3D time-of-flight," IEEE J. of Selected Topics in Quantum Electronics, **20**(6):3804810 (2014).
10. F. Villa, B. Markovic, S. Bellisai, D. Bronzi, A. Tosi, F. Zappa, S. Tisa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, "SPAD smart-pixel for time-of-flight and time-correlated single-photon counting measurements," Photonic Journal, **4**(3), 795–804 (2012).
11. G. Intermite, R. E. Warburton, A. McCarthy, X. Ren, F. Villa, A. J. Waddie, M. R. Taghizadeh, Y. Zou, F. Zappa, A. Tosi, and G. S. Buller, "Enhancing the fill-factor of CMOS SPAD arrays using microlens integration," Proceedings SPIE 9504, Photon Counting Applications 2015, doi:10.1117/12.2178950 (2015).
12. D. Tamborini, B. Markovic, F. Villa, and A. Tosi, "16 channel module based on a monolithic array of single photon detectors and 10 ps time-to-digital converters," IEEE Journal of Selected Topics in Quantum Electronics, **20**(6):3802908 (2014).

13. C. Niclass, C. Favi, T. Klute, M. Gersbach, and E. Charbon, "A 128 x 128 single-photon image sensor with column-level 10-Bit time-to-digital converter array," IEEE Journal of Solid-State Circuits, **43**(12), 2977–2989 (2008).
14. C. Veerappan, J. Richardson, R. Walker, D.-U. Li, M.W. Fishburn, Y. Maruyama, D. Stoppa, F. Borghetti, M. Gersbach, R. K. Henderson, and E. Charbon, "A 160x128 single-photon image sensor with on-pixel 55ps 10b time-to-digital converter," International Solid-State Circuits Conference, 312–314 (2011).
15. M. Gersbach, Y. Maruyama, E. Labonne, J. Richardson, R. Walker, L. Grant, R. Henderson, F. Borghetti, D. Stoppa, and E. Charbon, "A parallel 32×32 time-to-digital converter array fabricated in a 130 nm imaging CMOS technology," Proceedings of the 35th European Solid-State Circuits Conf., 196–199 (2009).
16. R. Lange and P. Seitz, "Solid-state time-of-flight range camera," IEEE Journal of Quantum Electronics, **37**(3), 390–397 (2001).
17. D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockerde, "100,000 frames/s 64x32 single-photon detector array for 2D imaging and 3D ranging," IEEE Journal of Selected Topics in Quantum Electronics, **20**(6):3804310 (2014).
18. A. M. Wallace and G. S. Buller, "3D imaging and ranging by time-correlated single photon counting," Computing & Control Engineering Journal, **12**(4), 157–168 (2001).
19. Hewlett-Packard, inc., "Time interval averaging," Application Note 162-1 (USA).
20. F. Zappa, S. Tisa, A. Tosi, and S. Cova, "Principles and features of single-photon avalanche diode arrays," Sensors and Actuators A: Physical, **140**(1), 103–112 (2007).
21. S. Tisa, F. Zappa, A. Tosi, and S. Cova, "Electronics for single photon avalanche diode arrays," Sensors and Actuators A: Physical, **140**(1), 113–122 (2007).
22. F. Villa, D. Bronzi, Y. Zou, C. Scarcella, G. Boso, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, U. Paschen, and W. Brockherde, "CMOS SPADs with up to 500 μm diameter and 55% detection efficiency at 420 nm," Journal of Modern Optics, **61**(2), 102–115 (2014).
23. A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernández, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," Applied Optics, **48**(32), 6241–6251 (2009).
24. M. Wahl, "Time-correlated single photon counting," PicoQuant GmbH Technical Note (2014).
25. S. Kurtti and J. Kostamovaara, "Pulse width time walk compensation method for a pulsed time-of-flight laser rangefinder," I2MTC 2009 – Int. Instrum. and Meas. Technology Conf., Singapore, doi:10.1109/imtc.2009.5168610 (2009).
26. D. R. Reilly and G. S. Kanter, "High speed lidar via GHz gated photon detector and locked but unequal optical pulse rates," Opt. Express 22 (13), 15718-15723 (2014).
27. P. A. Hiskett, C. S. Parry, A. McCarthy, and G. S. Buller, "A photon-counting time-of-flight ranging technique developed for the avoidance of range ambiguity at GHz clock rates," Optics Express, **16**(8), 13685–13698 (2008).
28. H. E. Andersen, S. E. Reutebuch, and R. J. McGaughey, "A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods," Can. J. Remote Sensing, **32**(5), 355–366 (2006).

1. Introduction

During the last few decades, the capability to measure time intervals with sub-nanosecond resolution opened the way to the acquisition of 3D depth-resolved images, with sub-centimeter precision, through the measurement of the Time-of-Flight (TOF) of a laser pulse shone toward objects in the scene and returning back to a detector. Nowadays, TOF-based 3D-ranging systems are becoming widespread in several fields of science and everyday life, such as automotive safety applications [1], gesture recognition for user interaction [2], 3D scanners for virtual prototyping [3], long-range security surveillance. Single-photon sensitive detectors provide the extremely valuable chance to drastically reduce the illumination power compared to traditional detectors (i.e. to improve the system's eye safety), thus increasing both achievable maximum depth-range and frame-rate.

Even many other applications, including Fluorescence Lifetime Imaging Microscopy (FLIM) [4] and Fluorescence Correlation Spectroscopy (FCS) [5], profitably make use of *photon-timing* (i.e. TOF) information to reconstruct very faint (single-photon level) and very fast (tens of picosecond) optical signals by means of Time-Correlated Single-Photon Counting (TCSPC) [6] technique. Others are exploiting TOF and single photon sensitivity to measure other events, such as scintillations for time-resolved Positron Emission Tomography [7].

