Time-of-Flight Camera – An Introduction

Larry Li   *Sensing Solutions*

1. Introduction

3D Time-of-Flight (TOF) technology is revolutionizing the machine vision industry by providing 3D imaging using a low-cost CMOS pixel array with an active modulated light source. Compact construction, easy-of-use, together with high accuracy and frame-rate makes TOF cameras an attractive solution for a wide range of applications. In this article, we will cover the basics of TOF operation, and compared TOF with other 2D/3D vision technologies. Then various applications that benefit from TOF, such as gesturing and 3D scanning/printing are explored. Finally, resources that help readers get started with the Texas Instruments 3D TOF solution are included.

2. Theory of Operation

A 3D time-of-flight (TOF) camera operates by illuminating the scene with a modulated light source, and observing the reflected light. Phase shift between the illumination and the reflection is measured and translated to distance. Figure 1 illustrates the basic TOF concept. Typically, the illumination is provided by a solid-state laser or a LED operating in the near-infrared (NIR) invisible to the human eyes. An imaging sensor designed to respond to the same spectrum receives the light and converts the photonic energy to electrical charges. Note that the light entering the sensor has both the ambient light component and the reflected component.
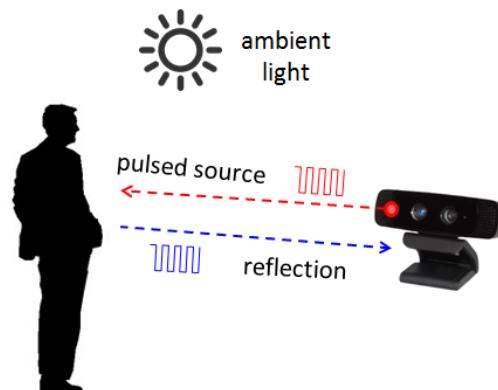

Figure 1: 3D time-of-flight camera operation.

In order to detect the phase shift between the illumination and the reflection, the illuminating source can be **pulsed** or modulated by a **continuous-wave (CW)**, typically sinusoid or square wave. Square wave modulation is popular because it can be easily realized with digital circuits [4].

Pulsed modulation can be achieved by integrating photoelectrons from the reflected light, or by starting a fast counter at the first detection of the reflection. The latter require a fast photo-detector, usually a *single-photo avalanche detector* (SPAD). This counting approach requires very fast electronics, since to achieve 1 millimeter accuracy require timing a light pulse of 6.6 picosecond in duration. This level of accuracy is nearly impossible to achieve in silicon at room temperature [1].
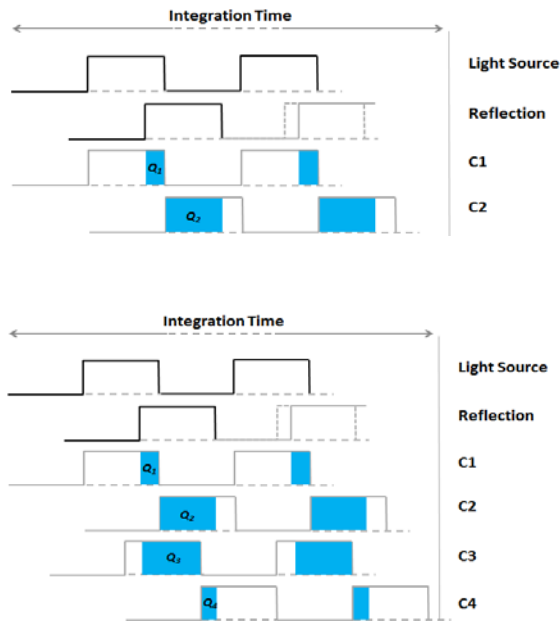
Figure 2: Two time-of-flight methods: pulsed (top) and continuous-wave (bottom).

