Machine Vision and Applications manuscript No.
(will be inserted by the editor)

An Overview of Depth Cameras and Range Scanners Based on Time-of-Flight Technologies

Radu Horaud · Miles Hansard · Georgios Evangelidis · Clément Ménier

Received: date / Accepted: date

Abstract Time-of-flight (TOF) cameras are sensors that can measure the depths of scene-points, by illuminating the scene with a controlled laser or LED source, and then analyzing the reflected light. In this paper we will first describe the underlying measurement principles of time-of-flight cameras, including: (i) pulsed-light cameras, which measure *directly* the time taken for a light pulse to travel from the device to the object and back again, and (ii) continuous-wave modulated-light cameras, which measure the phase difference between the emitted and received signals, and hence obtain the travel time *indirectly*. We review the main existing designs, including prototypes as well as commercially available devices. We also review the relevant camera calibration principles, and how they are applied to TOF devices. Finally, we discuss the benefits and challenges of combined TOF and color camera systems.

This work has received funding from the French Agence Nationale de la Recherche (ANR) under the MIXCAM project ANR-13-BS02-0010-01, and from the European Research Council (ERC) under the Advanced Grant VHIA project 340113.

R. Horaud and G. Evangelidis
INRIA Grenoble Rhône-Alpes
Montbonnot Saint-Martin, France
E-mail: radu.horaud@inria.fr, georgios.evangelidis@inria.fr

M. Hansard
School of Electronic Engineering and Computer Science
Queen Mary University of London, United Kingdom
E-mail: miles.hansard@qmul.ac.uk

C. Ménier
4D View Solutions
Grenoble, France
E-mail: clement.menier@4dviews.com

Keywords LIDAR · range scanners · single photon avalanche diode · time-of-flight cameras · 3D computer vision · active-light sensors

1 Introduction

During the last decades, there has been a strong interest in the design and development of range-sensing systems and devices. The ability to remotely measure range is extremely useful and has been extensively used for mapping and surveying, on board of ground vehicles, aircraft, spacecraft and satellites, and for civil as well as military purposes. NASA has identified range sensing as a key technology for enabling autonomous and precise planet landing with both robotic and crewed space missions, e.g., [Amzajerdian et al., 2011]. More recently, range sensors of various kinds have been used in computer graphics and computer vision for 3-D object modeling [Blais, 2004], Other applications include terrain measurement, simultaneous localization and mapping (SLAM), autonomous and semi-autonomous vehicle guidance (including obstacle detection), as well as object grasping and manipulation. Moreover, in computer vision, range sensors are ubiquitous in a number of applications, including object recognition, human motion capture, human-computer interaction, and 3-D reconstruction [Grzegorzek et al., 2013].

There are several physical principles and associated technologies that enable the fabrication of range sensors. One type of range sensor is known as LIDAR, which stands either for *"Light Imaging Detection And Ranging"* or for *"LIght and raDAR"*. LIDAR is a remote-sensing technology that estimates range (or distance, or depth) by illuminating an object with a collimated

laser beam, followed by detecting the reflected light using a photodetector. This remote measurement principle is also known as *time of flight* (TOF). Because LIDARs use a fine laser-beam, they can estimate distance with high resolution. LIDARs can use ultraviolet, visible or infrared light. They can target a wide range of materials, including metallic and non-metallic objects, rocks, vegetation, rain, clouds, and so on – but excluding highly specular materials.

The vast majority of range-sensing applications require an array of depth measurements, not just a single depth value. Therefore, LIDAR technology must be combined with some form of scanning, such as a rotating mirror, in order to obtain a row of horizontally adjacent depth values. Vertical depth values can be obtained by using two single-axis rotating mirrors, by employing several laser beams with their dedicated light detectors, or by using mirrors at fixed orientations. In all cases, both the vertical field of view and the vertical resolution are inherently limited. Alternatively, it is possible to design a scannerless device: the light coming from a single emitter is diverged such that the entire scene of interest is illuminated, and the reflected light is imaged onto a two-dimensional array of photodetectors, namely a TOF *depth camera*. Rather than measuring the intensity of the ambient light, as with standard cameras, TOFcameras measure the reflected light coming from the camera's own light-source emitter.

Therefore, both TOF range scanners and cameras belong to a more general category of LIDARs that combine a single or multiple laser beams, possibly mounted onto a rotating mechanism, with a 2D array of light detectors and time-to-digital converters, to produce 1-D or 2-D arrays of depth values. Broadly speaking, there are two ways of measuring the time of flight [Remondino and Stoppa, 2013], and hence two types of sensors:

- *Pulsed-light* sensors directly measure the round-trip time of a light pulse. The width of the light pulse is of a few nanoseconds. Because the pulse irradiance power is much higher than the background (ambient) irradiance power, this type of TOF camera can perform outdoors, under adverse conditions, and can take long-distance measurements (from a few meters up to several kilometers). Light-pulse detectors are based on *single photon avalanche diodes* (SPAD) for their ability to capture individual photons with high time-of-arrival resolution [Cova et al., 1981], approximatively 10 picoseconds (10^{-11} s).
- *Continuous-wave* (CW) modulation sensors measure the phase differences between an emitted continuous sinusoidal light-wave signal and the backscattered signals received by each photodetector [Lange and Seitz, 2001]. The phase difference between emitted and received signals is estimated via cross-correlation (demodulation). The phase is directly related to distance, given the known modulation frequency. These sensors usually operate indoors, and are capable of short-distance measurements only (from a few centimeters to several meters). One major shortcoming of this type of depth camera is the phase-wrapping ambiguity [Hansard et al., 2013].

This paper overviews pulsed-light (section 2) and continuous wave (section 4) range technologies, their underlying physical principles, design, scanning mechanisms, advantages, and limitations. We review the principal technical characteristics of some of the commercially available TOF scanners and cameras as well as of some laboratory prototypes. Then we discuss the geometric and optical models together with the associated camera calibration techniques that allow to map raw TOF measurements onto Cartesian coordinates and hence to build 3D images or point clouds (section 6). We also address the problem of how to combine TOF and color cameras for depth-color fusion and depth-stereo fusion (section 7).

2 Pulsed-Light Technology

As already mentioned, pulsed-light depth sensors are composed of both a light emitter and light receivers. The sensor sends out pulses of light emitted by a laser or by a laser-diode (LD). Once reflected onto an object, the light pulses are detected by an array of photodiodes that are combined with time-to-digital converters (TDCs) or with time-to-amplitude circuitry. There are two possible setups which will be referred to as *range scanner* and *3D flash LIDAR camera*:

- A *TOF range scanner* is composed of a single laser that fires onto a single-axis rotating mirror. This enables a very wide field of view in one direction (up to 360°) and a very narrow field of view in the other direction. One example of such a range scanner is the Velodyne family of sensors [Schwarz, 2010] that feature a rotating head equipped with several (16, 32 or 64) LDs, each LD having its own dedicated photodetector, and each laser-detector pair being precisely aligned at a predetermined vertical angle, thus giving a wide vertical field of view. Another example of this type of TOF scanning device was recently developed by the Toyota Central R&D Laboratories: the sensor is based on a single laser combined with a

multi-facet polygonal (rotating) mirror. Each polygonal facet has a slightly different tilt angle, as a result each facet of the mirror reflects the laser beam into a different vertical direction, thus enhancing the vertical field-of-view resolution [Niclass et al., 2013].
– A *3D flash LIDAR camera* uses the light beam from a single laser that is *spread* using an optical diffuser, in order to illuminate the entire scene of interest. A 1-D or 2-D array of photo-detectors is then used to obtain a depth image. A number of sensor prototypes were developed using single photon avalanche diodes (SPADs) integrated with conventional CMOS timing circuitry, e.g., a 1-D array of 64 elements [Niclass et al., 2005], 32×32 arrays [Albota et al., 2002, Stoppa et al., 2007], or a 128×128 array [Niclass et al., 2008]. A 3-D *Flash LIDAR camera*, featuring a 128×128 array of SPADs, described in [Amzajerdian et al., 2011], is commercially available.

2.1 Avalanche Photodiodes

One of the basic elements of any TOF camera is an array of photodetectors, each detector has its own timing circuit to measure the range to the corresponding point in the scene. Once a light pulse is emitted by a laser and reflected onto an object, only a fraction of the optical energy is received by the detector – *the energy fall-off is inversely proportional to the square of the distance*. When an optical diffuser is present, this energy is further divided among multiple detectors. If a depth precision of a few centimeters is needed, the timing precision must be less than a nanosecond.[1] Moreover, the bandwidth of the detection circuit must be high, which also means that the noise is high, thus competing with the weak signal.

The vast majority of pulse-light receivers are based on arrays of *single photon avalanche diodes* which are also referred to as *Geiger-mode avalanche photodiodes* (G-APD). A SPAD is a special way of operating an avalanche photodiode (APD), namely it produces a fast electrical pulse of several volts amplitude in response to the detection of a single photon. This electrical pulse then generally trigers a digital CMOS circuit integrated into each pixel. An integrated SPAD-CMOS array is a compact, low-power and all-solid-state sensor [Niclass et al., 2008].