In this paper, we present the results achieved by a camera based on 32×32 SPAD (Single-Photon Avalanche Diode [8]) detectors and a 350 ps resolution TDC (Time-to-Digital Converter) integrated into each pixel. The array is able to detect single-photons with a frame-rate as high as 100,000 fps. The chip offers the double feature to measure both the light signal intensity by counting the single-photons detected within a frame (*photon-counting* mode) and also the arrival time of the first photon per pixel in each frame (*photon-timing* mode). We present the merging of 2D *photon-counting* images and 3D *photon-timing* maps into combined frames, in order to augment image information. Furthermore, we show how to mitigate intrinsic limitations of TDCs, like short Full-Scale Range (FSR), refolding, and long frame-time, extending the signal-to-noise ratio and decreasing the required optical pulse energy. Finally, we demonstrate the fast scanning of large scenes, for increasing the effective number of pixels, and the detection of objects hidden by front obstacles, thanks to the *photon-counting* sensitivity and the *photon-timing* TOF information of the camera.

Section 2 describes the camera architecture. *Section 3* shows how to exploit unwarping and refolding to improve TDC performance. *Section 4* presents how to merge 3D and 2D data and, also, scan the target, in order to augment image richness. Eventually, *Section 5* draws conclusions and perspectives.

2. Camera architecture

We described the detailed microelectronic design of the SPAD chip in [9]. It consists of an array of 32×32 smart-pixels and global on-chip electronics for clocking and read-out, fabricated in a 0.35 μm CMOS process. Each pixel includes a 30 μm diameter SPAD with very good noise performance if compared to the current state-of-art in CMOS SPADs (120 dark counts/s per pixel at 35 °C, corresponding to just 0.17 Hz/μm^2 dark count rate density) and a high-linearity 350 ps resolution TDC [10]. A microlens array is also under development [11], in order to mitigate the effects of the low fill-factor of 3.14%, caused by the integration of one TDC per pixel. Other similar chips can be found in literature, with either a lower number of pixels [12], a multiplexed TDCs architecture (which limits the maximum frame-rate) [13], or a much higher dark count rate density [14-15].

The chip is readout by means of a commercial *Opal-Kelly XEM6010* board, which includes a Xilinx Spartan-6 FPGA, 128 MiB SDRAM, and a USB 2.0 interface. We developed a custom interface board, for synchronizing the camera to an external pulsed laser, and a power supply board. Fig. 1 shows the overall assembly and the compact ($6 \times 6 \times 8$ cm^3) housing; the only external components needed for 3D TOF ranging are a sub-nanosecond pulsed laser and a Personal Computer, as shown in Fig. 2.
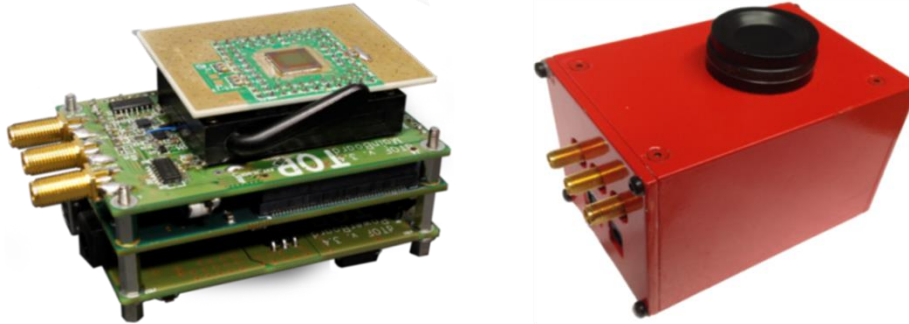

Fig. 1. Boards assembly (left) and housing (right) of the SPAD camera. Note the SPAD chip on top, the intermediate interface board with SMA connectors, the FPGA digital processing board underneath with USB, and the power supply board at the bottom. The metallic box is provided with a standard C-Mount connector for lenses.
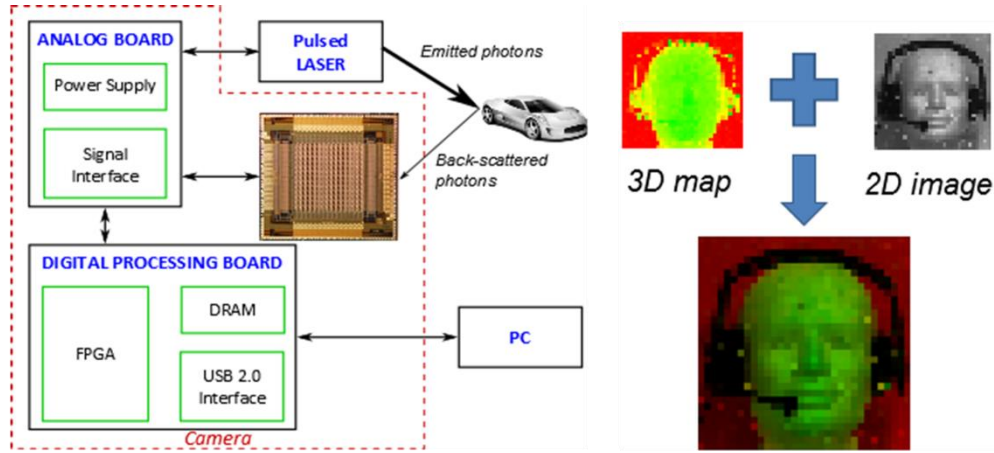
Fig. 2. Typical Time-of-Flight 3D ranging setup (left): thanks to the single-photon sensitivity of the SPAD array chip, just some of the many photons emitted from a pulsed laser need to be reflected back by the target and be detected. Each frame consists of 1024 (i.e. 32×32) data of either 2D "intensity" (*photon-counting*) or 3D "ranging" (*photon-timing*) maps of the objects in the scene. Thanks to post-processing, it is possible to merge 2D and 3D information (right).

