A survey of modern 3D depth sensors

You-Sin Lin
National Taiwan University

Abstract

Statement: This report does not have any double assignment or any completed work before this semester.

I surveyed several sensors of their sensing principles and properties on the market. I will introduce the sensing principles of different sensors, compare their properties, and give several example of tasks that those sensors can be utilized. Finally, introduce an new application of structured-light method — Single-Photon Structured-Light.

1. Introduction

Since 3D sensing is becoming more and more popular nowadays, we often see 3D sensors being utilized in different devices. And there are many kinds of sensors with different sensing principles and properties. In this section, I will introduce what is 3D sensing, where is 3D sensing utilized and why selection of sensors is an important problem.

1.1. What is 3D sensing

3D sensing refers to the process of using 3D sensor to convert the depth data of a physical object into numerical data. Those data can then be used to reconstruct 3D point clouds or bounding boxes.

1.2. Where is 3D sensing utilized

There are many industries that utilized 3D sensing nowadays, such as self-driving car, medical, and Face-ID on our mobile device. And different sensors are used in different industries due to different advantages and disadvantages. For example, self-driving cars use LiDAR, and the medical industry uses structured-light.

1.3. Why is selection of sensors an important problem

Each 3D sensing task requires a different working distance, accuracy, and sensor size, etc. So, different sensors are utilized in different tasks, and the selection of sensors is an important problem.

2. Related work

2.1. Modern 3D depth sensors

There are several types of modern 3D sensors: **Passive Visual** (Camera), **LiDAR**, **Radar**, **Ultrasonic**, **GPS** (Global Positioning System), **IMU** (Inertial Measurement Unit). And we will focus on passive visual and LiDAR sensors in this report.

2.2. Passive Visual: ToF (Time of Flight) method

Shoot an laser or infrared at an object and measure the round trip time. Thus, the distance is speed of light times the round trip time divided by 2. ToF method has several features:

- **Real-Time:** ToF sensors can get depth data without triangulation, so ToF method has lower latency than Stereo-Vision and Structured-Light method.
- **Error $>$ 1 mm:** The error of ToF method is mainly caused by hardware constraint and usually greater than 1 mm.

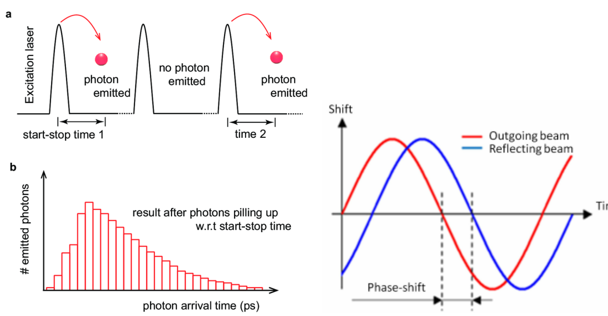
According to different emitter and receivers of the sensor, ToF method is subdivided into **dToF** (direct ToF) and **iToF** (indirect ToF).

- **dToF**: Emit multiple laser pulses in a short time and receive with SPAD (Single Photon Avalanche Diode) sensors. SPAD sensors can give a signal at the moment of receiving a photon, unlike the CMOS sensor, which needs to accumulate over a threshold for a period of time before there is a signal. And dToF sensors use TCSPC (Time Correlated Single Photon Counting) to count the time of receiving photons, and use the time with the largest number as the time of flight and calculate the depth. Figure 1a shows the TCSPC process.
- **iToF**: Emit frequency modulation continuous infrared and receive with CMOS sensors. And calculate time based on phase angle difference between emit and receive, so iToF is also called phase-based ToF. Figure 1b shows the schematic diagram of phase angle

difference. This is the equation of how iToF get depth data:

$$d = \frac{1}{2} \cdot \frac{C \cdot \phi}{2\pi \cdot f}, \tag{1}$$

Where d is the distance of object, C is speed of light, ϕ is the phase angle difference between emit and receive , finally f is the modulation frequency.



(a) TCSPC (b) Phase angle difference

Figure 1. dToF / iToF features

As we can see from the above equation (1) that only ϕ and f will affect the error, while ϕ is determined by the receiver and f is determined by the emitter. So, the source of error of dToF and iToF is mainly caused by hardware constraint, such as emitter, receiver and clock time. In dToF, they are laser, SPAD, TDC (Time to Digital Converter) respectively, while in iToF is emitter circuit modulation accuracy, CIS (Camera Image Sensor) and clock.

