Low Power Time-of-Flight 3D Imager System in Standard CMOS

Priyanka Kumar, Edoardo Charbon, R. Bogdan Staszewski, A. Borowski

Delft University of Technology, The Netherlands
Email: {P.Kumar-1, E.Charbon, R.B.Staszewski}@tudelft.nl, andreborowski@hotmail.com

This paper proposes a novel time-of-flight (ToF) sensor with high dynamic range and fine depth resolution for medium and long range 3D imaging applications. It focuses on low power consumption while maintaining human-eye safety requirements. A pixel array architecture is employed where each pixel consists of a single-photon avalanche diode (SPAD) as a photo-detector along with a dedicated time-to-digital converter (TDC) for fast image acquisition. The imager is designed to achieve millimeter-level depth resolution with maximum range up to 30 m operating at a maximum frame rate of 1000 fps. It targets security applications, primarily facial recognition, but is also suitable for automotive vision and robotics.

I. INTRODUCTION

Three-dimensional (3D) imaging is of interest for its ability to capture objects as we see them in the real world. In the past, 3D imaging techniques have been mostly restricted to research domains and low volume applications. However, the advancement in technology along with the recent outbreak of low-cost solid-state 2D imagers has fostered the idea of building low-cost and compact 3D imaging systems [1]. Such systems have various applications ranging from face-recognition to machine vision, from 3D gaming to nuclear security and from molecular imaging to land and sea surveying. This paper proposes a high resolution, monolithic time-of-flight (ToF) 3D imager sensitive to a single photon and featuring very low power consumption.

II. IMAGER ARCHITECTURE

The proposed 3D imaging system consists of a modulated laser light source and a matrix of pixels capable of detecting single photons along with their ToF at an accuracy of several picoseconds. The depth of the object is derived by computing the ToF of photons that are emitted from the laser source and reflected back from the target object. In this work, a single-photon avalanche diode (SPAD) is used to detect the incident photons. The main advantage of a SPAD is that it produces a digital signal for a photon impingement with outstanding timing resolution. The digital signal can be processed at ultra-high speed with picosecond accuracy in a CMOS circuit, such as a time-to-digital converter (TDC), to measure the time-of-arrival of the photon reflected by the target. The time-interval between photon emission and its detection is measured to extract the depth information of the object. The depth or range to a point on the object is given by: $\text{Depth} = c\,\frac{\text{ToF}}{2}$, where *c* is the speed of light. A depth map created by such measurements forms the 3D image of an object.

Typical facial recognition systems for security or surveillance applications require about 1 mm of depth resolution and a range of about 25-30 m. This translates into a large dynamic range for the time interval measurement system with millimeter resolution. Also, with a large pixel array of 1000x1000 pixels for high lateral resolution, the power consumption of each pixel becomes critical. This is because the whole system must not dissipate more than 1 Watt of power to avoid excessive heating and an eventual burnout. The desired imager specifications are achieved with an array of SPAD-TDC pixels operating at a frame rate of 1000 fps.

A. Imager Setup

The setup is operated in time-correlated single-photon counting (TCSPC) regime [2]. The ToF estimation is obtained by averaging the histograms of the reflected photons. The pulsed laser source emits a cone beam to illuminate the scene of interest, as shown in Figure 1. The cone can be perceived as a collection of discs incident on the object of interest. Assuming partial collimation, for a given laser power emitted from the source, the intensity of light varies with the radius, *r,* of the disk is:

$$\text{Intensity} = \frac{\text{Power emitted from source}}{\text{Area}} = \frac{\text{P}}{\pi \text{r}^2}$$

The impinging photon energy on any point on the object is: $E_0 = \frac{hc}{\lambda}$, where $h = 6.62 \cdot 10^{-34}$ J s is Planck's constant and $\lambda = 0.78$-0.9 μm is the wavelength. A CMOS SPAD photo-detector is sensitive to this near-infrared wavelength and meets eye-safety requirements. To achieve sensitivity to higher wavelengths such as 1.5 μm for increased eye safety, a Germanium layer can be employed in the design of the SPAD photo-detector [3].