The *pulsed* method is straightforward. The light source illuminates for a brief period of time (Δt), and the reflected energy is sampled at every pixel in parallel using two out-of-phase windows, C_1 and C_2, of the same Δt duration. Electrical charges accumulated during these samples, Q_1 and Q_2, are measured and used to compute distance using the formula:

$$d = \frac{1}{2} c \, \Delta t \left(\frac{Q_2}{Q_1 + Q_2} \right). \qquad \text{Eq. 1}$$

The CW method takes multiple samples per measurement, with each sample phase-shifted 90 degrees, for a total of four samples. Using this technique, the phase angle between illumination and reflection, φ, and the distance, d, can be calculated by

$$\varphi = \arctan \left(\frac{Q_3 - Q_4}{Q_1 - Q_2} \right), \qquad \text{Eq. 2}$$

$$d = \frac{c}{4\pi f} \varphi. \qquad \text{Eq. 3}$$

It follows that the pixel intensity amplitude (*A)* and offset **(***B)* can be computed by:

$$A = \frac{\sqrt{(Q_1 - Q_2)^2 + (Q_3 - Q_4)^2}}{2}, \qquad \text{Eq. 4}$$

$$B = \frac{Q_1 + Q_2 + Q_3 + Q_4}{4}. \qquad \text{Eq. 5}$$

In all equations above, **c** is the speed-of-light constant.

At first glance, the complexity of the CW calculations as compared to the pulsed method may seem unjustified, but a closer look at the CW equations reveals that the terms, $(Q_3 - Q_4)$ and $(Q_1 - Q_2)$ eliminates the effect of any constant offset from the measurements, such as that from the ambient light. Furthermore, the quotient of these in the phase equation reduces the effect of any constant gains from the distance measurements, such as circuit amplification and attenuation, or the reflected intensity.

The reflected amplitude (A) and offset (B) do have an impact the depth measurement accuracy. The depth measurement variance can be approximated by:

$$\sigma = \frac{c}{4\sqrt{2}\pi f} \cdot \frac{\sqrt{A+B}}{c_d A} \qquad \text{Eq. 6}$$

The modulation contrast, c_d, describes how efficient the imaging silicon separates and collects the photoelectrons. The reflected amplitude, A, is a function of the optical power. The offset, B, is a function of the ambient light. One may inferred from Equation 6 that high reflected amplitude, high modulation frequency and high modulation contrast will contribute to increased accuracy. High offset, on the other hand, can lead to saturation and reduces accuracy. At high frequency, the modulation contrast can begin attenuate due to the photoelectron mobility within the silicon. This puts a practical upper limit on the modulation

frequency.   TOF sensors with high roll-off frequency generally can deliver higher accuracy.

The fact that the CW measurement is based on phase, which wraps around every 2π, means the distance will also alias. The distance where the aliasing occurs is called the ambiguity distance, d_{amb}, and is defined as:

$$d_{\mathrm{amb}} = \frac{c}{2f} \qquad \textbf{Eq. 7}$$

Since the distance wraps, d_{amb} is also the maximum measurable distance. If one wishes to extend the measurable distance, one may reduce the modulation frequency, but at the cost of reduced accuracy, as is according to Equation 6.

Instead of accepting this compromise, advanced TOF systems deploy multi-frequency technique to extend the distance without reducing the modulation frequency.   Multi-frequency techniques work by adding one or more modulation frequencies to the mix.   Each modulation frequency will have a different ambiguity distance, but true location is the one where the different frequencies agree.   The dual-frequency concept is illustrated below.
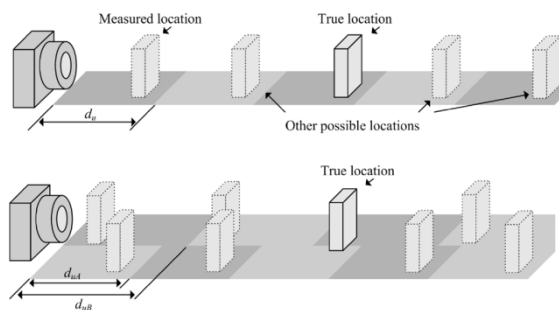
Figure 3: Extending distance using a multi-frequency technique.