The elementary building block of semiconductor diodes and hence of photodiodes is the $p-n$ junction, namely the boundary between two types of semiconductor materials, p-type and n-type. This is created by *doping* which is a process that consists of adding impurities into an extremely pure semiconductor for the purpose of modifying its electrical (conductivity) properties. Materials conduct electricity if they contain mobile charge carriers. There are two types of charge carriers in a semiconductor: *free electrons* and *electron holes*. When an electric field exists in the vicinity of the junction, it keeps *free electrons* confined on the n-side and *electron holes* confined on the p-side.

There are two types of $p-n$ diodes: forward-bias and reverse-bias. In forward-bias, the p-side is connected with the positive terminal of an electric power supply and the n-side is connected with the negative terminal. Both electrons and holes are pushed towards the junction. In reverse-bias, the p-side is connected with the negative terminal and the n-side is connected with the positive terminal. Otherwise said, the voltage on the n-side is higher than the voltage on the p-side. In this case both electrons and holes are pushed away from the junction.

A photodiode is a semiconductor diode that converts light into current. The current is generated when photons are absorbed in the photodiode. Photodiodes are similar to regular semiconductor diodes. A photodiode is designed to operate in reverse bias. An APD is a variation of a $p-n$ junction photodiode. When an incident photon of sufficient energy is absorbed in the region where the field exists, an electron-hole pair is generated. Under the influence of the field, the electron drifts to the n-side and the hole drifts to the p-side, resulting in the flow of *photocurrent* (i.e., the current induced by the detection of photons) in the external circuit. When a photodiode is used to detect light, the number of electron-hole pairs generated per incident photon is at best unity.

An APD detects light by using the same principle [Aull et al., 2002]. The difference between an APD and an ordinary $p-n$ junction photodiode is that an APD is designed to support high electric fields. When an electron-hole pair is generated by photon absorption, the electron (or the hole) can accelerate and gain sufficient energy from the field to collide with the crystal lattice and generate another electron-hole pair, losing some of its kinetic energy in the process. This process is known as *impact ionization*. The electron can accelerate again, as can the secondary electron or hole, and create more electron-hole pairs, hence the term "avalanche".

After a few transit times, a competition develops between the rate at which electron-hole pairs are being generated by *impact ionization* (analogous to a birth

[1] As the light travels at 3×10^{10} cm/s, 1 ns (or 10^{-9} s) corresponds to 30cm.

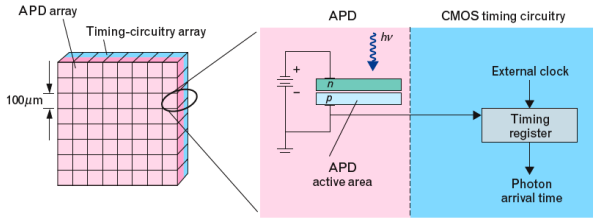

Fig. 1 The basic structure of an array of time of flight detectors consists of Geiger-mode APDs (pink) bonded to complementary-metal-oxide-semiconductor (CMOS) timing circuitry (blue). A photon of energy hv is absorbed in the APD active area. The gain of the APD resulting from the electron avalanche is great enough that the detector generates a pulse that can directly trigger the 3.3 V CMOS circuitry. No analog-to-digital converter is needed. A digital logic latch is used to stop a digital timing register for each pixel. The time of flight is recorded in the digital value of the timing register.

rate) and the rate at which they exit the high-field region and are *collected* (analogous to a death rate). If the magnitude of the reverse-bias voltage is below a value known as the *breakdown voltage*, collection wins the competition, causing the population of electrons and holes to decline. If the magnitude of the voltage is above the breakdown voltage, impact ionization wins. This situation represents the most commonly known mode of operation of APDs: measuring the intensity of an optical signal and taking advantage of the internal gain provided by impact ionization. Each absorbed photon creates on average a finite number M of electron-hole pairs. The internal gain M is typically tens or hundreds. Because the average photocurrent is strictly proportional to the incident optical flux, this mode of operation is known as linear mode.

2.2 Single Photon Avalanche Diodes

The fundamental difference between SPADs (also referred to as Geiger-mode APD) and conventional APDs is that SPADs are specifically designed to operate with a reverse-bias voltage well above the breakdown voltage. A SPAD is able to detect low intensity incoming light (down to the single photon) and to signal the arrival times of the photons with a jitter of a few tens of picoseconds [Cova et al., 1981]. Moreover, SPADs behave almost like digital devices, hence subsequent signal processing can be greatly simplified. The basic structure of an array of time-of-flight detectors consists of SPADs bonded to CMOS timing circuitry. A SPAD outputs an analog voltage pulse, that reflects the detection of a single photon, and that can directly trigger the CMOS circuitry. The latter implements time-to-digital converters to compute time-interval measurements between a start signal, global to all the pixels, and photon arrival times in individual pixels.

3 LIDAR Cameras

3.1 Velodyne Range Scanners

Whereas most LIDAR systems have a single laser that fires onto a rotating mirror and hence are only able to view objects in a single plane, the high-definition HDL-64E LIDAR range scanner from Velodyne[2] uses a rotating head featuring 64 semiconductor lasers. Each laser operates at 905 nm wavelength, has a beam divergence of 2 mrad, and fires 5 ns light pulses at up to 20,000 Hz. The 64 lasers are spread over a vertical field of view, and coupled with 64 dedicated photo detectors for precise ranging. The laser-detector pairs are precisely aligned at vertical angles to give a 26.8° vertical field of view. By spinning the entire unit at speeds of up to 900 rpm (15 Hz) around its vertical axis, a 360° field-of-view is generated.

Fig. 2 From left to right: VLP-16, HDL-32E, and HDL-64E high-definition range scanners manufactured by Velodyne. These rotating scanners feature a complete 360° horizontal field of view as well as 16, 32 and 64 laser-detector pairs spread over a vertical field of view. The sensors range is from 1 m and up to 120 m (depending on the material properties of the scanned object) with a accuracy of 2 cm.

This allows the sensor to achieve data collection rates that are an order of magnitude higher than most conventional designs. Over 1.3 million data points are generated each second, independent of the spin rate. Velodyne also commercializes 32 and 16 laser scanners with reduced vertical resolution. The range scanners are shown on figure 2 and their main parameters and specifications are summarized on Table 1.

[2] `http://velodynelidar.com/index.html`

A mathematical model of the Velodyne HDL-64E range scanner was developed in [Glennie, 2007, Glennie and Lichti, 2010] together with a calibration model and a practical method for estimating the model parameters. Extensive experiments using this model show that the actual noise level of the range measurements is 3.0 to 3.5 cm, which is double the manufacturer specification. Subsequently, the same authors analyzed the temporal stability (horizontal angle offset) of the scanner [Glennie and Lichti, 2011].

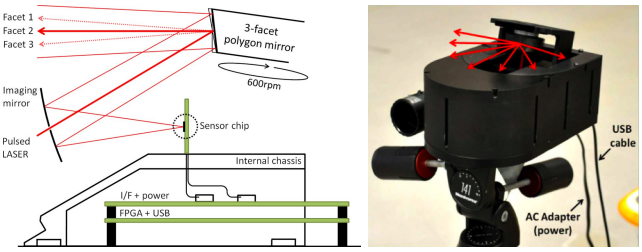

Fig. 3 A simplified diagram of a depth sensor system developed by Toyota (left) and a view of the complete system (right).

3.2 Toyota's Hybrid LIDAR Camera

Fig. 3 (left) shows a simplified diagram of the depth sensor system recently developed at the Toyota Central R&D Labs, Japan [Niclass et al., 2013]. A 870 nm pulsed laser source with a repetition rate of 200 kHz emits an optical beam with 1.5° and 0.05° of divergence in the vertical and horizontal directions, respectively. While the optical pulse duration is 4 ns full-width at half-maximum (FWHM), the mean optical power is 40 mW. The laser beam is coaxially aimed at the three-facet polygonal mirror through an opening in the center of an imaging concave mirror. Each facet of the polygonal mirror has a slightly different tilt angle.

As a result, in one revolution of 100 ms, the polygonal mirror reflects the laser beam into three vertical directions at $+1.5°$, $0°$, and $-1.5°$, thus covering, together with the laser vertical divergence, a contiguous vertical FOV of 4.5°. During the 170° horizontal scanning, at one particular facet, back-reflected photons from the targets in the scene are collected by the same facet and imaged onto the CMOS sensor chip at the focal plane of the concave mirror. The chip has a vertical line sensor with 32 macro-pixels. These pixels resolve different vertical portions of the scene at different facet times, thus generating an actual vertical resolution of 96 pixels. Since each macro-pixel circuit operates in full parallelism, at the end of a complete revolution, 1020×32 distance points are computed. This image frame is then repartitioned into 340×96 actual pixels at 10 FPS. An optical near-infrared interference filter (not shown in the figure) is also placed in front of the sensor for background light rejection.