Most of the reflected photons from the object are lost and only a few are detected. This loss is referred to as 'link attenuation' and its magnitude is calculated as the ratio of the number of photons incident on the imager to those reflected by the object. It may be assumed that the light is reflected

978-1-4673-1260-8/12/$31.00 ©2012 IEEE

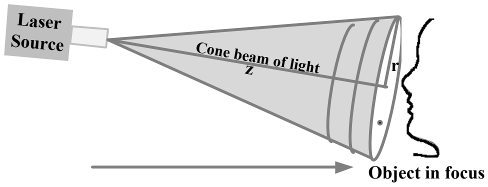

Figure 1 - Cone beam hitting the target.

from the object equally in all directions in the form of a hemisphere with radius equal to distance of the object from the imager 'z'. Considering a nominal pixel pitch of 25 µm in 65 nm CMOS technology for compact system, the link attenuation for a 1000x1000 pixel array is given by:

$$\frac{\text{Energy}_{\text{imager}}}{\text{Energy}_{\text{reflection}}} = \frac{\text{Area}_{\text{imager}}}{\text{Area}_{\text{hemisphere}}} = \frac{(25\mu .25\mu)\text{x}10^6}{4\pi .30^2} = 5.5\ x\ 10^{-8}$$

This corresponds to a large link-attenuation of 60 - 80 dB for a range of 30 m. implying that the laser pulses emitted from the source must have very high intensity. A modified TCSPC technique is proposed to solve this problem.

B. Proposed technique for TCSPC

In conventional TCSPC, the photons reflected by a target are detected and time-stamped. The experiment is repeated a high number of times to acquire relevant statistics. In order to avoid pile-up effects, it is desirable to detect far less than one photon per SPAD per pulse. On the other hand, for maximum efficiency, it must be ensured that as many photons are detected per SPAD per pulse as possible. To meet these contradictory requirements, it is important to control the intensity of light reaching the detector, which in turn determines the desired emitted laser energy per pulse. Figure 2 depicts the relation between the number of effective photons per SPAD per pulse and the optical energy of pulse incident on a square matrix of detector. For one photon detection per SPAD per pulse, the optical energy received on the imager should be ~1 pJ. Hence, the emitted laser energy per pulse needs to be 100 µJ, with 80 dB of link attenuation.

A laser pulse of 100 µJ is extremely powerful and may be hazardous to the human eye. A microlens may be incorporated in the detector to increase the imager efficiency by about 5-20 times [4]. Figure 2 shows the relation between energy incident on the imager and number of photons per SPAD per pulse with and without the use of a 20x microlens. However, the mechanical integration of a microlens increases the cost of the system.

In order to achieve one photon detection per SPAD per pulse or less without the use of eye-hazardous high-energy pulses, we propose the use of a high frequency *train of pulses* where the intensity of each pulse is much lower than the desired intensity. Figure 3 presents the conceptual idea. A closely spaced train of pulses compensates for the link attenuation and low photon detection probability. For example, emitting 1000 pulses, each with energy of 100 nJ, implies that the total laser energy emitted in each burst is 100 µJ. Furthermore, constructing closely-spaced (high-frequency) and low intensity pulses is easier than a single powerful pulse.

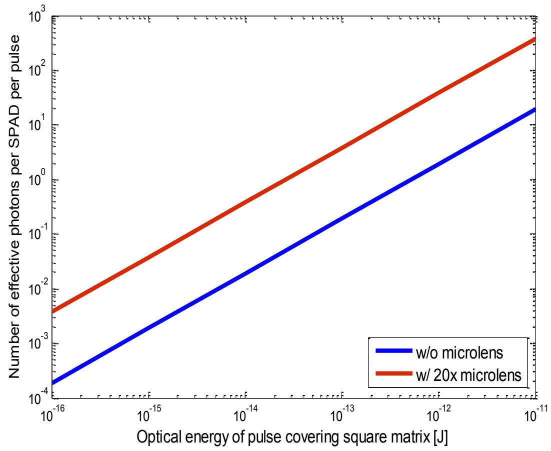

Figure 2 – Effective photons per SPAD per pulse vs. energy incident on a square matrix of detector.