3. Point Cloud

In 3D TOF cameras, distance is measured for every pixel in an addressable array, resulting in a ***depth map***. A depth map is a collection of 3D points (each point also known as a ***voxel***). As an example, a QVGA sensor will have a depth map of 320 x 240 x 3 voxels.   2D representation of a depth map is a gray-scale image, as is illustrated by the soda cans example in Figure 4—the brighter the intensity, the closer the voxel. Figure 4 shows the depth map of a group of soda cans.

Figure 4: Depth map of soda cans.

Alternatively, a depth map can be rendered in a three-dimensional space as a collection of points, or ***point-cloud***.   The voxels can be mathematically connected to form a mesh for texture mapping. The texture can come from a color image of the same subject, creating a life-like 3D avatar, as is illustrated in Figure 5.
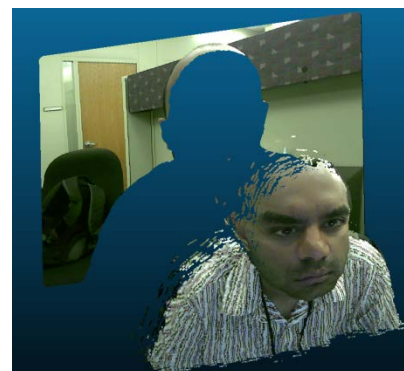

Figure 5: Avatar formed from point-cloud.

4. Other Vision Technologies

Time-of-flight technology is not the only 3D vision technology available. In this section, we will compare TOF versus 2D and other 3D vision technologies. A table summarizing the comparison is included at the end of this section.

2D Vision

Most machine vision systems deployed today are based on 2D vision, usually a single digital camera. Such approach is very effective for planar applications where the lighting is closely controlled, such as automated inspections. In these applications, defects are usually detected using well-known 2D image processing techniques, such as edge detection, template matching and morphology open/close , to extract important features that can be compared to a database for pass-fail determination. To detect defect along the z-axis, additional 1D sensor or 3D vision are often used.

2D vision could be used in unstructured environment as well with the aid of advanced image processing algorithms to get around complications caused by varying illumination and shading conditions. Take the images in Figure 6 for example--these images are from the same face, but under very different lighting conditions. The shading difference can make face recognition difficult even for humans. In contrast, computer recognition using point cloud data from TOF sensors is unaffected by shading, since illumination is provided by the TOF sensor itself, and the depth measurement is derived from the phase, not intensity.
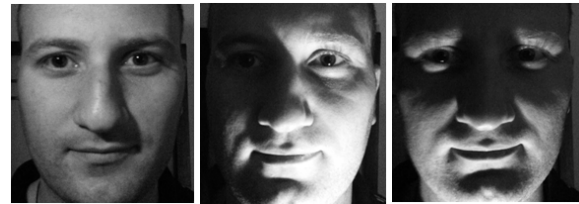

Figure 6: Same face, different shading.

3D Vision

Robust 3D vision overcomes many problems of 2D vision, as the depth measurement can be used to easily separate foreground from background. This is particularly useful for gesture recognition, where the first step is to segment the parts of the subject relevant to gesture recognition from those that are irrelevant.
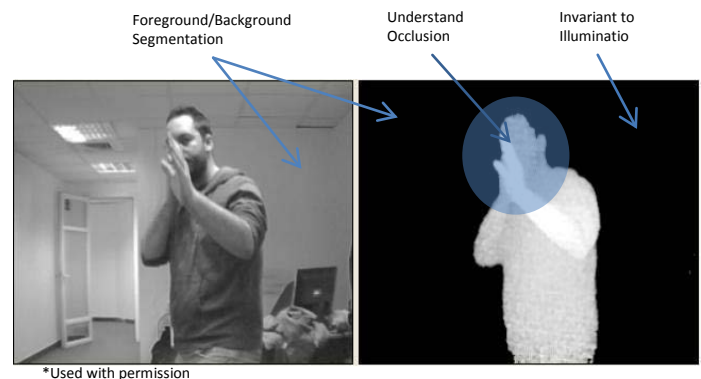

Figure 7: Advantages of 3D vision over 2D.

In the next two subsections we will compare the TOF technology with two other 3D vision technologies: *stereo vision* and *structured-light*.