The system electronics consists of a rigid-flex head-sensor PCB, a laser driver board, a board for signal interface and power supply, and a digital board comprising a low-cost FPGA and USB transceiver. Distance, intensity, and reliability data are generated on the FPGA and transferred to a PC at a moderate data rate of 10 Mbit/s. The system requires only a compact external AC adapter from which several other power supplies are derived internally.

3.3 3D Flash LIDAR Cameras

A 3D Flash LIDAR is another name used to designate a sensor that creates a 3D image (a depth value at each pixel) from a single laser pulse that is used to *flood-illuminate* the targeted scene or objects. The main difference between a LIDAR camera and a standard LIDAR device is that there is no need of a mechanical scanning mechanism, e.g., rotating mirror. Hence, a Flash LIDAR may well be viewed as a 3D *video* camera that delivers 3D images at up to 30 FPS. The general principle and basic components are shown of Fig. 4. Flash LIDARs use a light-pulse emitted by a single laser, that is reflected onto a scene object. Because the reflected light is further divided among multiple detectors, the energy fall-off is considerable. Nevertheless, the fact that there is no need for scanning represents a considerable advantage. Indeed, each individual SPAD is exposed to the optical signal for a long period of time, typically of the order of ten milliseconds. This allows for a large number of illumination cycles that can be averaged to reduce the various effects of noise.

The MIT Lincoln Laboratory reported the development of 3D Flash LIDAR long-range (up to 500 m) camera prototypes based on a short-pulse (1 ns) microchip laser, transmitting at a wavelength of 532 nm, and SPAD/CMOS imagers [Albota et al., 2002, Aull et al., 2002]. Two LIDAR camera prototypes were developed at MIT Lincoln Laboratory, one based on a 4×4 pixels SPAD/CMOS sensor combined with a two-axis rotating mirror, and one based on a 32×32 pixels SPAD/CMOS sensor.

Advanced Scientific Concepts Inc.[3] developed a 3D Flash LIDAR prototype [Stettner et al., 2008] as well

[3] `http://www.advancedscientificconcepts.com/index.html`

Model	Resolution (H×V)	Range/Accuracy	FOV	Frame rate	Points/second	Laser	Pulse width
HDL-64E	0.08° × 0.4°	2 – 120 m / 2 cm	360° × 26.8°	5-15 Hz	1,300,000	905 nm	10 ns
HDL-32E	0.08° × 1.33°	2 – 120 m / 2 cm	360° × 31.4°	5-20 Hz	700,000	905 nm	10 ns
VLP-16	0.08° × 1.87°	2 – 100 m / 2 cm	360° × 30°	5-20 Hz	300,000	905 nm	10 ns
Toyota	0.05° × 1.5°	not specified	170° × 4.5°	10 Hz	326,400	870 nm	4 ns

Table 1 The principal characteristics of the Velodyne LIDAR range scanners and of Toyota's LIDAR prototype that can operate outdoors. The maximum range depends on the material properties of the targeted object and can vary from 50 m (for pavement) to 120 m (for cars and trees). All these range scanners use class 1 (eye safe) semiconductor lasers.

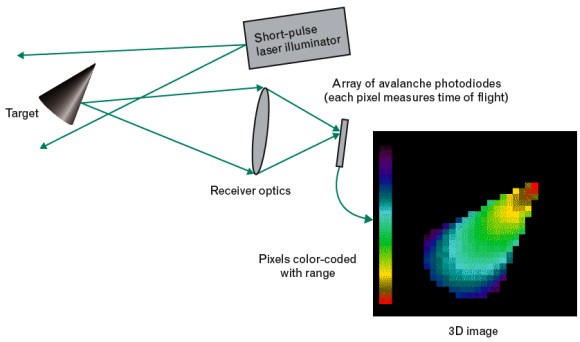

Fig. 4 This figure illustrates the basic principle of a Flash LIDAR camera. An eye-safe laser flood-illuminates an object of interest.

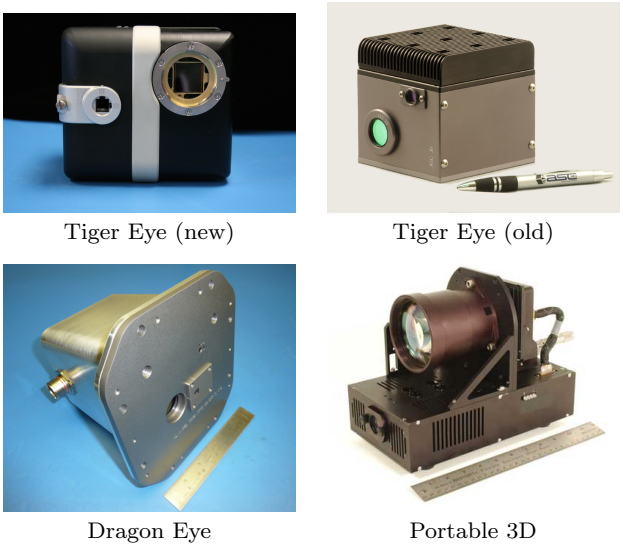

Fig. 5 3D Flash LIDAR depth cameras manufactured and commercialized by Advanced Scientific Concepts, Santa Barbara, CA. These cameras use a single pulsed-light diffused laser beam and can operate outdoors in adverse conditions at up to 30FPS.

as a number of commercially available LIDAR cameras, e.g., Fig. 5. The TigerEye/TigerCup[4] is truly a 3D video camera. It is equipped with an eye-safe 1570 nm laser, with a CMOS 128×128 pixels sensor, and it delivers images at 10-30 FPS. It has an interchangeable lens such that its range and field of view (FOV) can vary: 3° × 3° FOV and range up to 1100 meters, 8.6° × 8.6° FOV and range up to 450 meters, 45° × 22° FOV and range up to 150 meters, and 45° × 45° FOV and range up to 60 meters.

The DragonEye/GoldenEye[5] 3D Flash LIDAR space camera delivers both intensity and depth videos at 10 FPS. It has a 128×128 SPAD based sensor and its FOV is of 45° × 45° which is equivalent of a 17 mm focal length and it can range up to 1500 m. The DragonEye was tested and used by NASA for precision navigation and safe landing [Amzajerdian et al., 2011]. The specifications of these cameras are summarized in Table 2.

3.4 Other LIDAR camera

Recently, Odos Imaging[6] announced the commercialization of a high-resolution pulsed-light time-of-flight camera, Fig. 6. The camera has a resolution of 1280×1024 pixels, a range up to 10 m, a frame rate of 30 FPS and up to 450 FPS, depending on the required precision (Table 2. It can be used both indoor and outdoor (for outdoor applications it may require additional filters). One advantage of this camera is that it delivers both depth and standard monochrome images. Another LIDAR camera is Basler's pulsed-light camera based on a Panasonic TOF CCD sensor. The main characteristics of these cameras are summarized in Table 2.

4 Continuous-Wave Technology

All these depth sensors share some common characteristics, as follows [Lange and Seitz, 2001, Remondino and Stoppa, 2013]:

[4] `http://www.advancedscientificconcepts.com/products/tigercub.html`

[5] `http://www.advancedscientificconcepts.com/products/portable.html`

[6] `http://www.odos-imaging.com/`

Camera	Resolution	Range	Mult. cameras	FOV	FPS	Laser	Indoor/out
TigerEye-1	128×128	1100 m	not specified	3° × 3°	10-30	1570 nm	no/yes
TigerEye-2	128×128	450 m	not specified	8.6° × 8.6°	10-30	1570 nm	no/yes
TigerEye-3	128×128	150 m	not specified	45° × 22°	10-30	1570 nm	no/yes
TigerEye-4	128×128	60 m	not specified	45° × 45°	10-30	1570 nm	no/yes
DragonEye	128×128	1500 m	not specified	45° × 45°	10-30	1570 nm	no/yes
Real.iZ VS-1000	1280×1024	10 m	possible	45° × 45°	30-450	905 nm	yes/yes
Basler	640×480	6.6 m	not specified	57° × 43°	15	not specified	yes/yes

Table 2 This table summarizes the main features of commercially available 3D Flash LIDAR cameras. The accuracy of the depth measurements announced by the manufacturers are not reported in this table as the precision of the measurements depend on a lot of factors, such as the surface properties of the scene objects, illumination conditions, frame rate, etc.

Fig. 6 The high-resolution real.iZ pulsed-light LIDAR camera manufactured by Odos Imaging.

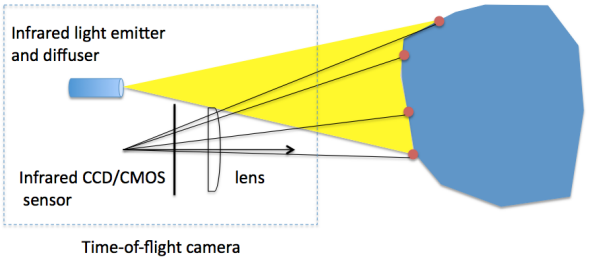

Fig. 7 This figure shows the image formation principle of 3D Flash LIDARs and continuous-wave TOF cameras.