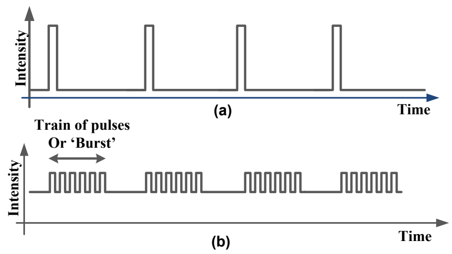

Figure 3- Train of pulses

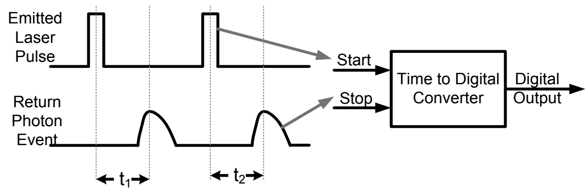

Figure 4 - Basic concept of TDC in 3D imaging.

C. Imager architecture

A SPAD photo-detector is an n-p junction biased above breakdown, thus operating in Geiger mode [5]. In this mode of operation, the device becomes sensitive to single photons. Each detected photon generates a digital signal that is then sent to the TDC. In the prior art, sharing of resources was common in detectors. In [1], a single TDC was shared by a full array while a TDC per four columns was employed in [2] to enable parallel image acquisition. In this work, an in-pixel TDC as in [6] calculates the ToF of a photon incident on a SPAD photo-detector. The proposed imager is an array of 1000x1000 such pixels. The output of TDCs in the pixel array presents the ToF estimation for constructing 3D image. The technique used for ToF estimation is discussed in Section III.

III. Time of Flight Estimation

The time interval measurement for the proposed 3D imaging system is performed by the TDC as shown in

Figure 4. In order to measure a maximum range of 30 m with a resolution better than 5 mm, a TDC with a dynamic range of 14 bits is required. Achieving this dynamic range with a simple delay-line based TDC requires a large number of delay elements and, consequently, leads to large are and power consumption for a 1000x1000 pixel array. In order to limit the number of delay elements, coarse-fine TDC architecture is one of the techniques generally employed.

1) Coarse-Fine Architecture based Time Interval Measurement

The conventional method for measuring a time interval is based on a coarse-fine technique [7] as shown in Figure 5. It uses a coarse TDC to count the whole number of reference clock cycles, while a fine TDC performs interpolation (from the photon event edge to the next clock edge) to achieve high time resolution.

The choice of clock frequency used for the coarse TDC presents a trade-off between area, speed, power, and jitter. A higher clock frequency requires a faster coarse counter that consumes more power but needs fewer interpolation-levels in the fine TDC. A low clock frequency reduces the power consumed in the coarse counter at the cost of higher delay elements in the fine TDC. These trade-offs are summarized for two values of clock-frequency in Table 1.

In this design, we propose the use of a fine TDC in each pixel to enable independent measurements of ToF for each SPAD event to avoid loss of information and also enable faster image acquisition. But to achieve a reasonable pixel fill factor, the area occupied by the fine-TDC cannot be very large. To achieve this goal for a 1000x1000 pixel array, the frequency of the global clock cannot be very high. Furthermore, randomness in photon detections in the pixels implies that the coarse counters in each pixel always need to be in operation. Therefore, a coarse-fine TDC architecture is not suitable for large pixel arrays as it is power hungry by design. Although this can be reduced to some extent by sharing of coarse counters between a group of pixels, it is still an inefficient time measurement scheme in terms of area and power.

2) Proposed Scheme for Time Interval Measurement

To reduce power consumption, an event-driven paradigm is used based on a START-STOP configuration, where the photon arrival triggers the START signal while the following edge of a configurable clock works as the STOP signal as shown in Figure 6. The TDC is active only when a SPAD event occurs, thereby reducing the overhead of an always-running counter as in the coarse-fine TDC approach.

This approach also incorporates a military concept called reconnaissance to lower the power consumption. As shown in Figure 7, the range of object is estimated in the first few measurements and the clock frequency is adjusted according to the distance of the object from the imager. For instance, if the object is at a distance of 5 m, the clock period is set to an equivalent time of ~33 ns. When the clock frequency is set

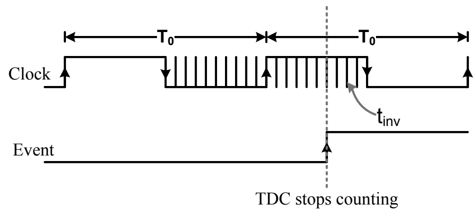

Figure 5 - Principle of a coarse-fine TDC [7].