The chip can be operated in two different modes: *photon-counting* for 2D "intensity" acquisitions and *photon-timing* for either 3D TOF ranging or TCSPC (e.g. fluorescence time decays) measurements. In *photon-counting* mode, an in-pixel 6-bit counter allows to count up to 63 photons per frame and per pixel, with a frame-rate as high as 100,000 fps (given a readout time of 9.7 ns per pixel). Such 2D movies provide frames with "grey" levels proportional to objects' brightness (both spontaneously emitted and reflected, due to laser illumination); the FPGA can further improve dynamic range by adding together consecutive frames, while lowering frame-rate. In *photon-timing* mode, each in-pixel 10-bit TDC computes the time delay between the local START signal (triggered by the detection of a photon in that pixel) and the global STOP signal (synchronous with the emitted laser pulse). In this way, it is possible to reconstruct the photons' TOF, hence the distance between camera and targets. Thanks to this "reversed" START-STOP configuration, TDC conversions only take place when a photon is detected, therefore reducing power dissipation. In both modes, dead times are minimized thanks to the possibility to perform readout concurrently with the acquisition of a new frame.

Compared to phase-resolved (also called "indirect" TOF) ranging systems, based just on *photon-counting* (see for example [16] and [17]), this "direct" TOF approach does not need to reconstruct the full optical waveform, and therefore requires less photons to perform the measurement; the drawbacks are the necessity for a more expensive sub-nanosecond pulsed laser and a more sophisticated in-pixel *photon-timing* electronics. We operated the in-pixel TDCs in order to provide an LSB (Least Significant Bit) of about 5.6 cm distance resolution (i.e. LSB = 350 ps), while reaching a maximum Full-Scale Range (FSR) of 55 meters (i.e. FSR = 360 ns), with a 10-bit discretization. Note that the FSR does not limit the maximum distance to be within 55 m from the camera. In fact, with a priori coarse knowledge of the absolute target distance, even the 3D profile of objects located kilometers away from the camera can be easily observed and ranged, by properly instructing the FPGA to introduce a time-delay between laser excitation and TDC global STOP trigger. Instead, FSR only limits the depth-range to be shorter than 55 m at the target distance. A simple scanning along the depth-axis (e.g. by using a programmable delayed trigger) can also be employed to reach longer measurement range, at the expense of frame-rate. Furthermore, FSR can be also increased by reducing system's main clock frequency (otherwise at 180 MHz), at the expense

of correspondingly larger (i.e. poorer) LSB resolution: for example, we also tested the camera at 50 MHz, resulting in LSB = 1.25 ns and FSR = 1.28 μs, i.e. 19 cm resolution over a 190 m range.

In principle, each pixel could provide the distance information with just one detected photon. However, both dark counts of SPADs (due to ignitions caused by electron-hole thermal generations or tunneling) and background light can trigger spurious TDC conversions. Given the impossibility to distinguish between signal photons and background events (both photons and noise ignitions), a set of repetitive measurements is required. By exploiting the TCSPC technique [18] it is possible to repeat the pulsed laser excitation, to record each TOF measurement, and to build a histogram of arrival times. Thereafter, the computation of the histogram centroid, t_{TOF}, provides the average TOF, hence the distance information $d=c \cdot t_{TOF}/2$, where c is the speed of light. Since background events are spread all over the 1024 bins and contribute as a flat distribution to the histogram, we experimentally verified that even with background intensity much higher than the useful signal (up to 100 times), the measurement uncertainty is limited to less than 20% LSB (i.e., less than 1 cm).

The number of frames to average (i.e. the number of data to accumulate in the histogram) must be chosen as a trade-off between final desired frame-rate, actual signal intensity, background level, and desired final depth precision. The latter can be estimated by means of the well-known equation $\sigma_{FINAL} = \sigma_{SINGLE-SHOT} / \sqrt{N}$ [19], where $\sigma_{SINGLE-SHOT}$ is the single-shot precision and N the number of valid frames, i.e. the number of frames where a signal photon was detected by that pixel. Fig. 3 shows an example of a TOF histogram for a single pixel with a strong background compared to reflected photons (40:1) due to non-filtered ambient light. However, the impact of background shot noise is drastically reduced by increasing the number of accumulated frames, as shown in Fig. 3 where a 32× increase, from 512 (left plot) to 16,384 (right plot) frames in each histogram, leads to an increased precision by about $\sqrt{32}=5.6\times$).

Note that the single-shot precision is a combination of the SPAD time jitter [20], the front-end electronics jitter [21] and the TDC resolution. In our case, it is dominated mostly by the 350 ps LSB of the TDC, since the measured full-width at half maximum of the SPADs is better than 100 ps (as we characterized in [22]). According to the previous equation, the precision σ_{FINAL} can be indefinitely increased by computing histograms on higher number of frames: for instance, with 350 ps single-shot precision and N = 10,000 frames (i.e. 100 ms total acquisition), the final precision is 0.5 mm. However, in a real scenario, the precision could be limited by other issues, like thermal drifts or laser instabilities.
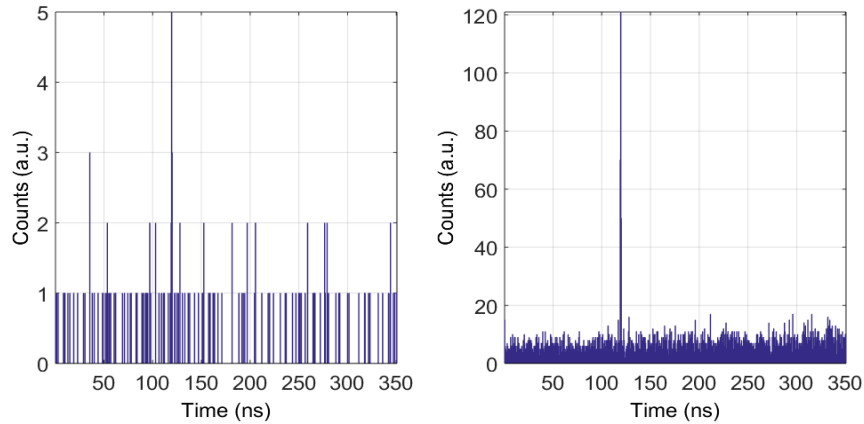

Fig. 3. Time-of-Flight histograms of one of the 1024 pixels, with 40:1 ratio between background and signal photons, computed on 512 frames (left) and 16,384 frames (right) per histogram. With the latter, a sub-centimeter precision is reached through centroid computation, while the former provides a poorer precision, very much limited by shot-noise.