- The transmitter, a light emitter (generally a LED, or light-emitting diode) sends light onto an object and the time the light needs to travel from the illumination source to the object and back to the sensor is measured.
- In the case of *continuous-wave* (CW), the emitted signal is a sinusoidally modulated light signal.
- The received signal is *phase-shifted* due to the round trip of the light signal. Moreover, the received signal is affected by the object's reflectivity, attenuation along the optical path and background illumination.
- Each pixel independently performs demodulation of the received signal and therefore is capable of measuring both its phase delay as well as amplitude and offset (background illumination).

The imaging principle of a CW-TOF camera is shown on Fig. 7.

4.1 Demodulation Principle

Let $s(t)$ and $r(t)$ be the optical powers of the emitted and received signals respectively:

$$s(t) = a_1 + a_2 \cos(2\pi f t), \tag{1}$$
$$r(t) = A \cos(2\pi f t - 2\pi f \tau) + B, \tag{2}$$

where f is the modulation frequency, τ is the time delay between the emitted and received signals, $\phi = 2\pi f \tau$ is the corresponding phase shift, a_1 and a_2 are the offset and amplitude of the modulated emitted signal, A is the amplitude of the received signal, and B is the offset of the received signal due to background illumination (illumination other than the emitter itself). The cross-correlation between the powers of the emitted and received signals can be written as:

$$\mathcal{C}(x) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} s(t) r(t-x) \mathrm{d}t. \tag{3}$$

By substituting $s(t)$ and $r(t)$ with their expressions (1) and (2) and by developing the terms, we obtain:

$$\begin{aligned}\mathcal{C}(x,\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} \Big(& a_2 B \cos(2\pi f t) \\ & + a_2 A \cos(2\pi f t) \cos(2\pi f t - 2\pi f(\tau + x)) \\ & + a_1 A \cos(2\pi f t - 2\pi f(\tau + x)) \Big) \mathrm{d}t \\ & + a_1 B. \end{aligned} \tag{4}$$

Using the identities

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} \cos t \, \mathrm{d}t = 0$$
$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} \cos t \cos(t-u) \, \mathrm{d}t = \frac{1}{2} \cos u,$$

we obtain:

$$\mathcal{C}(x,\tau) = \frac{a_2 A}{2}\cos(2\pi f(x+\tau)) + a_1 B. \quad (5)$$

Using the notations $\psi = 2\pi f x$ and $\phi = 2\pi f \tau$, we can write:

$$\mathcal{C}(\psi,\phi) = \frac{a_2 A}{2}\cos(\psi+\phi) + a_1 B. \quad (6)$$

Let's consider the values of the correlation function at four equally spaced samples within one modulation period, $\psi_0 = 0, \psi_1 = \pi/2, \psi_2 = \pi$, and $\psi_3 = 3\pi/2$, namely $C_0 = \mathcal{C}(0,\phi), C_1 = \mathcal{C}(\pi/2,\phi), C_2 = \mathcal{C}(\pi,\phi)$, and $C_3 = \mathcal{C}(3\pi/2,\phi)$, e.g., Fig. 8. These four sample values are sufficient for the unambiguous computation of the offset B, amplitude A, and phase ϕ [Lange and Seitz, 2001]:

$$\phi = \arctan\left(\frac{C_3 - C_1}{C_0 - C_2}\right) \quad (7)$$

$$A = \frac{1}{a_2}\sqrt{(C_3 - C_1)^2 + (C_0 - C_2)^2} \quad (8)$$

$$B = \frac{1}{4a_1}(C_0 + C_1 + C_2 + C_3) \quad (9)$$

4.2 Pixel Structure

An electro-optical demodulation pixel performs the following operations, [Büttgen and Seitz, 2008]:

- light detection, the incoming photons are converted into electron charges;
- demodulation (based on correlation),
- clocking, and
- charge storage.

The output voltage of the storage capacitor, after integration over a short period of time T_i, is proportional to the correlation (R is the optical responsivity of the detector):

$$V(x) = \frac{RT_i}{C_S}C(x) \quad (10)$$

Hence, the received optical signal is converted into a photocurrent.

The samples are the result of the integration of the photocurrent of a duration $\Delta t < 1/f$. In order to increase the signal-to-noise ratio of one sampling process, the samples C_0 to C_3 are the result of the summation over many modulation periods (up to hundreds of thousands). A pixel with four shutters feeding four charge storage nodes allows the simultaneous acquisition of the four samples needed for these computations. The four shutters are activated one at a time for a time equal to $T/4$ (where T is the modulation period), and the shutter activation sequence is repeated for the whole integration time T_i which usually includes hundreds of thousands of modulation periods.

4.3 Depth Estimation from Phase

A depth value d **at each pixel** is computed with the following formula:

$$d = \frac{1}{2}c\tau \quad (11)$$

where c is the light speed and τ is the time of flight. Since we measure the phase $\phi = 2\pi f\tau$, we obtain:

$$d = \frac{c}{4\pi f}\phi = \frac{c}{4\pi f}\arctan\left(\frac{C_3 - C_1}{C_0 - C_1}\right) \quad (12)$$

Nevertheless, because the phase is defined up to 2π, there is an inherent **phase wrapping ambiguity** in measuring the depth:

- Minimum depth: $d_{\min} = 0 \;\; (\phi = 0)$
- Maximum depth: $d_{\max} = \frac{c}{2f} \;\; (\phi = 2\pi)$.

The depth at each pixel location (i,j) can be written as a function of this wrapping ambiguity:

$$d(i,j) = \left(\frac{\phi(i,j)}{2\pi} + n(i,j)\right) d_{\max} \quad (13)$$

where $n = 0, 1, 2, \ldots$ is the number of wrappings. This can also be written as:

$$d(i,j) = d_{TOF}(i,j) + n(i,j)d_{\max} \quad (14)$$

where d is the *real* depth value and d_{TOF} is the *measured* depth value. It is important to stress that the number of wrappings is not the same at each pixel. Let's consider a modulation frequency $f = 30$ MHz, the unambiguous range of the camera is in this case from $d_{\min} = 0$ to $d_{\max} = 5$ meters. The ambiguity decreases as the modulation frequency decreases, but in the same time the accuracy decreases as well. Several methods were proposed in the literature to solve for the phase wrapping ambiguity [Ghiglia and Romero, 1994, Oprişescu et al., 2007, Bioucas-Dias and Valadão., 2007, Payne et al., 2009, McClure et al., 2010, Droeschel et al., 2010b,a, Choi et al., 2010, Choi and Lee, 2012]

To summarize, the following remarks can be made concerning these depth-camera technologies:

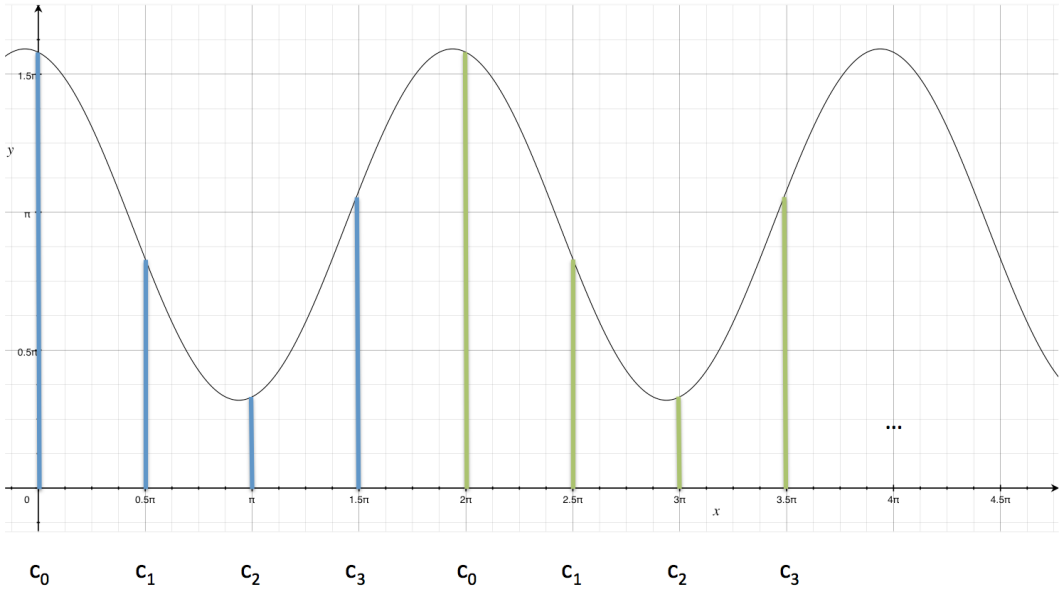

Fig. 8 This figure shows the general principle of the four-bucket method that estimates the demodulated optical signal at four equally-spaced samples in one modulation period. A CCD/CMOS circuit achieves light detection, demodulation, and charge storage. The demodulated signal is stored at four equally spaced samples in one modulation period. From these four values, it is then possible to estimate the phase and amplitude of the received signal as well as the amount of background light (offset).