Table 1 – Performance of coarse-fine TDC over two frequencies.

Parameters	f = 10 GHz	f = 2.5 GHz
Area	1x	4x (longer interpolation period, more delay taps in TDC)
TDC Resolution	Equal to an inverter delay	
Jitter	$\sqrt{5}\,\sigma_{inv}$	$\sqrt{20}\,\sigma_{inv}$
Speed of operation	Higher	Lower
Clock Distribution	Complex, more power/effort in distribution	Simpler, less power/effort in distribution

correctly, an incident photon event triggers the TDC which then counts up to the following clock edge of the system clock as shown in Figure 6. In this case, the TDC is active only when photon detection occurs, contrary to the coarse-fine TDC architecture.

The ToF measurement in this scheme is given by:

$$\text{Total time} = [\mathrm{T_{clock}} - (\mathrm{N_{output}} * \mathrm{t_{inv}})]$$

The maximum clock period, in this technique, can be equivalent to the maximum range of imager, say 30 m or 200 ns, while the minimum clock period corresponds to only 1 m or 6.6 ns. This implies that the maximum system clock frequency is ~150 MHz which results in much lower power for a clock distribution circuit than a 10 GHz clock frequency. Moreover, the TDC is in operation for a much smaller period of time thereby saving power. For minimal power consumption, the time interval for which the TDC in each pixel is in operation can be fixed to an average of 5 ns corresponding to a distance of about 1 m. The time interval of 5 ns is good enough to capture an image in a depth range varying about ~1 m. This is a reasonable assumption for a 3D camera to operate in low power mode.

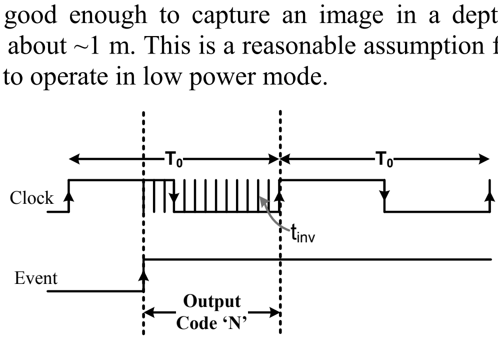

Figure 6 - ToF measurement technique for lower power consumption.

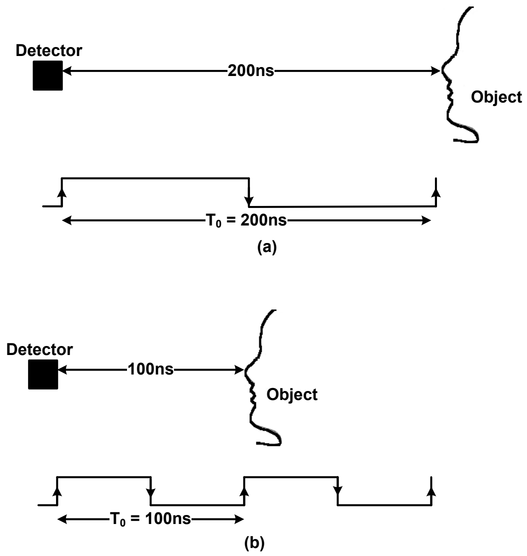

Figure 7 – The concept of Reconnaissance: (a) Lowest clock frequency when object is 30m (~200 ns) apart, (b) clock frequency increased when an object at closer distance is detected leading to lower power consumption.

A potential issue in this approach is aliasing from objects outside the fixed range of the imager. But the intensity of photons incident on the imager from further distance is much lower and these readings can be ignored in post-processing.