3. Augmenting 3D depth performance

For long-distance TOF ranging, with hundreds of meters target distance, a considerably high laser pulse energy (in the order of some tens of μJ) has to be employed with linear-mode photodetectors. Instead, an example of low-power long-distance 3D ranging measurement obtained thanks to a single-photon single-pixel scanned detector was performed in [23] by McCarthy et al., by using a 1 mW average illumination power to obtain 1 ms per-pixel dwell time (i.e. they acquired a scanned 32×32 equivalent pixels image at 1 fps). For a non-scanned, high frame-rate, multi-pixel imaging system like the one presented in this paper, the drawback is the need of higher optical power, due to the increased area (i.e. Field-of-View, FoV) to be illuminated.

Moreover, when using a single-photon imager in *photon-timing* mode, the readout time is typically much longer than the actual TDC active time, thus limiting the actual laser repetition rate to the inverse of the dwell time duration. In our case, each pixel is readout in just 9.7 ns, hence the full 1024 pixel frame takes about 10 μs to be serially read out, corresponding to a maximum frame-rate of 100,000 fps. Therefore, the actual maximum photon detection rate is 100,000 photons/s per pixel at the most, e.g. achievable with a 100 kHz pulsed laser. It could be possible to almost saturate this rate, by increasing the pulse energy in order to have an almost 100% detection probability per each pulse. However, this will result in a distortion of the acquired histogram, which will no longer look like to the optical waveform to be measured, due to photons pile-up [24], introducing the so-called time walk error in 3D ranging measurements, because of which the centroid position depends both on target distance and on object brightness. To prevent this error, and avoid that a first detected photon (the one triggering the TDC) could hide others belonging to the same pulse, the detection ratio should be limited to less than about 5%. Time walk error can be also corrected in post-processing if full knowledge of the shape of the emitted optical pulse is provided [25].

One possibility to avoid time walk error without decreasing signal intensity is to shine several pulses and repeat TOF measurements many times within the same frame. In this way, the probability to detect a photon in each measurement can be kept below the 5% limit, whereas the overall collection probability in each frame can be increased with no distortions, eventually achieving the theoretical limit of one photon per each frame. In this situation, in fact, the 5% limit has to be considered within each laser pulse, and not on the overall frame.

In order to increase the chance to detect photons in *photon-timing* mode, with no pile-up distortion and with no higher pulse energy, we conceived the following two techniques.

3.1 Enhanced sensitivity through TDC unwarping

One advantage in using high repetition rate and low pulse energy lasers is the broad availability of commercial products. Unfortunately, this hits with the typical constraints of the minimum frame duration and the FSR of the TDC. Fig. 4 (a) shows a typical case, where the TDC range is shorter (e.g. 360 ns of our in-pixel TDCs) than the time required to readout the full frame (e.g. 10 μs in our case). In this case, since just one TDC conversion can take place in each frame, the effective laser repetition rate is limited by the frame-rate (100,000 fps in our case).

The first technique we propose exploits a higher repetition-rate laser with sub-nanosecond laser pulses that shine the scene more than just once per frame (see Fig. 4 (b)). For example, a 100 MHz (i.e. repetition period is 10 ns) laser could provide 36 excitations within the 360 ns FSR of the TDCs, thus resulting in a 36× increment of detection probability. If pile-up (which should be evaluated on the single laser pulse and not on the overall frame) is negligible [24], the TOF histogram will show equally spaced peaks, which can be eventually unwarped and summed together, thus obtaining larger detection probability and, therefore, better depth resolution. The drawback is a reduced maximum depth-range, set by $1/f_{laser}$ instead of the complete FSR of TDCs, if no smart processing like the ones presented in [26] and [27] is
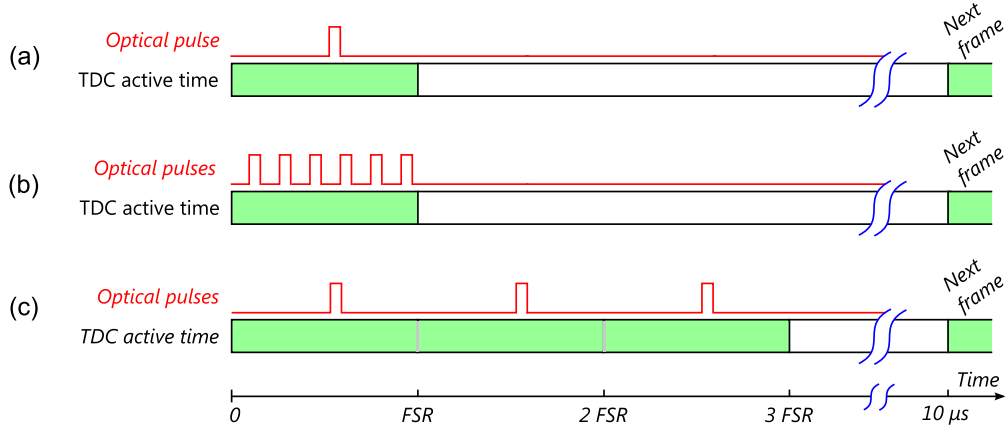
Fig. 4. For frame durations longer than the TDC FSR (a), the 10 μs readout time limits the laser repetition rate. Such a limitation can be overcome (b) by employing a laser repetition period much shorter than the FSR and then unwarping the histograms or (c) by using a laser repetition period equal to the FSR, thus automatically folding repetitions back into the same histogram. Note that frame duration is always fixed to 10 μs, limited by the readout time, which is performed concurrently with dwell time.
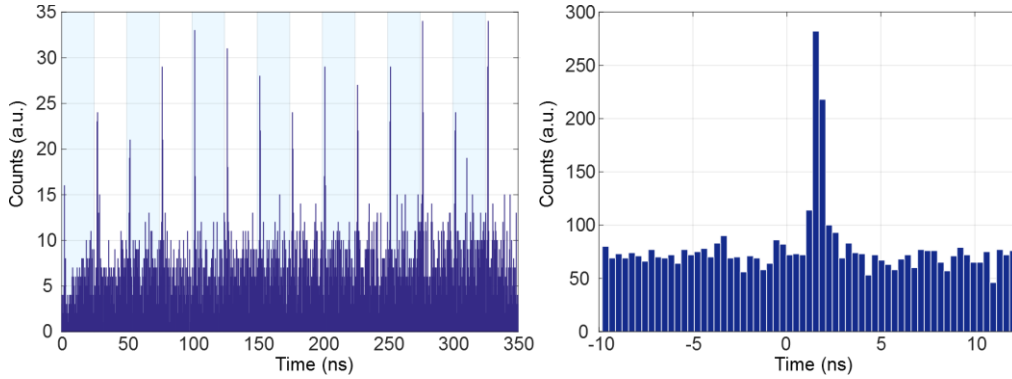