– A CW-TOF camera works at a very precise modulation frequency. Consequently, it is possible to simultaneously and synchronously use several CW-TOF cameras, either by using a different modulation frequency for each one of the cameras, e.g., six cameras in the case of the SR4000 (Swiss Ranger), or by encoding the modulation frequency, e.g., an arbitrary number of SR4500 cameras.
– In order to increase the signal-to-noise ratio, and hence the depth accuracy, CW-TOF cameras need a relatively long integration time (IT), over several time periods. In turn, this introduces *motion blur* [Hansard et al., 2013] (chapter 1) in the presence of moving objects. Because of the need of long IT, fast shutter speeds (as done with standard cameras) cannot be envisaged.

To summarize, the sources of errors of these cameras are: demodulation, integration, temperature, motion blur, distance to the target, background illumination, phase wrapping ambiguity, light scattering, and multiple path effects. A quantitative analysis of these sources of errors is available in [Foix et al., 2011]. In the case of several cameras operating simultaneously, interferences between the different units is an important issue.

5 TOF Cameras

In this section we review the characteristics of some of the commercially available cameras. We selected those camera models for which technical and scientific documentation is readily available. The main specifications of the overviewed camera models are summarized in Table 3.

5.1 The SR4000/SR4500 Cameras

The SR4000/4500 cameras, figure 9, are manufactured by Mesa Imaging, Zurich, Switzerland.[7] They are continuous-wave TOF cameras that provide depth, amplitude, and confidence images with a resolution of 176×144 pixels. In principle, the cameras can work at up to 30 FPS but in practice more accurate depth measurements are obtained at 10-15 FPS.

The modulation frequency of the SR4000 camera can be selected by the user. The camera can be operated at:

– 29 MHz, 30 MHz, or 31 MHz corresponding to a maximum depth of 5.17 m, 5 m and 4.84 m respectively.

[7] `http://www.mesa-imaging.ch/`

Fig. 9 The SR4000 (left) and SR4500 (right) CW-TOF cameras manufactured by Mesa Imaging.

Fig. 10 The DS311 (left) and DS325 (rigtht) CW-TOF cameras manufactured by SoftKinetic.

– 14.5 MHz, 15.0 MHz, or 15.5 MHz corresponding to a maximum depth of 10.34 m, 10 m and 9.67 m respectively.

This allows the simultaneous and synchronous use of up to six SR4000 cameras to be used together with any number of color cameras.

The modulation frequency of SR4500 is of 13.5 MHz which allows a maximum depth of 9 m. Moreover, an arbitrary number of SR4500 cameras can be combined together because the modulation frequency is encoded differently for each unit.

5.2 The Kinect v2 RGB-D Camera

The Kinect color and depth (RGB-D) camera, manufactured by Microsoft, was recently upgraded to Kinect v2. Unlike the former version that was based on structured-light technology, the latter uses a time-of-flight sensor [Payne et al., 2014, Bamji et al., 2015] and was mainly designed for gaming [Sell and O'Connor, 2014]. Kinect-v2 achieves one of the best image resolution among TOF cameras commercially available. Moreover, it uses multiple modulation frequencies (10-130 MHz) thus achieving an excellent compromise between depth accuracy and phase unwrapping, i.e. Section 4.3 above. In [Bamji et al., 2015] it is reported that the Kinect v2 can measure depth in the range 0.8-4.2 m with an accuracy of 0.5% of the measured range. Several recent articles evaluate the Kinect v2 sensor for mobile robotics [Fankhauser et al., 2015] and in comparison with the structured-light version [Sarbolandi et al., 2015].

It is interesting to note that Kinect v2 is heavier than its predecessor (970 g instead of 170 g) requires higher voltage (12 V instead of 5 V) and power usage (15 W instead of 2.5 W).

5.3 Other CW TOF cameras

The following depth cameras are based on the same continuous wave demodulation principles (see Table 3 for a summary of the characteristics of these cameras):

– DS311 and DS325 cameras, figure 10, manufactured by SoftKinetic,[8]
– E70 and E40 manufactured by Fotonic,[9]
– TOF sensor chip manufactured by PMD.[10],
– The D-imager manufactured by Panasonic has a range up to 15 cm. It was discontinued in March 2015.[11]

6 Calibration of Time-of-Flight Cameras

Both pulsed-light and continuous-wave TOF cameras can be modeled as pinhole cameras, using the principles of projective geometry. The basic projection equation is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{15}$$

This implies that the homogeneous coordinates of an image-point $\boldsymbol{p} = (x, y, 1)^\top$ are projectively equal to the scene-coordinates $\boldsymbol{P} = (X, Y, Z)^\top$, specifically:

$$Z\boldsymbol{p} = \boldsymbol{P}. \tag{16}$$

In practice, a realistic model of the projection process involves the intrinsic, extrinsic, and distortion parameters, as described below [Zhang, 2000, Hartley and Zisserman, 2003, Bradski and Kaehler, 2008].

6.1 Intrinsic Parameters

A digital camera records the image in pixel-units, which are related to the coordinates $(x, y)^\top$ in (15) by

$$\begin{aligned} u &= \alpha_u x + u_0 \\ v &= \alpha_v y + v_0. \end{aligned} \tag{17}$$

[8] http://www.softkinetic.com/
[9] http://www.fotonic.com/
[10] http://www.pmdtec.com/
[11] http://www2.panasonic.biz/es/densetsu/device/3DImageSensor/en/

Camera	Resolution	Range	Mult. cameras	FOV	Max FPS	Illumination	Indoor/out
SR4000	176×144	0−5 or 0−10 m	6 cameras	43° × 34°	30	LED	yes/no
SR4500	176×144	0−9 m	many cameras	43° × 34°	30	LED	yes/no
DS311	160×120	0.15−1 or 1.5−4.5 m	not specified	57° × 42°	60	LED	yes/no
DS325	320×240	0.15−1 m	not specified	74° × 58°	60	diffused laser	yes/no
E70	160×120	0.1−10 m	4 cameras	70° × 53°	52	LED	yes/yes
E40	160×120	0.1−10 m	4 cameras	45° × 34°	52	LED	yes/yes
Kinect v2	512×424	0.8−4.2 m	not specified	70° × 60°	30	LED	yes/no

Table 3 This table summarizes the main features of commercially available CW TOF cameras. The accuracy of the depth measurements announced by the manufacturers are not reported in this table as the precision of the measurements depend on a lot of factors, such as the surface properties of the scene objects, illumination conditions, frame rate, etc.

In these equations we have the following *intrinsic parameters*:

- Horizontal and vertical factors, α_u and α_v, which encode the change of scale, multiplied by the focal length.
- The image center, or *principal point*, expressed in pixel units: (u_0, v_0).

Alternatively, the intrinsic transformation (17) can be expressed in matrix form, $\boldsymbol{q} = \mathbf{A}\boldsymbol{p}$ where $\boldsymbol{q} = (u, v, 1)$ are pixel coordinates, and the 3×3 matrix $\mathbf{A}$ is defined as

$$\mathbf{A} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{18}$$

Hence it is possible to express the *direction* of a visual ray, in camera coordinates, as

$$\boldsymbol{p} = \mathbf{A}^{-1}\boldsymbol{q}. \tag{19}$$

A TOF camera further allows the 3D *position* of point $\boldsymbol{P}$ to be estimated, as follows. Observe from equation (16) that the Euclidean norms of $\boldsymbol{P}$ and $\boldsymbol{p}$ are proportional:

$$\|\boldsymbol{P}\| = Z\|\boldsymbol{p}\|. \tag{20}$$

The TOF camera measures the distance d from the 3D point $\boldsymbol{P}$ to the optical center,[12] so $d = \|\boldsymbol{P}\|$. Hence the Z coordinate of the observed point is

$$Z = \frac{\|\boldsymbol{P}\|}{\|\boldsymbol{p}\|} = \frac{d}{\|\mathbf{A}^{-1}\boldsymbol{q}\|}. \tag{21}$$

We can therefore obtain the 3D coordinates of the observed point, by combining (16) and (19), to give

$$\boldsymbol{P} = \frac{d}{\|\mathbf{A}^{-1}\boldsymbol{q}\|}\mathbf{A}^{-1}\boldsymbol{q}. \tag{22}$$

Note that the point is recovered in the camera coordinate system; the transformation to a common 'world' coordinate system is explained in the following section.

[12] In practice it measures the distance to the image sensor and we assume that the offset between the optical center and the sensor is small

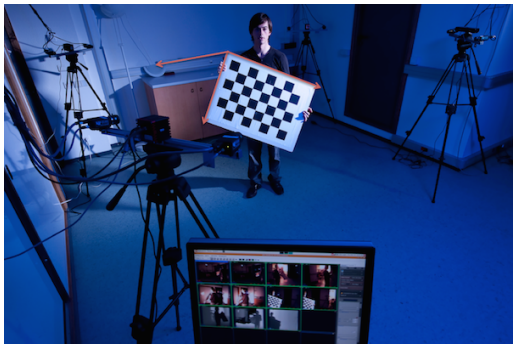

Fig. 11 This figure shows a setup for TOF calibration. The calibration board is the same one used for color camera calibration and it can be used to estimate the lens parameters as well.