B. Improving Resolution

The uncertainty of the detector's depth resolution is determined by two error sources – TDC quantization noise and non-linearity, and SPAD jitter. We assume that the dark counts and the background counts can be ignored, in most cases, since they are uniformly distributed in time and the signal-to-background ratio in the time region of interest is sufficiently high to validate the assumption. The TDC quantization noise is half of a delay tap in a delay line which is about 10-12 ps for an inverter based delay line in 65 nm CMOS. The SPAD jitter corresponds to the timing uncertainty associated with the detection of a photon. This is roughly 60-100 ps and, therefore, is the dominant noise contributor in this system.

To reduce the timing uncertainty due to these two sources, averaging is employed where N measurements averaged together reduce the uncertainty by $\sqrt{N}$ times. The TCSPC measurement yields a histogram for each pixel. To avoid holding the histogram in memory, averaging is performed during the collection of the measurement samples. Two approaches to averaging are used in this system for improved resolution – spatial and temporal.

Spatial averaging implies averaging among several SPADs, i.e., multiple SPADs are grouped in a single pixel and correspond to a single point on the target. The time-stamps from this group of SPADs are averaged to improve the depth resolution at a cost of lateral resolution (Figure 8). Temporal averaging refers to averaging of measurements from a single SPAD occurring at different instants of time. As discussed in

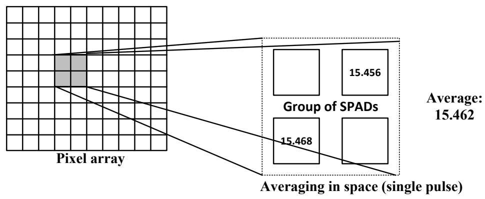

Figure 8 - Spatial averaging in the pixel array.

earlier sections, this work employs the concept of '*train of pulses'*. The measurement of ToF in all the pixels during a burst constitutes a frame.

Assuming that the object of interest does not move over the duration of multiple bursts, which in the order of µs, the ToF measurements of consecutive frames can be averaged to reduce the error and improve resolution. However, multiple frames being merged together by averaging lead to lower frame rate. Therefore, temporal averaging results in higher depth resolution but at the cost of lower frame rate.

IV. CONCLUSION

We have proposed a new time-of-flight sensor capable of computing the depth map of a scene in real time. The sensor was designed to minimize power consumption and to assure eye safety using a series of techniques implemented at a system level. In this architecture, a pixel array is implemented in 65 nm CMOS, where each pixel comprises a single-photon avalanche diode (SPAD) and a dedicated time-to-digital converter (TDC) for parallel time-of-flight (ToF) acquisition. Millimeter-level depth resolution at ranges up to 30 m is achieved with a maximum frame rate of 1000 fps.

V. ACKNOWLEDGEMENT

This work was supported, in part, by Dutch Technology Foundation STW Valorization Grant.

VI. REFERENCES

[1] C. Niclass, et al., "Design and Characterization of a CMOS 3-D Image Sensor based on Single Photon Avalanche Diodes", IEEE Journal of Solid-State Circuits, 40(9), 1847-1854, 2005.

[2] C. Niclass, et. al., "A 128 x 128 single-photon image sensor with column-level 10-bit time-to-digital converter array" , IEEE 1. of Solid-State Circuits, vol.43 pp. 2977-2989, 2008.

[3] A. Sammak, et al., "A CMOS Compatible Ge-on-Si APD Operating in Proportional and Geiger Modes at Infrared Wavelengths," in IEEE International Electron Devices Meeting, 2011.

[4] S. Donati, G. Martini, M. Norgia, "Microconcentrators to recover fill-factor in image photodetectors with pixel on-board processing circuits", Opt. Express 15(26), 18066-18075, 2007.

[5] S. Cova et al., "Towards Picosecond Resolution with Single-Photon Avalanche Diodes", Rev. Sci. Instr., 52(3), 408-412, 1981.

[6] C. Veerappan, et, al., "A 160x128 Single-Photon Image Sensor with On-Pixel, 55ps 10b Time-to-Digital Converter", IEEE Intl. Solid-State Circuits Conference, 312-314, 2011.

[7] R. B. Staszewski, et al.; , "1.3 V 20 ps time-to-digital converter for frequency synthesis in 90-nm CMOS," Circuits and Systems II: Express Briefs, IEEE Transactions on , vol.53, no.3, pp. 220- 224, 2006.