➢ **Stereo Vision vs. TOF**

Stereo vision generally uses two cameras separated by a distance, in a physical arrangement similar to the human eyes. Given a point-like object in space, the camera separation will lead to measurable disparity of the object positions within the two camera

images. With a simple pin-hole camera model, the object position in each image can be computed, which we will represented them by α and β. With these angles, the depth, z, can be computed.

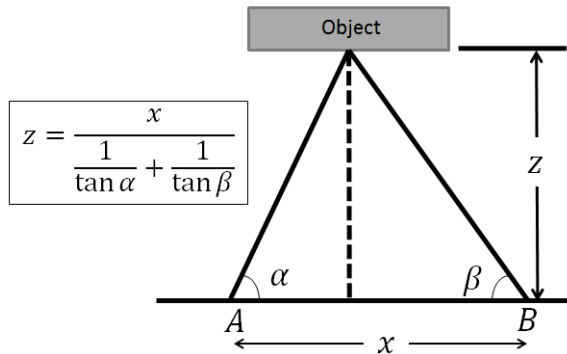$$z = \frac{x}{\frac{1}{\tan \alpha} + \frac{1}{\tan \beta}}$$

Figure 8: Stereopsis--depth through disparity measurement.

A major challenge in stereo vision is solving the **correspondence problem**: *giving a point in one image, how to find the same point in the other camera?* Until the correspondence can be established, disparity, and therefore depth, cannot be determined. Solving the correspondence problem require complex, computationally intensive algorithms for feature detection and template matching. Feature extraction and matching also require sufficient intensity or color variation in the image. This requirement render stereo vision less effective if the subject lacks these variations—for example, measuring the distance to a uniformed colored wall. TOF does not have this limitation because it does not depend on color or texture to measure distance. In stereo vision **depth resolution error is a quadratic function of the distance**; by comparison, TOF, which works off the reflected light, is also sensitive to distance. However, for TOF this shortcoming is remedied by increasing the illumination energy as necessary.

Stereo vision has some advantages. It is relatively **affordable to implement**, as most common off-the-shelf cameras can be used. Also, due to its human-like physical configuration, stereo vision is **well-suited for capturing images for stereoscopic presentation to humans,** so that both humans and machine can be looking at the same images.

➢ **Structured-Light vs. TOF**

Structure-light works by projecting known patterns onto the subject and inspecting the pattern distortion. Sequential projection of coded or phase-shifted patterns is often required to extract a single depth frame, which leads to **lower frame rate**. Low frame rate means the **subject must remain relatively still** during the projection sequence to avoid excessive blurring. The reflected pattern is sensitivity to optical interference from the environment, therefore, structured-light tends is more suited for indoor application under controlled environment. A major advantage of structured-light is that it can achieve very high spatial (X-Y) resolution by using off-the-shelf DLP and HD cameras. **Error! Reference source not found.** shows the structured-light concept.
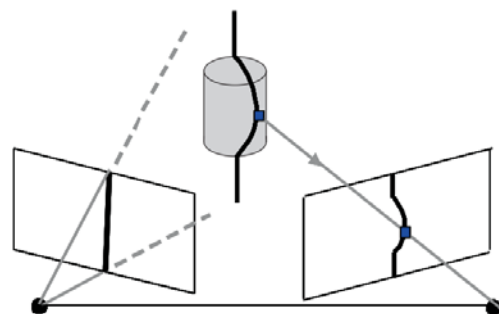

Figure 9: Structured light concept.

By comparison, TOF is less sensitivity to environmental lighting condition, more technology has lower resolution, but it is rapidly improving. The comparison of TOF camera with stereo vision and structured-light is summarized in Table 1. The key takeaway is that **TOF is cost-effective, mechanically compact depth imaging**

mechanically compact.   The current TOF

system that is generally unaffected by varying environmental illumination and requires almost no processing to extract point cloud data. This combination makes TOF well-suited for a wide range of application.