Fig. 5. High repetition rate pulsed lasers can be profitably employed to reduce the required pulse energy: (left) histogram of one pixel, obtained with a 40 MHz laser (25 ns repetition period) and 360 ns TDC FSR and (right) corresponding unwarping histogram. Note the resulting higher peak intensity, but the reduced depth range of 3.75 m (corresponding to 25 ns).

performed. For instance, in Fig. 5 we used a 40 MHz laser to enhance sensitivity (or to correspondingly decrease the required pulse energy) by a factor 15×, with a final depth range of 3.75 m (given the 25 ns laser repetition period).

3.2 Enhanced sensitivity through TDC refolding

The second technique we implemented in the camera to enhance sensitivity, but with no reduction of the depth range, exploits TDC refolding (see Fig. 4 (c)). If no photon is detected within the first FSR of the TDC, the TDC is still kept active (i.e., the STOP signal is provided only at the end of the frame) until a photon from a subsequent laser pulse (or a background or DCR event) is detected and triggers the conversion. In this way, thanks to the TDC refolding, once the FSR is reached, photons from different pulses are marked with the same TOF value, with no need of unwarping through post-processing. Of course the pulsed laser repetition period must be chosen to be an integer sub-multiple of the TDC's FSR.
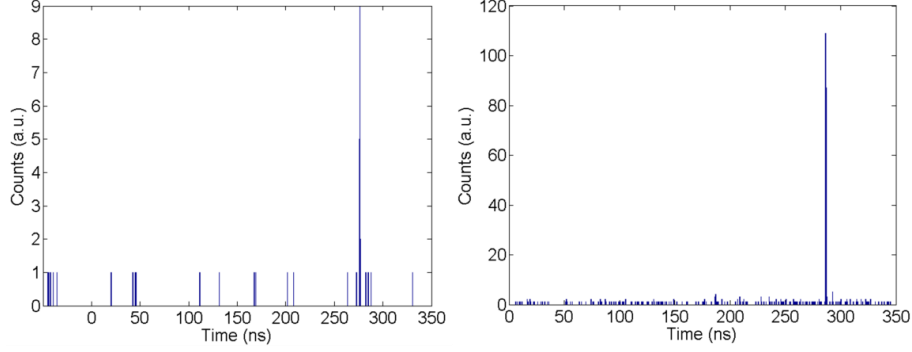
Fig. 6. Comparison between a TOF histogram without (left) and with (right) TDC refolding, when using a 360 ns TDC FSR and a 2.86 MHz (i.e. 1/360 ns) pulsed laser. Note the signal increase by a factor 10x obtained by making the TDCs refold 10 times.

Even in this case, the overall 3D system sensitivity is enhanced compared to the one achieved with a single laser pulse. Furthermore, as the chip readout is concurrent with the acquisition of the successive frame, the measurement duty-cycle (given as TDC active period over frame duration) can be increased from the original 3.6% (360 ns FSR over 10 μs frame duration) up to a theoretical 100%. Eventually, the maximum duty-cycle could be limited by power dissipation of the imager chip, mostly set by the many TDCs running together within an array. For our chip, we verified that thermal dissipation limits the maximum achievable duty-cycle to about 35%. Hence, with the present camera, the photon detection probability improves by a factor 10×, corresponding to a $\sqrt{10} = 3.16×$ signal-to-noise ratio improvement, as shown in Fig. 6.