6.2 Extrinsic Parameters

A rigid transformation from the arbitrary *world coordinate frame* $\boldsymbol{P}_w = (X_w, Y_w, Z_w)^\top$ to the camera frame can be modelled by a rotation and translation. This can be expressed in homogeneous coordinates as:

$$\begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \boldsymbol{T} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{P}_w \\ 1 \end{pmatrix}. \tag{23}$$

The 3×3 matrix $\mathbf{R}$ has three degrees of freedom, which can be identified with the angle and normalized axis of rotation. Meanwhile, the 3×1 translation vector is defined by $\boldsymbol{T} = -\mathbf{R}\boldsymbol{C}$, where $\boldsymbol{C}$ contains the world coordinates of the camera-centre. Hence there are six *extrinsic parameters* in the transformation (23). The equation can readily be inverted, in order to obtain world coordinates from the estimated point $\boldsymbol{P}$ in equation (22).

6.3 Lens Distortion Model

A commonly used lens distortion model widely used for color cameras, [Bradski and Kaehler, 2008, Gonzalez-Aguilera et al., 2011], can be adopted for TOF cameras

as well: the observed distorted point (x_l, y_l) results from the displacement of (x, y) according to:

$$\begin{pmatrix} x_l \\ y_l \end{pmatrix} = l_\rho(r) \begin{pmatrix} x \\ y \end{pmatrix} + \boldsymbol{l}_\tau(x, y) \tag{24}$$

where $l_\rho(r)$ is a scalar radial function of $r = \sqrt{x^2 + y^2}$, and $\boldsymbol{l}_\tau(x, y)$ is a vector tangential component. These are commonly defined by polynomial functions

$$l_\rho(r) = 1 + \rho_1 r^2 + \rho_2 r^4 \quad \text{and} \tag{25}$$

$$\boldsymbol{l}_\tau(x, y) = \begin{bmatrix} 2xy & r^2 + 2x^2 \\ r^2 + 2y^2 & 2xy \end{bmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} \tag{26}$$

such that the complete parameter-vector is $[\rho_1\ \rho_2\ \tau_1\ \tau_2]$. The images can be undistorted, by numerically inverting (26), given the lens and intrinsic parameters. The projective linear model, described in sections (6.1–6.2), can then be used to describe the complete imaging process, with respect to the undistorted images.

It should also be noted, in the case of TOF cameras, that the outgoing infrared signal is subject to optical effects. In particular, there is a radial attenuation, which results in strong vignetting of the intensity image. This can be modeled by a bivariate polynomial, if a full photometric calibration is required [Lindner et al., 2010, Hertzberg and Frese, 2014].

6.4 Depth Distortion Models

TOF depth estimates are also subject to systematic nonlinear distortions, particulalry due to deviation of the emitted signal from the ideal model described in Section 4.1. This results in 'wiggling' error of the average distance estimates, with respect to the true distance [Foix et al., 2011, Fursattel et al., 2016]. Because this error is systematic, it can be removed by reference to a precomputed look-up table [Kahlmann et al., 2006]. Another possibility is to learn a mapping between raw depth values, estimated by the sensor, and corrected values. This mapping can be performed using regression techniques applied to carefully calibrated data. A random regression forest is used in [Ferstl et al., 2015] to optimize the depth measurements supplied by the camera. A kernel regression method based on a Gaussian kernel is used in [Kuznetsova and Rosenhahn, 2014] to estimate the depth bias at each pixel. Below we describe an efficient approach, which exploits the smoothness of the error, and which uses a B-spline regression [Lindner et al., 2010] of the form:

$$d'(x, y) = d(x, y) - \sum_i^n \beta_i B_{i,3}\big(d(x, y)\big) \tag{27}$$

where $d'(x, y)$ is the corrected depth. The spline basis-functions $B_{i,3}(d)$ are located at n evenly-spaced depth control-points d_i. The coefficients β_i, $i = 1, \ldots, n$ can be estimated by least-squares optimization, given the known target-depths. The total number of coefficients n depends on the number of known depth-planes in the calibration procedure.

6.5 Practical Considerations

To summarize, the TOF camera parameters are composed of the pinhole camera model, namely the parameters $\alpha_u, \alpha_v, u_0, v_0$ and the lens distortion parameters, namely $\rho_1, \rho_2, \tau_1, \tau_2$. Standard camera calibration methods can be used with TOF cameras, in particular with CW-TOF cameras because they provide an amplitude+offset image, i.e. Section 4, together with the depth image: Standard color-camera calibration methods, e.g. OpenCV packages, can be applied to the amplitude+offset image. However, the low-resolution of the TOF images implies specific image processing techniques, such as [Hansard et al., 2014, Kuznetsova and Rosenhahn, 2014]. As en example, Fig. 12 shows the depth and amplitude images of a calibration pattern gathered with the SR4000 camera. Here the standard corner detection method was replaced with the detection of two pencils of lines that are optimally fitted to the OpenCV calibration pattern.

TOF-specific calibration procedures can also be performed, such as the depth-wiggling correction [Lindner et al., 2010, Kuznetsova and Rosenhahn, 2014, Ferstl et al., 2015]. A variety of TOF calibration methods can be found in a recent book [Grzegorzek et al., 2013]. A comparative study of several TOF cameras based on an error analysis was also proposed [Fursattel et al., 2016].

One should however be aware of the fact that depth measurement errors may be quite difficult to predict due to the unknown material properties of the sensed objects and to the complexity of the scene. In the case of a scene composed of complex objects, multiple-path distortion may occur, due to the interaction between the emitted light and the scene objects, e.g. the emitted light is backscattered more than once. Techniques for removing multiple-path distortions were recently proposed [Freedman et al., 2014, Son et al., 2016].

7 Combining Multiple TOF and Color Cameras

In addition to the image and depth-calibration procedures described in section 6, it is often desirable to

Fig. 12 The depth (left) and amplitude (middle) images of an OpenCV calibration pattern grabbed with an SR4000 camera. The actual image resolution is of 176 × 144 pixels. The amplitude image (middle) has undergone lens undistortion using the model briefly described in section 6.3. The calibration pattern detected in the amplitude image (right) using the method of [Hansard et al., 2014].

combine data from multiple devices. There are several reasons for this, depending on the application. Firstly, current depth cameras are pinhole devices, with a single optical centre (16). This means that the resulting point-cloud data is viewpoint dependent; in particular, depth discontinuities will give rise to holes, when the point cloud is viewed from a different direction. Secondly, it is often desirable to combine multiple point clouds, in order to reduce the amount and the anisotropy of sensor noise. Thirdly, typical TOF devices do not capture colour data; hence it may be necessary to cross-calibrate one or more RGB cameras with the depth sensors.

Both pulsed-light and continuous-wave TOF cameras work at a precise wavelength in the (near-)infrared domain. They are equipped with an optical band-pass filter properly tuned onto the wavelength of the light emitter. This allows, in principle, simultaneous capture from multiple TOF cameras with each signaling at its own modulation frequency, so that interference is unlikely. As mentioned above, not all of the TOF manufacturers allow a user-defined modulation frequency (see Table 3). Because of different spectra, TOF cameras do not interfere with typical color cameras, and can be easily combined. In either combination, however, the cameras must be synchronized properly, that is, a common clock signal should trigger all of the cameras.

The simplest possible combination is one TOF and one color camera. Some of the available TOF cameras, e.g., Microsoft Kinect v2 and SoftKinetic DS311/DS325 have a built-in color camera with its own sensor and lens, which is mounted a few centimeters away from the TOF camera. Note that only real.iZ by Odos Imaging uses the same pixel to acquire both color and depth measurements, thus eliminating the additional calibration and registration. [13].

Another possible setup is to use one TOF camera and two color cameras [Gudmundsson et al., 2008, Zhu et al., 2008, Gandhi et al., 2012, Evangelidis et al., 2015, Mutto et al., 2015], e.g., Fig. 14. The advantage of using two color cameras is that they can be used as a stereoscopic camera pair. Such a camera pair, once calibrated, provides dense depth measurements (via a stereo matching algorithm) when the scene is of sufficient texture and lacks repetitive patterns. However, untextured areas are very common in man-made environments, e.g. walls, and the matching algorithm typically fails to reconstruct such scenes. While TOF cameras have their own limitations (noise, low resolution, etc.) that were discussed above, they provide good estimates regardless of the scene texture. This gives rise to *mixed* systems that combine active-range and the passive-parallax approaches and overcome the limitations of each approach alone. In particular, when a high-resolution 3D map is required, such a mixed system is highly recommended. Roughly speaking, sparse TOF measurements are used as a regularizer of a stereo matching algorithm towards a dense high-resolution depth map [Evangelidis et al., 2015].