Table 1: Comparison of 3D Imaging Technologies

CONSIDERATIONS	STEREO VISION	STRUCTURED-LIGHT	TIME-OF-FLIGHT (TOF)
Software Complexity	High	Medium	Low
Material Cost	Low	High	Medium
Compactness	Low	High	Low
Response Time	Medium	Slow	Fast
Depth Accuracy	Low	High	Medium
Low-Light Performance	Weak	Good	Good
Bright-Light Performance	Good	Weak	Good
Power Consumption	Low	Medium	Scalable
Range	Limited	Scalable	Scalable
APPLICATIONS			
Game		X	X
3D Movies	X		
3D Scanning		X	X
User Interface Control			X
Augmented Reality	X		X

5. Applications

There is a wide range of applications that benefits from the 3D-TOF technology—from automotive to industrial to healthcare, to smart advertising, gaming and entertainment. 3D-TOF could also serve as excellent input device to both stationary and portable computing devices.   In automotive, the TOF sensor could enable autonomous driving and increased safety. In industrial, the TOF sensor could be used as HMI and for enforcing safety envelopes, so that humans and robots can work in close proximity. In smart advertising, digital signage could be highly interactive, targeting media contents to the specific audience.   In

healthcare, gestures offer non-contact human-machine interactions, creating a more sanitary operating environment. The gesture capability of TOF sensing is particularly well-suited for the consumer electronics, particularly in gaming, portable computing, and home entertainment. The TOF's natural interface provides an intuitive gaming interface especially for first-person video games. This same interface could also replace the remote controls, mice and touch screens.   Generally speaking, the TOF applications can be categorized into Gesture and Non-Gesture. Gesture based applications emphasize human interactions, while non-

gesture based applications emphasize on   measurements.

Figure 10: TOF technology applies to a wide range of applications.

Gesture based Applications

Gesture based applications translate human movements (faces, hands or whole-body) into symbolic directives to enable natural interactions with gaming consoles, smart televisions, or portable computing devices. For examples, channel surfing can be done by waving of hands, presentation slides can be scrolled through finger pointing.   These applications usually require fast response time, low- to medium-range, centimeter-level accuracy and power consumption.

Figure 11: Gesture recognition using a 3D-TOF sensor.

Non-Gesture based Applications

3D-TOF technology can be used in non-gesture based applications as well. For instance, in *automotive*, TOF camera can increase safety by alerting the driver when it detects people and objects in the vicinity of the car, and in computer assisted driving. In *robotics and automation*, TOF camera can help detect product defects and enforce safety envelopes required for humans and robots to work in close proximity. With *3D printing* rapidly becoming popular and affordable, TOF cameras can be used to perform *3D scanning* to enable "3D copying" capability. In all of these applications, spatial accuracy is important.

* Used with permission

6. How to Get Started

Hardware

Texas Instruments 3D-TOF solution is based on the CW method, but has added many enhancements for power reduction and increased accuracy. The current generation is a 3-chip solution, comprised a TOF imaging sensor (OPT81x0), the analog front-end (VSP5324) and a TOF controller (OPT91xx). The solution is available also as a camera development kit (CDK). Figure 12 shows the system block diagram.
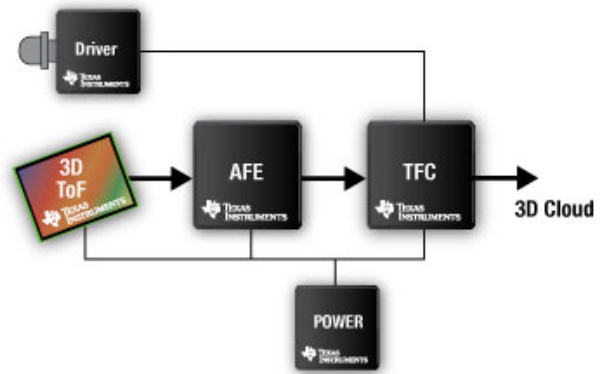
Figure 12: TI 3D-TOF chip set.

3D TOF Sensor Array

The 3D TOF sensor is an addressable CMOS pixel array based on the CAPD technology [2], which supports high pixel modulation frequency (>50MHz), and up to 5x increase in signal-to-noise ratio (SNR). Through advanced silicon fabrication process, the pixel device material is carefully tuned to respond to specific optical spectrum (850nm-870nm) and to optimize the photoelectron collection and transport to the readout channels.