Fig. 7 presents the single-shot precision (evaluated as the Full-Width at Half Maximum, FWHM of the TCSPC histogram, which reflects the total amount of jitter) for the three modalities presented in Fig. 4: as long as the laser is sufficiently stable, all three modes have a comparable single-shot precision, limited by the intrinsic TDC time-jitter of about 600 ps FWHM. A slight increase of the FWHM can be noted when employing refolding mode, caused by long-term instabilities of the TDC clock oscillator, which becomes more evident when dwell time becomes longer. As anticipated in Section 2, the total ranging precision is influenced both by the single-shot precision and the square root of the number of acquired valid conversions, which can be increased by employing mode (b) or (c). We performed a test acquisition of a white wall at 30 fps, with a final root-mean square precision of 80 ps for mode (a) and 30 ps for mode (c).
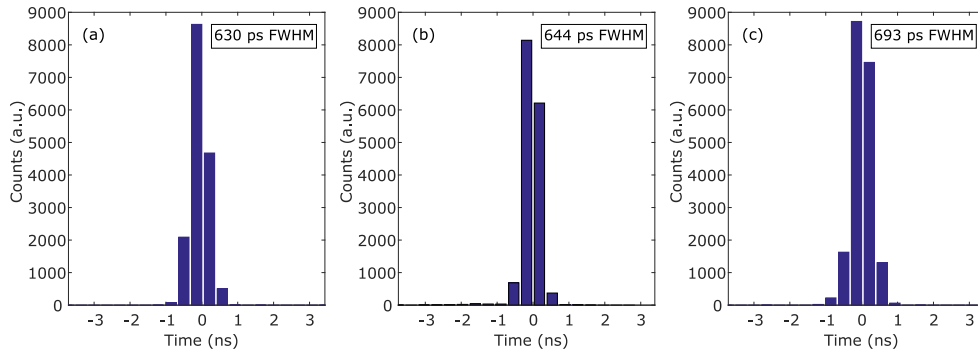

Fig. 7 Single-shot precision for the 3 modalities: traditional mode (a), TDC unwarping (b) and TDC refolding (c). The single-shot precision is comparable for modes (a) and (b), and about 10% worse in mode (c), due to the oscillator long-term jitter.
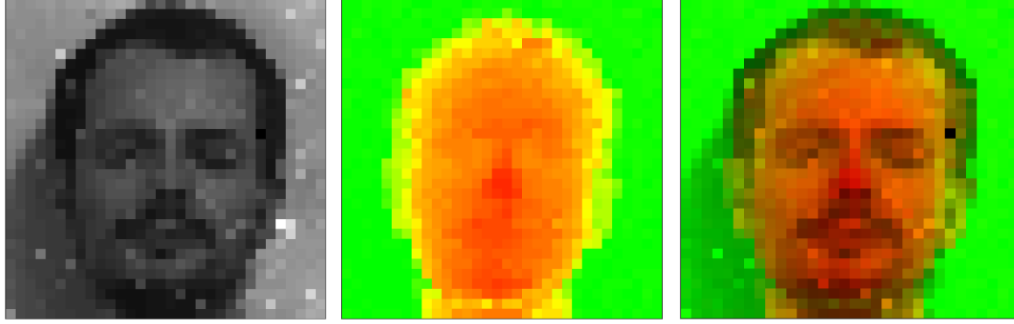
Fig. 8. 2D (left) and 3D (center) data merge (right): the 3D distance information is rendered on a green-to-red scale and each pixel brightness is adjusted by means of the 2D image content, to obtain the final merged image with more details (e.g. hair color and shape, moustaches, etc.).

4. 2D and 3D acquisitions

4.1 Enhanced quality through 2D plus 3D merging

As the sensor is able to detect both 2D and 3D images of the scene (see Fig. 2 right), we combined the two features into an augmented data image. We superimposed the 2D "intensity" image to its 3D "depth" map, thus obtaining an improved 3D model of the object under observation. The concept is better illustrated in Fig. 8, in which the intensity of the false-color scale of the 3D information (obtained through pixel by pixel TOF centroid computation in photon timing mode) is adjusted by means of the 2D image (obtained through the photon-counting mode), in order to have better dynamics with respect to the one achievable by simply estimating the TCSPC peak intensity. More specifically, the 3D distance map is rendered in a green-to-red scale (as visible in Fig. 8 center), whereas the 2D image is inserted (Fig. 8 right) by converting the RGB (Red-Greed-Blue) value computed for each pixel into HSL (Hue-Saturation-Lightness) color space and adjusting each pixel's lightness value. For pixels with a not sufficient signal-to-noise ratio (e.g. in which the number of noise events is higher than 100 times the signal photons), the 3D information is not added to the image and the pixel is rendered in gray scale, highlighting that the SNR was not enough to extract 3D information with proper confidence.
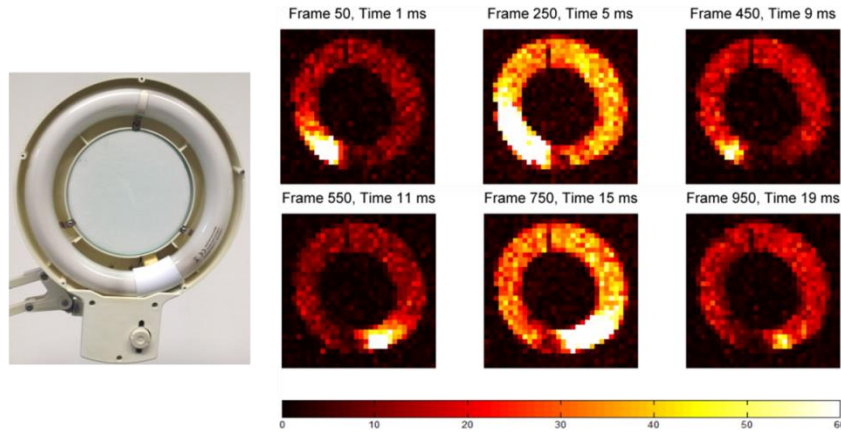
4.2 Very high frame-rate 2D acquisitions

Fig. 9. Frames from a 2D movie acquired at 50,000 fps showing the fast gas discharge propagation through a circular neon lamp tube (left). Note the discharge ignition (frame 50 and 550), propagation (250 and 750) and quench (450 and 950) moving periodically clockwise and then counter-clockwise. The false color scale is the number of photons in each 20 µs frame. The oscillation repeats at the 50 Hz mains supply frequency.