Given that TOF cameras can be modeled as pin-hole cameras, one can use multiple-camera geometric models to *cross-calibrate* several TOF cameras or any combination of TOF and color cameras. Recently, an earlier TOF-stereo calibration technique [Hansard et al., 2011] was extended to deal with an arbitrary number of cameras [Hansard et al., 2015]. It is assumed that a set of calibration vertices can be detected in the TOF images, and back-projected into 3D. The same vertices are re-

[13] There has been an attempt at a similar architecture in [Kim et al., 2012]; this 3D and color camera is not commercially available

constructed via stereo-matching, using the high resolution RGB cameras. Ideally, the two 3D reconstructions could be aligned by a rigid 3D transformation; however, this is not true in practice, owing to calibration uncertainty in both the TOF and RGB cameras. Hence a more general alignment, via a 3D projective transformation, was derived. This approach has several advantages, including a straightforward SVD calibration procedure, which can be refined by photogrammetric bundle-adjustment routines. An alternative approach, initialized by the factory calibration of the TOF camera, is described by [Jung et al., 2014].

A different cross-calibration method can be developed from the constraint that depth points lie on calibration planes, where the latter are also observed by the colour camera [Zhang and Zhang, 2011]. This method, however, does not provide a framework for calibrating the intrinsic (section 6.1) or lens (section 6.3) parameters of the depth camera. A related method, which does include distortion correction, has been demonstrated for Kinect v1 [Herrera et al., 2012].

Finally, a more object-based approach can be adopted, in which dense overlapping depth-scans are merged together in 3D. This approach, which is ideally suited to handheld (or robotic) scanning, has been applied to both Kinect v1 [Newcombe et al., 2011] and TOF cameras [Cui et al., 2013].

8 Conclusions

Time of flight is a remote-sensing technology that estimates range (depth) by illuminating an object with a laser or with a photodiode and by measuring the travel time from the emitter to the object and back to the detector. Two technologies are available today, one based on pulsed-light and the second based on continuous-wave modulation. Pulsed-light sensors measure directly the pulse's total trip time and use either rotating mirrors (LIDAR) or a light diffuser (Flash LIDAR) to produce a two-dimensional array of range values. Continuous-wave (CW) sensors measure the phase difference between the emitted and received signals and the phase is estimated via demodulation. LIDAR cameras usually operate outdoor and their range can be up to a few kilometers. CW cameras usually operate indoor and they allow for short-distance measurements only, namely up to 10 meters. Depth estimation based on phase measurement suffer from an intrinsic phase-wrapping ambiguity. Higher the modulation frequency, more accurate the measurement and shorter the range.

Generally speaking, the spatial resolution of these sensors is a few orders of magnitude (10 to 100) less than video cameras. This is mainly due to the need to capture sufficient backscattered light. The depth accuracy depends on multiple factors and can vary from a few centimeters up to several meters. TOF cameras can be modeled as pinhole cameras and therefore one can use standard camera calibration techniques (distortion, intrinsic and extrinsic parameters). One advantage of TOF cameras over depth sensors based on structured light and triangulation, is that the former provides an amplitude+offset image. The amplitude+offset and depth images are gathered by the same and unique sensor, hence one can use camera calibration techniques routinely used with color sensors to calibrate TOF cameras. This is not the case with triangulation-based range sensors for which special-purpose calibration techniques must be used. Moreover, the relative orientation between the infrared camera and the color camera must be estimated as well.

TOF cameras and TOF range scanners are used for a wide range of applications, from multimedia user interfaces to autonomous vehicle navigation and planetary/space exploration. More precisely:

- Pulsed-light devices, such as the Velodyne, Toyota, and Advanced Scientific Concepts LIDARs can be used under adverse outdoor lighting conditions, which is not the case with continuous-wave systems. These LIDARs are the systems of choice for autonomous vehicle driving and for robot navigation (obstacle, car, and pedestrian detection, road following, etc.). The Toyota LIDAR scanner is a laboratory prototype, at the time of writing.
- The 3D Flash LIDAR cameras manufactured by Advanced Scientific Concepts have been developed in collaboration with NASA for the purpose of planet landing [Amzajerdian et al., 2011]. They are commercially available.
- The SR4000/45000 cameras are used for industrial and for multimedia applications. Although they have limited image resolution, these cameras can be easily combined together into multiple TOF and camera systems. They are commercially available.
- The Kinect v2, SoftKinetic, Basler and Fotonic cameras are used for multimedia and robotic applications. They are commercially available, some of them at a very affordable price. Additionally, some of these sensors integrate an additional color camera which is internally synchronized with the TOF camera, thus yielding RGB-D data. Nevertheless, one shortcoming is that they cannot be easily externally synchro-

Fig. 13 A single system (left), comprising a time-of-flight camera in the centre, plus a pair of ordinary color cameras. Several (four) such systems can be combined together and calibrated in order to be used for 3D scene reconstruction (right).

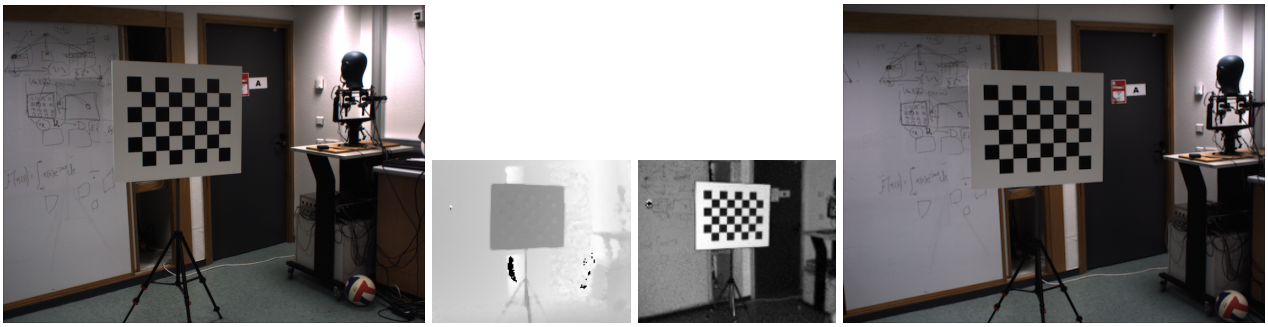

Fig. 14 Calibration images from synchronized captures. The greyscale images provided by the color camera pair are shown onto the left and onto the right. The middle smaller images correspond to enlarged depth and amplitude images provided by the TOF camera.

nized in order to build multiple-camera TOF-TOF or TOF-color systems.
- The real.iZ is a prototype developed by Odos Imaging. It is a very promising camera but its commercial availability is not clear and the time of the writing of this paper.

In conclusion, time-of-flight technologies are used in many different configurations, for a wide range of applications. The refinement, commoditization, and miniaturization of these devices is likely to have an increasing impact on everyday life, in the near future.

References

1. M. A. Albota, B. F. Aull, D. G. Fouche, R. M. Heinrichs, D. G. Kocher, R. M. Marino, J. G. Mooney, N. R. Newbury, M. E. O'Brien, B. E. Player, et al. Three-dimensional imaging laser radars with Geiger-mode avalanche photodiode arrays. *Lincoln Laboratory Journal*, 13(2):351–370, 2002.
2. F. Amzajerdian, D. Pierrottet, L. Petway, G. Hines, and V. Roback. Lidar systems for precision navigation and safe landing on planetary bodies. In *International Symposium on Photoelectronic Detection and Imaging 2011*, pages 819202–819202. International Society for Optics and Photonics, 2011.
3. B. F. Aull, A. H. Loomis, D. J. Young, R. M. Heinrichs, B. J. Felton, P. J. Daniels, and D. J. Landers. Geiger-mode avalanche photodiodes for three-dimensional imaging. *Lincoln Laboratory Journal*, 13(2):335–349, 2002.
4. S. C. Bamji, P. O'Connor, T. Elkhatib, S. Mehta, B. Thompson, L. A. Prather, D. Snow, O. C. Akkaya, A. Daniel, D. A. Payne, et al. A 0.13 μm cmos system-on-chip for a 512$\times$ 424 time-of-flight image sensor with multi-frequency photo-demodulation up to 130 mhz and 2 gs/s adc. *IEEE Journal of Solid-State Circuits*, 50(1):303–319, 2015.
5. J. M. Bioucas-Dias and G. Valadão. Phase unwrapping via graph cuts. *IEEE Transactions on Image Processing*, 16(3):698–709, 2007.
6. F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1), 2004.
7. G. Bradski and A. Kaehler. *Learning OpenCV*. O'Reilly, 2008.
8. B. Büttgen and P. Seitz. Robust optical time-of-flight range imaging based on smart pixel structures. *IEEE*

Transactions on Circuits and Systems I: Regular Papers, 55(6):1512–1525, 2008.