Finally, I compared the properties of dToF and iToF as shown in the Table 1. You can think of dToF as a LiDAR, and iToF as a normal ToF camera, so you can better understand this table. Distance ambiguity means that different depths may calculate the same phase angle difference, resulting in the same depth, which only occurs in iToF.

	dToF	**iToF**
Working Distance	Longer	Shorter
Precision	Limit to pulse period only	Lower when distance increase
Error	Laser, SPAD, TDC	Clock, Emitter, CIS
Distance Ambiguity	No	Yes
Space Resolution	Smaller	Larger
Anti-interference	Stronger	Weaker
Power	Lower (emit pulse)	Higher (emit continuous light)
Cost	Higher	Lower

Table 1. dToF / iToF comparison

The following Table 2 shows the properties of an iToF camera available on the market —LUCID Helios 2, which will be compared at Section 3.

2.3. Passive Visual: Stereo Vision method

Shoot the object with two calibrated cameras in different poses at the same time, and get the depth data by triangulation. Stereo Vision-Based method has several features:

Resolution	**Accuracy (up to 0.5 m)**	**Working distance**	**Dimensions**	**Sensor size**	**Power consumption**	**Price**
640 x 480 (30FPS)	± 3.6 (mm)	0.3–8.3 (m)	60 x 60 x 77.5 (mm)	$\frac{1}{2}$"	< 12W	$ 1495

Table 2. LUCID Helios 2

- **Triangulation-based:** With the corresponding feature points in the images captured by two different pose cameras, and calibration parameters for cameras, we can get the depth data of the object. Because each feature point on the images corresponds a line in 3D space. And the point where the two lines represented by corresponding feature points intersect is the 3D coordinates of the object.
- **Higher measurement accuracy than ToF method:** Since Stereo-Vision method uses 2 cameras, the resolution is obviously higher than of the ToF sensor, so the accuracy is also higher.
- **Performance depends on the richness of the scenario features:** When the number of feature points in the scene decreases, the number of corresponding feature points also decreases, thus the performance will decrease either.

The following Table 3 shows the properties of an Stereo-Vision method camera available on the market —Stereolabs ZED 2i, which will be compared at Section 3.

Resolution	**Accuracy (up to 0.5 m)**	**Working distance**	**Dimensions**	**Sensor size**	**Power consumption**	**Price**
1920 x 1080 (30FPS) / 662 x 376 (100FPS)	± 5 (mm)	0.3–20 (m)	175.25 x 30.25 x 43.1 (mm)	$\frac{1}{3}$"	< 1.9W	$ 499

Table 3. Stereolabs ZED 2i

2.4. Passive Visual: Structured-Light method

Project the pattern on the object by projector, and capture it with camera. Then extract the feature points from the image and correspond to the feature points in the projector image. Finally, obtain the point cloud by triangulation. The following are several features of Structured-Light method:

- **Triangulation-based**
- **Less dependent on scenario features than Stereo Vision method:** Since Structured-Light method project encoded patterns on the object, it has more feature points, better robustness to ambient light than classical Stereo-Vision method.
- **High accuracy, Short working distance:** In order to clearly identify patterns, the working distance of structured light is usually short, but with self-defined patterns, this method can achieve higher depth accuracy than classical Stereo-Vision method.

Table 4 shows the properites of an Structured-Light method camera on the market — Astra Mini, which will be compared at Section 3.

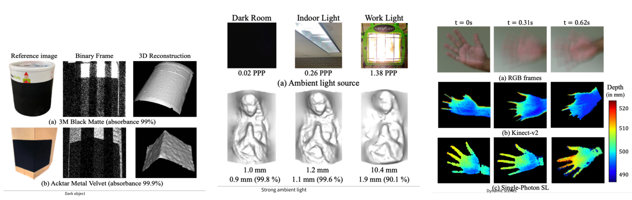
Resolution	Accuracy (up to 0.5 m)	Working distance	Dimensions	Power consumption	Price
640 x 480 (30FPS)	± 1∼3 (mm)	0.6 ∼5 (m)	80 x 20 x 20 (mm)	< 2.4W	$ 170

Table 4. Astra Mini

Next, I will briefly introduce the newly Structured Light application — **Single-photon structured light** proposed by Varun Sundar et al [1].

They replaced the structured light camera with a SPAD sensor array, and used the DMD projector to project the image line encoded pattern so that the frequency reaches 20k Hz. Achieving a result that cannot be achieved by the original structured light method. In their experiments, they overcome the three situations where the original structured light method cannot effectively identify the depth:

- **Dark objects:** Because the SPAD sensor array is used, as long as the sporadic reflected photons are received, the pattern can be obtained to calculate the depth. Their experiment result on dark object is shown in Figure 2a.
- **Strong ambient light:** With their Hybrid encoded patterns, they are more robust to ambient light than BCH encoded patterns. The result is shown in Figure 2b.
- **Dynamic scenes:** With high-speed sensor array and DMD projector, 20,000 images can be obtained per second, which can effectively reconstruct the depth of dynamic scenes. Their experiment result on dynamic scenes is shown Figure 2c.