In photon-counting mode, the camera is able to acquire 2D images of the target scene with a frame-time as short as 10 μs, therefore being able to reach very high frame-rates, up to 100,000 fps. Fig. 9 shows six frames from a 50,000 fps 2D video of a neon lamp. The frames illustrate the temporal behavior of the gas discharge inside the tube: the spark originates at the positive electrode (bottom-left corner of the lamp, frame 50), then spreads through the tube (frame 250) and eventually quenches-off (frame 450); periodically after one half-period (10 ms) the whole process repeats the other way round (frame 550-950). Given the 50 Hz neon mains frequency, the whole process repeats every 20 ms, i.e. every 1,000 frames.

4.3 Enhanced x-y resolution through intra-pixel scanning

For 3D ranging video acquisitions, we employed a simple non-confocal optical setup, in which the active illumination was supplied by a 750 nm wavelength, 100 ps full-width at half maximum (FWHM), 90 mW (class 4) pulsed four wave mixing laser, with a 40 MHz repetition rate allowing to exploit TDC unwarping as described in Fig. 4 (b). As already introduced in Section 3, the relatively high 90 mW optical power is required to encompass the field-of view of all the pixels simultaneously, whereas scanned single point detectors like the one presented in [23] can employ a much lower optical power, at the expense of frame-rate. The adjustable laser divergence, ranging from 15 to 180 mrad, allowed the illumination field-of-view to be varied from $15 \times 15 \, cm^2$ up to $300 \times 300 \, cm^2$, with 8 m maximum target distance. The detection field-of-view was matched to the illumination by accordingly modifying the lens focal length, ranging from 25 mm up to 250 mm (f/4.8). Both camera and illuminator were mounted on three axes motorized micro-translators, for focus (z axis) adjustment and for scene scanning (along x and y axes). The SPAD photon detection efficiency at 750 nm wavelength is 10%.
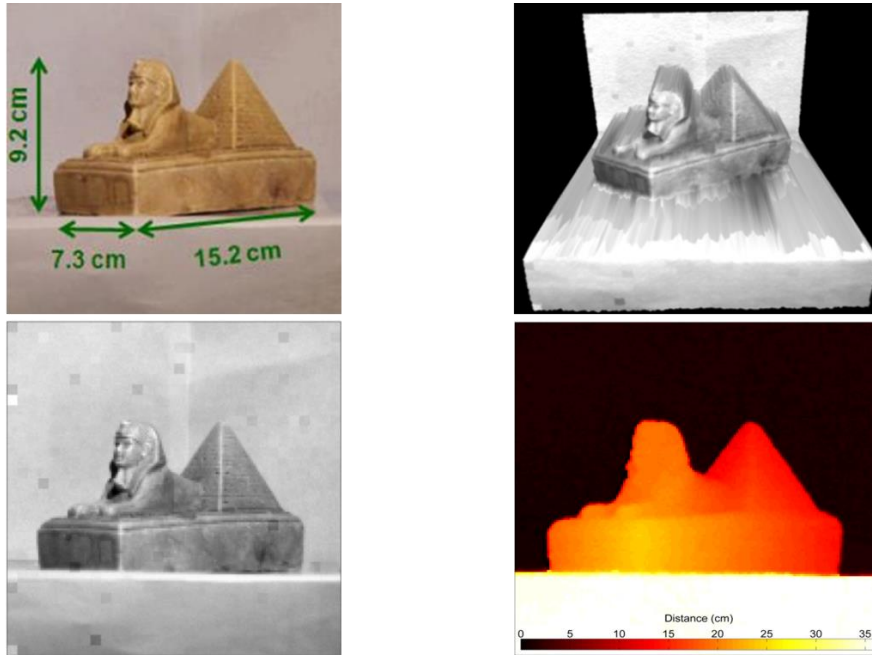

Fig. 10. Object under observation as acquired with a conventional CCD camera (top left) and as acquired by the 32×32 SPAD camera with 5×5 intra-pixel scans, in 2D *photon-counting* mode (bottom left) and in 3D *photon-timing* mode (bottom right). The 2D and 3D merge (top right) gives a 25,600 pixel resolution image with more detailed depths and shades. The false-color scale represents the depth compared to background (bottom right). The statue was at 8 m from the camera and the overall dwell time was 150 ms.
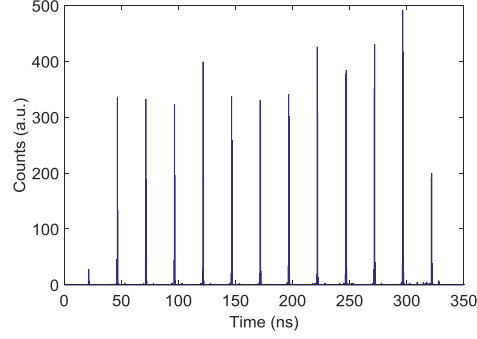
Fig. 11 Arrival time histogram for a typical pixel for the acquisition shown in Fig. 10. The histogram is performed on 16,384 frames, in which each pixel collected about 8,000 signal and 50 background photons.

In order to improve x-y spatial resolution, we performed intra-pixel scanning by moving the camera with 30 μm steps, i.e. a submultiple of the 150 μm pixel pitch. Fig. 10 shows a small statue acquired with our 32×32 SPAD-array in 2D and 3D modes, with a 5×5 intra-pixel scanning, thus obtaining a 25 kpixel (160×160) image in an overall 150 ms dwell time (6 ms per each scan, without considering readout time). The histogram for a typical pixel is shown in Fig. 11: the acquisition has been taken in dim light conditions (< 50 lux), obtaining about 50 background counts and 8,000 signal photons per pixel.

4.4 Enhanced field-of-view through inter-pixel scanning

We also performed inter-pixel scanning in order to enlarge the field-of-view, by shifting the camera with 4.8 mm steps, i.e. the size of the 32×32 array active area (i.e. 32×150 μm). In Fig. 12, a statue with $20 \times 10 \times 10$ cm^3 size was acquired with both intra- and inter-pixel scans: the final lateral resolution of 51,200 pixels (160×320) was achieved by means of a 5×5 intra-pixel scan and a 1×2 inter-pixel, in just 300 ms total dwell time (as in the previous acquisition, the dwell time for each scan was 6 ms, without considering dead times between each frame).