9. O. Choi and S. Lee. Wide range stereo time-of-flight camera. In *Proceedings IEEE International Conference on Image Processing*, 2012.
10. O. Choi, H. Lim, B. Kang, Y. S. Kim, K. Lee, J. D. K. Kim, and C. Y. Kim. Range unfolding for time-of-flight depth cameras. In *Proceedings IEEE International Conference on Image Processing*, 2010.
11. S. Cova, A. Longoni, and A. Andreoni. Towards picosecond resolution with single-photon avalanche diodes. *Review of Scientific Instruments*, 52(3):408–412, 1981.
12. Y. Cui, S. Schuon, S. Thrun, D. Stricker, and C. Theobalt. Algorithms for 3d shape scanning with a depth camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1039–1050, 2013.
13. D. Droeschel, D. Holz, and S. Behnke. Probabilistic phase unwrapping for time-of-flight cameras. In *Joint 41st International Symposium on Robotics and 6th German Conference on Robotics*, 2010a.
14. D. Droeschel, D. Holz, and S. Behnke. Multifrequency phase unwrapping for time-of-flight cameras. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010b.
15. G. D. Evangelidis, M. Hansard, and R. Horaud. Fusion of Range and Stereo Data for High-Resolution Scene-Modeling. *IEEE Trans. PAMI*, 37(11):2178 – 2192, 2015.
16. P. Fankhauser, M. Bloesch, D. Rodriguez, R. Kaestner, M. Hutter, and R. Siegwart. Kinect v2 for Mobile Robot Navigation: Evaluation and Modeling. In *International Conference on Advanced Robotics*, Istanbul, Turkey, July 2015.
17. D. Ferstl, C. Reinbacher, G. Riegler, M. Rüther, and H. Bischof. Learning depth calibration of time-of-flight cameras. Technical report, Graz University of Technology, 2015.
18. S. Foix, G. Alenya, and C. Torras. Lock-in Time-of-Flight (ToF) Cameras: A Survey. *IEEE Sensors*, 11 (9):1917–1926, Sept 2011.
19. D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt. Sra: Fast removal of general multipath for tof sensors. In *European Conference on Computer Vision*, pages 234–249. Springer, 2014.
20. P. Fursattel, S. Placht, C. Schaller, M. Balda, H. Hofmann, A. Maier, and C. Riess. A comparative error analysis of current time-of-flight sensors. *IEEE Transactions on Computational Imaging*, 2(1):27 – 41, 2016.
21. V. Gandhi, J. Cech, and R. Horaud. High-resolution depth maps based on TOF-stereo fusion. In *IEEE International Conference on Robotics and Automation*, pages 4742–4749, 2012.
22. D. C. Ghiglia and L. A. Romero. Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods. *Journal of Optical Society of America A*, 11(1):107–117, 1994.
23. C. Glennie. Rigorous 3D error analysis of kinematic scanning LIDAR systems. *Journal of Applied Geodesy*, 1(3):147–157, 2007.
24. C. Glennie and D. D. Lichti. Static calibration and analysis of the velodyne hdl-64e s2 for high accuracy mobile scanning. *Remote Sensing*, 2(6):1610–1624, 2010.
25. C. Glennie and D. D. Lichti. Temporal stability of the velodyne hdl-64e s2 scanner for high accuracy scanning applications. *Remote Sensing*, 3(3):539–553, 2011.
26. D. Gonzalez-Aguilera, J. Gomez-Lahoz, and P. Rodriguez-Gonzalvez. An automatic approach for radial lens distortion correction from a single image. *IEEE Sensors*, 11(4):956–965, 2011.
27. M. Grzegorzek, C. Theobalt, R. Koch, and A. Kolb. *Time-of-Flight and Depth Imaging. Sensors, Algorithms and Applications*, volume 8200. Springer, 2013.
28. Sigurjon Arni Gudmundsson, Henrik Aanaes, and Rasmus Larsen. Fusion of stereo vision and time-of-flight imaging for improved 3D estimation. *Int. J. Intell. Syst. Technol. Appl.*, 5(3/4), 2008.
29. M. Hansard, R. Horaud, M. Amat, and S.K. Lee. Projective alignment of range and parallax data. In *IEEE Computer Vision and Pattern Recognition*, pages 3089–3096, 2011.
30. M. Hansard, S. Lee, O. Choi, and R. Horaud. *Time-of-Flight Cameras: Principles, Methods and Applications*. Springer, 2013.
31. M. Hansard, R. Horaud, M. Amat, and G. Evangelidis. Automatic detection of calibration grids in time-of-flight images. *Computer Vision and Image Understanding*, 121:108–118, 2014.
32. M. Hansard, G. Evangelidis, Q. Pelorson, and R. Horaud. Cross-calibration of time-of-flight and colour cameras. *Computer Vision and Image Understanding*, 134:105–115, May 2015.
33. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
34. D.C. Herrera, J. Kannala, and J. Heikila. Joint Depth and Color Camera Calibration with Distortion Correction. *IEEE Trans. PAMI*, 34(10):2058–2064, 2012.
35. Christoph Hertzberg and Udo Frese. Detailed modeling and calibration of a time-of-flight camera. In *ICINCO 2014 - Proc. International Conference on Informatics*

in Control, Automation and Robotics, pages 568–579, 2014.

36. J. Jung, J-Y. Lee, Y. Jeong, and I.S. Kweon. Time-of-flight sensor calibration for a color and depth camera pair. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.
37. T. Kahlmann, F. Remondino B, and H. Ingensand. Calibration for increased accuracy of the range imaging camera swissranger. In *ISPRS Archives*, pages 136–141, 2006.
38. S.-J. Kim, J. D. K. Kim, B. Kang, and K. Lee. A CMOS image sensor based on unified pixel architecture with time-division multiplexing scheme for color and depth image acquisition. *IEEE Journal of Solid-State Circuits*, 47(11):2834–2845, 2012.
39. A. Kuznetsova and B. Rosenhahn. On calibration of a low-cost time-of-flight camera. In *ECCV Workshops*, pages 415–427, 2014.
40. R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37 (3):390–397, 2001.
41. M. Lindner, I. Schiller, A. Kolb, and R. Koch. Time-of-flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114(12): 1318–1328, 2010.
42. S. H. McClure, M. J. Cree, A. A. Dorrington, and A. D. Payne. Resolving depth-measurement ambiguity with commercially available range imaging cameras. In *Image Processing: Machine Vision Applications III*, 2010.
43. C.D. Mutto, P. Zanuttigh, and G.M. Cortelazzo. Probabilistic ToF and Stereo Data Fusion Based on Mixed Pixels Measurement Models. *IEEE Trans. PAMI*, 37 (11):2260 – 2272, 2015.
44. R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *ISMAR*, 2011.
45. C. Niclass, A. Rochas, P.-A. Besse, and E. Charbon. 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.
46. C. Niclass, C. Favi, T. Kluter, M. Gersbach, and E. Charbon. A 128×128 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.
47. C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami. A 100-m range 10-frame/s 340 96-pixel time-of-flight depth sensor in 0.18-CMOS. *IEEE Journal of Solid-State Circuits*, 48(2):559–572, 2013.
48. Ş. Oprişescu, D. Fălie, M. Ciuc, and V. Buzuloiu. Measurements with TOF cameras and their necessary corrections. In *IEEE International Symposium on Signals, Circuits & Systems*, 2007.
49. A. Payne, A. Daniel, A. Mehta, B. Thompson, C. S. Bamji, D. Snow, H. Oshima, L. Prather, M. Fenton, L. Kordus, et al. A 512×424 CMOS 3D time-of-flight image sensor with multi-frequency photo-demodulation up to 130MHz and 2GS/s ADC. In *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pages 134–135. IEEE, 2014.
50. A. D. Payne, A. P. P. Jongenelen, A. A. Dorrington, M. J. Cree, and D. A. Carnegie. Multiple frequency range imaging to remove measurement ambiguity. In *9th Conference on Optical 3-D Measurement Techniques*, 2009.
51. F. Remondino and D. Stoppa, editors. *TOF Range-Imaging Cameras*. Springer, 2013.
52. H. Sarbolandi, D. Lefloch, and A. Kolb. Kinect range sensing: Structured-light versus time-of-flight Kinect. *CVIU*, 139:1–20, October 2015.
53. B. Schwarz. Mapping the world in 3D. *Nature Photonics*, 4(7):429–430, 2010.
54. J. Sell and P. O'Connor. The Xbox one system on a chip and Kinect sensor. *IEEE Micro*, 32(2):44–53, 2014.
55. K. Son, M.-Y. Liu, and Y. Taguchi. Automatic learning to remove multipath distortions in time-of-flight range images for a robotic arm setup. In *IEEE International Conference on Robotics and Automation*, 2016.
56. R. Stettner, H. Bailey, and S. Silverman. Three dimensional Flash LADAR focal planes and time dependent imaging. *International Journal of High Speed Electronics and Systems*, 18(02):401–406, 2008.
57. D. Stoppa, L. Pancheri, M. Scandiuzzo, L. Gonzo, G-F Dalla Betta, and A. Simoni. A CMOS 3-D imager based on single photon avalanche diode. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 54 (1):4–12, 2007.
58. C. Zhang and Z. Zhang. Calibration between depth and color sensors for commodity depth cameras. In *Proc. of the 2011 IEEE Int. Conf. on Multimedia and Expo*, ICME '11, pages 1–6, 2011.
59. Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.
60. J. Zhu, L. Wang, R. G. Yang, and J. Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *Proc. CVPR*, pages 1–8, 2008.