Analog Front-End (AFE)

The AFE (VSP5324) supports up to 4 differential inputs, each with a sample-and-hold front-end that also helps reject common-mode noise. A high-speed, low-power 12-bit ADC samples the input at up to 80 MSPS, and delivers the data serially over two differential serial data channels.

TOF Controller (TFC)

The TFC is a sophisticated state machine that synchronizes the operation of 3D TOF sensor, AFE and illumination. It scans the pixels and

calculates the depth per each pixel and performs the required de-aliasing, de-noising, frequency tuning and temperature compensation. It also handles high-rate input/output serialization/deserialization.

Point cloud data from the TFC can be accessed by external hosts through a variety of ways. For webcam-like applications, the TFC can interface to a USB controller, such as the FX2USB, and enumerate as a USB Video Class camera under the host's operating system. Alternatively, the TFC may interface with an embedded processor, such as OMAP™ and Sitara™ through a native camera interface or a parallel bus interface.

To support a wide range of applications demanding different illumination power, the power subsystem (PMIC) and laser/LED driver are intentionally kept as external components. To complete the solution, Texas Instruments also offers a wide selection of PMIC and LED devices. For more details on the TI 3D-TOF solution, please visit http://www.ti.com/3dtof.

Software

There are several software vendors that provide state-of-the-art gesture recognition libraries that will work with the point cloud generated by the TI solution. There are also open-source libraries available that have a rich set of features. The URLs for these software resources are listed below:

SoftKinetic: http://www.softkinetic.com

Omek Interactive: http://www.omekinteractive.com

Point-Cloud Library: http://www.pointclouds.org

OpenCV: http://www.opencv.org

7. Summary

In this article, we introduced the 3D-TOF technology, discussed the theory of operation, compared TOF with other 2D and 3D vision technologies and explained TOF's differentiated advantages. We also explored a wide range of applications which TOF enables, including those requiring gestures and those that do not. To help readers get started with TOF, we introduced Texas Instruments' 3D-TOF chipset and CDK, as well as third-party software resources.

References

[1] E. Charbon, M. Fisburn, R. Walker, R. Henderson, C. Niclass, *SPAD-Based Sensors*.

[2] D. Van Nieuwenhove, W. Van der Tempel, R. Grrotjans, M. Kuijk, "Time-of-flight Optical Ranging Sensor based on a Current Assisted Photonic Demodulator", *Proceedings Symposium, IEEE/LEOS Benelux Chapter*, 2006, Eindhoven.

[3] J. Geng, "Structured-Light 3D Surface Imaging: a Tutorial", *Advances in Optics and Photonics* 3, 128-160 (2011).

[4] M. Hansard, S. Lee, O. Choi, R. Horaud, "Time-of-Flight Cameras: Principles, Methods and Applications", *Springer Brief in Computer Science*, 2012.

IMPORTANT NOTICE

Products		Applications	
Audio	www.ti.com/audio	Automotive and Transportation	www.ti.com/automotive
Amplifiers	amplifier.ti.com	Communications and Telecom	www.ti.com/communications
Data Converters	dataconverter.ti.com	Computers and Peripherals	www.ti.com/computers
DLP® Products	www.dlp.com	Consumer Electronics	www.ti.com/consumer-apps
DSP	dsp.ti.com	Energy and Lighting	www.ti.com/energy
Clocks and Timers	www.ti.com/clocks	Industrial	www.ti.com/industrial
Interface	interface.ti.com	Medical	www.ti.com/medical
Logic	logic.ti.com	Security	www.ti.com/security
Power Mgmt	power.ti.com	Space, Avionics and Defense	www.ti.com/space-avionics-defense
Microcontrollers	microcontroller.ti.com	Video and Imaging	www.ti.com/video
RFID	www.ti-rfid.com		
OMAP Applications Processors	www.ti.com/omap	**TI E2E Community**	e2e.ti.com
Wireless Connectivity	www.ti.com/wirelessconnectivity		