Note that the presence of hot pixels causes some artefacts in the 2D acquisitions (see Fig. 12 left), whereas 3D information is not influenced (see Fig. 12 center and right). Hot pixels (i.e., pixels with a noise level much higher than the average) are very common in any SPAD array, though in our chip they were just 5% of the total pixel amount [22].
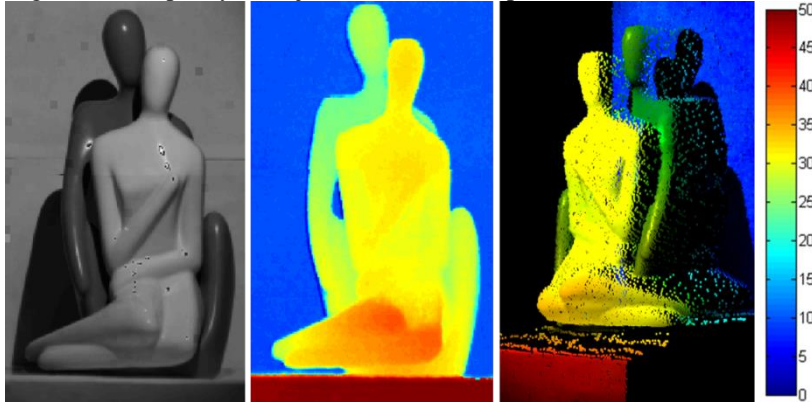

Fig. 12. Acquisition by the SPAD camera with 5×10 inter-pixel scans, in 2D photon-counting mode (left) and in 3D photon-timing mode (center). The 2D and 3D merge (right) gives a 51,200 pixel resolution image (160×320) with more detailed depths and shades. The false color scale represents the depth (in cm) compared to background. The target was distant 8 m from the SPAD camera. The dwell time was 300 ms.
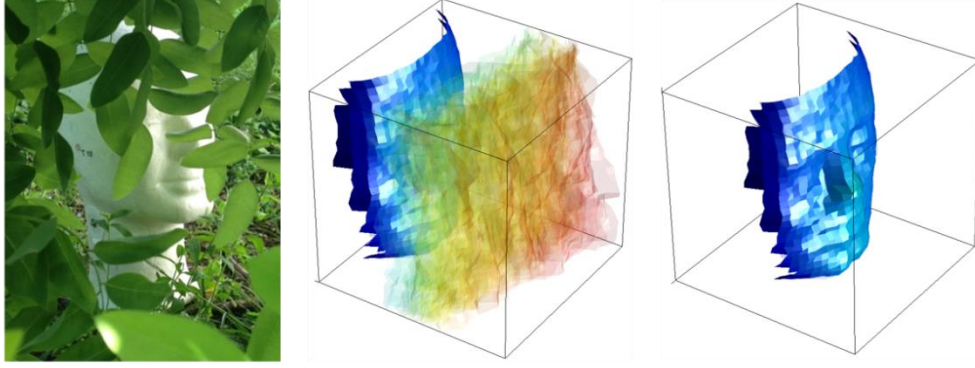
Fig. 13. Behind-foliage detection: photograph of the scene (left), image acquired by the SPAD TOF camera after reconstruction of the 3D scene (center), and after taking apart foliage contribution by thresholding photons arrival time (right). Tree and face were placed at 7 m distance from the SPAD camera.

4.5 Behind-foliage 3D vision

Differently from phase-resolved ("indirect" TOF) 3D ranging systems, direct TOF ranging systems can be easily employed even when translucent or dark moving obstacles are present between camera and target scene, thus partially obstructing the view. Such a capability can be exploited for instance in LiDAR-based measurements [28], which can profitably make use of the individual single-shot photons' TOF distance, or in military applications to detect targets camouflaged in vegetation.

Fig. 13 shows an example of scene in which a tree branch, waving under light wind, was inserted between camera and target scene. In this condition, photons reflected from the obstacle are easily distinguished from the ones hitting the target, thanks to their different arrival times. The returns from the two objects can therefore be elaborated independently and the complete scene can be reconstructed. Even if the distance between hidden face and tree was just 30 cm and the scene was 7 meters away from the SPAD camera, both objects are clearly visible, after few seconds of acquisition.

5. Conclusions

We presented a system based on a 32×32 pixel Single-Photon Avalanche Diode camera, able to acquire very high frame-rate (up to 100 kfps) 2D movies by counting single photons, and 3D ranging acquisitions by exploiting direct Time-Of-Flight (TOF) technique, thanks to 350 ps resolution Time-to-Digital Converters integrated into each pixel. We conceived two techniques that, without increasing the optical pulse energy, improve Signal-to-Noise ratio of 3D ranging acquisitions and mitigate a typical disadvantage of most TOF cameras. In fact, TOF cameras are usually able to record just one TDC conversion per frame, thus limiting the laser repetition rate to few hundreds of kHz. Thanks to the proposed techniques, it is possible to shine many laser pulses in each frame, reducing the maximum depth range or exploiting the refolding of the TDCs.

We also exploited the camera to acquire both high frame-rate (50 kfps) 2D videos of the evolution of a neon gas discharge inside a fluorescent lamp, and 2D + 3D static scanned images (with up to 51,200 pixels) of objects located at 8 meters distance, with centimeter depth resolution by employing a 90 mW pulsed laser. Thanks to the possibility to acquire both 2D and 3D acquisitions, we have been able to superimpose the 2D intensity images on the 3D depth maps in order to obtain complete description of the target scene.

Finally, we showed how the TOF technique can detect objects partially hidden by translucent shields, being able to distinguish in the time-domain (i.e. depth range) photons incoming from both the shield and from the target object.