(a) Dark object (b) Strong ambient light (c) Dynamic scenes

Figure 2. Single-photon structured light result

2.5. LiDAR

Finally, LiDAR uses the aforementioned dToF method, emits multiple laser beams horizontally across the scene to obtain depth data. And LiDAR has large working distance but is expensive. Table 5 shows the properites of an LiDAR on the market — SICK MRS6224R, which will be compared at Section 3.

Aperture angle	Angular resolution	Accuracy	Working distance	Dimensions	Power consumption	Price
120° x 15° (10FPS)	0.13° (H), 0.625° (V)	± 125 (mm)	0.5 ∼200 (m)	125.5 x 176 x 131 (mm)	20 W	$ 11928

Table 5. SICK MRS6224R

3. Comparison

Before starting the comparison, first introduce the properties used to compare various 3D sensors, including **resolution**, **accuracy**, **working distance**, **dimension** (size) , **power** and **price**.

The comparison is shown in Table 6. And the comparison radar graph is shown in Figure 3.

Property / Method	Resolution	Accuracy (up to 0.5 m)	Working distance	Dimensions	Power consumption	Price
ToF	640 x 480 (30FPS)	± 3.6 mm	0.3 ∼8.3 (m)	60 x 60 x 77.5 (mm)	<12W	$ 1495
Stereo Vision	1920 x 1080 (30FPS) / 662 x 376 (100FPS)	± 5 mm	0.3 ∼20 (m)	175.25 x 30.25 x 43.1 (mm)	<1.9W	$ 499
Structured-Light	640 x 480 (30FPS)	± 1∼3 mm	0.6 ∼5 (m)	80 x 20 x 20 (mm)	<2.4W	$ 170
LiDAR	924 x 64 (PC)	± 125 mm	0.6 ∼200(m)	125.5 x 176 x 131 mm	20 W	$ 11928

Table 6. 3D depth sensor comparison

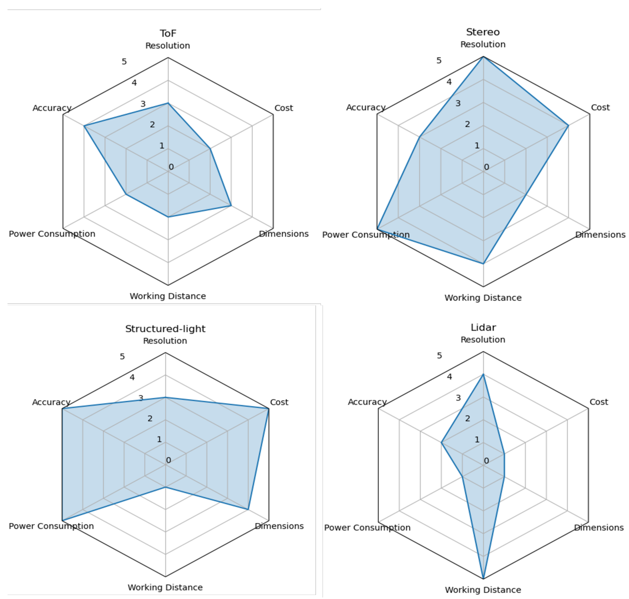


Figure 3. Comparison radar graph

First of all, the Stereo Vision camera has the best resolution because it can be implemented with two cameras. Next, the accuracy is compared to the size of the error at 0.5m. It can be seen that the error of Structured Light is the smallest, while the error of LiDAR is dozens of times that of other methods. In terms of working distance, LiDAR is far ahead of other methods and can reach 200 meters. Then there is the dimension comparison. The volume of the Structured Light camera is the smallest, while the volume of the LiDAR is significantly larger than other methods. In terms of power consumption, because Stereo Vision and Structured Light methods do not need to emit infrared rays or lasers, the power consumption is significantly lower than the other two methods. Finally, the price of structured light is the lowest when the resolution is similar.

4. Conclusion

Among ToF, Stereo Vision, Structured Light, LiDAR 3D sensors, ToF is the most basic 3D sensor. Although Stereo Vision is excellent in all aspects, its performance is limited by the number of features. Structured Light has higher accuracy and lower energy consumption, but its disadvantage is its short working distance. Finally, LiDAR has the advantage of having the longest working distance.

References

[1] Varun Sundar, Sizhuo Ma, Aswin C Sankaranarayanan, and Mohit Gupta. Single-photon structured light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17865–17875, 2022. 3