Introduction to Depth Sensing

Harish Venkataraman

Depth Cameras Arch & Tech Lead, Cameras and Depth, TED

Meta Inc.

Outline

- What is depth sensing?
- 3D sensing applications
- Technologies for 3D sensing
- Technological challenges/comparison
- Imaging for the metaverse

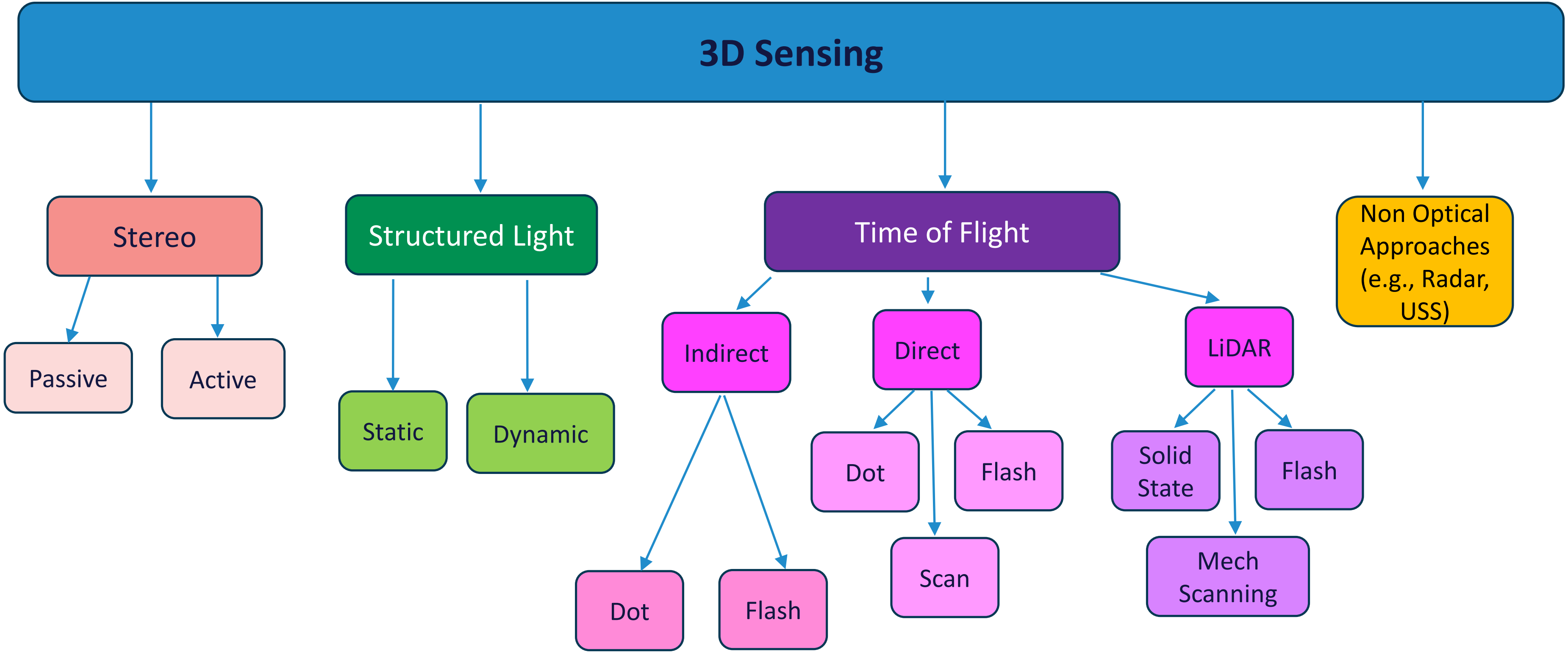
Introduction to Depth Sensing

- Depth sensing enables machines to perceive our environment in 3 dimensions
  - They see the environment around them in the way humans do
- Depth sensing is the act of measuring distance to objects, obstacles in the scene to generate 3D maps of the scene
- Specialized cameras/sensors are used for depth sensing
- Enables real time choices, decisions, experiences in many applications

Depth Sensing Applications

- Depth sensing & cameras are used in a variety of applications:
  - Smart phones (authentication, room scanning, low light auto focus)
  - Industrial applications (robots)
  - AR/VR applications
  - Autonomous vehicles
  - Agriculture (harvesting and weeding robots)
  - Home (robot vacuum cleaners)
  - Real estate (3D scanning)

The Tree of Light



Passive Stereo

- A target is imaged by two cameras
  - Now triangulation is between 2 images rather than illuminator and camera
- Disparity in pixel position imaging the object is calculated

Range estimate:

$$r = \frac{f \cdot b}{\|P1 - P2\|}$$

Precision estimate: (not SNR limited)

$$\Delta r = \frac{r^2}{f \cdot b} \Delta d$$

Key parameters:

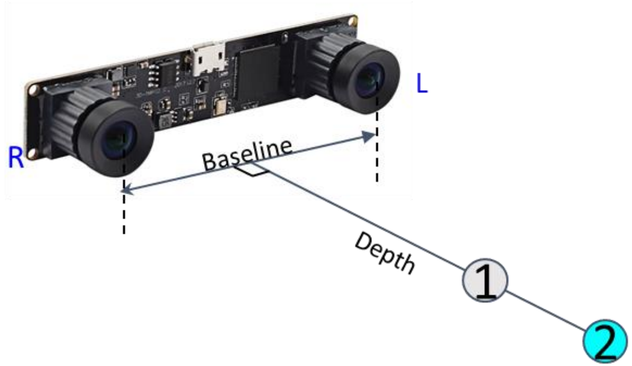
$\|P1 - P2\| = disparity\ (\mu m)$

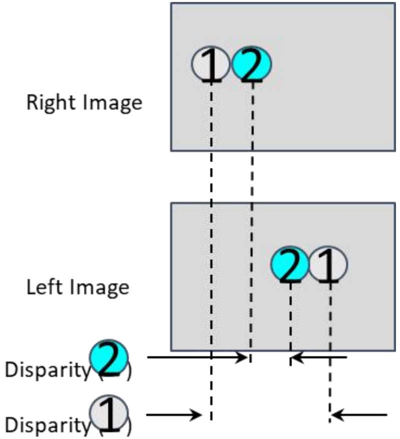
$f = focal\ length\ (mm)$

$b = baseline\ (mm)$

$r = range\ (m)$

$\Delta d = minimum\ disparity (\mu m)$





To minimize range errors:

- Space pixels apart
- Increase focal length
- Decrease pixel size
- Precisely co-register points on object

Active Stereo Depth Imaging

- In order to co-register features, a light pattern may be projected onto the FoV
  - This makes it possible to range feature-less or periodic objects
  - E.g., Walls, desk, floor, wallpaper
- Algorithmic improvements such as ML will improve performance
- Comes with computational burden

Precision estimate (SNR limited)

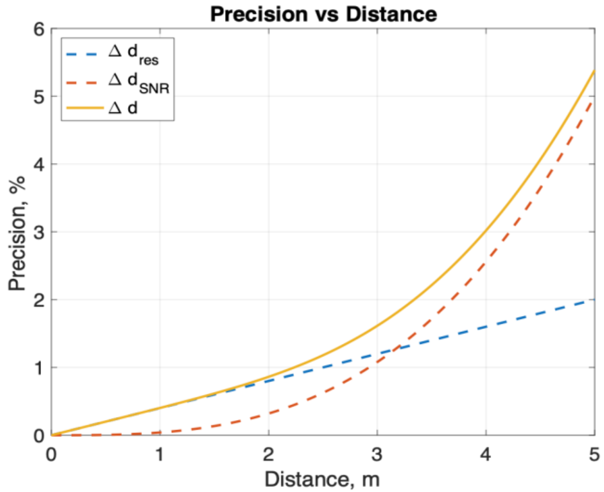
$$\Delta r = \frac{r^2}{f \cdot b}\sqrt{\Delta d_{res}^2 + \Delta d_{SNR}^2}$$

Constant: depends on pixel and algorithm

SNR dependent: Depends on range, ambient power, etc.

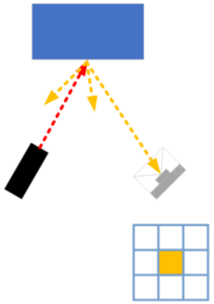
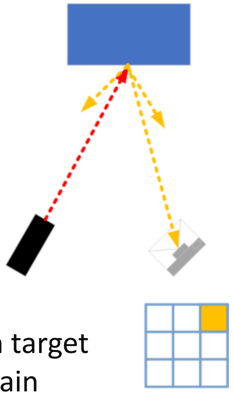
Key parameters

$$\|P1 - P2\| = disparity\ (\mu m)$$
$$f = focal\ length\ (mm)$$
$$b = baseline\ (mm)$$
$$r = range\ (m)$$
$$\Delta d = minimum\ disparity (\mu m)$$

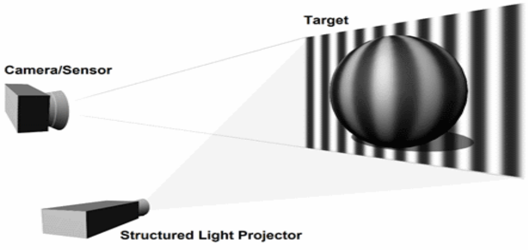


Structured Light (Static)

- A target is illuminated by alight source with a known pattern
- The angle in which it is viewed by camera depends on its position wrt illuminator and sensor
- Challenge is to uniquely localize the region in the object being viewed by each pixel
- Light and dark patterns create registration marks to improve this registration

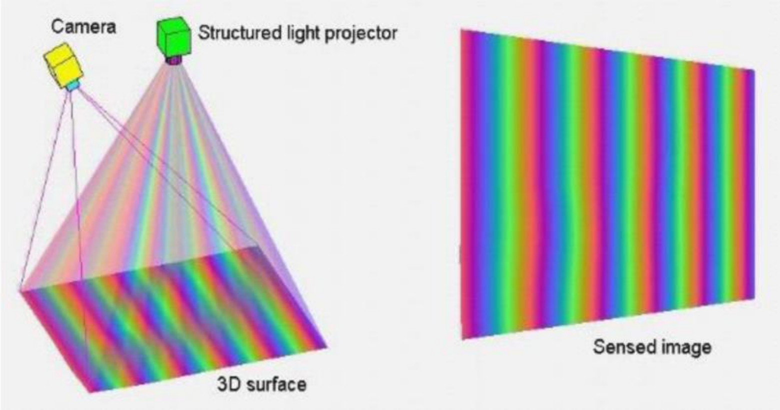
Laser light backscattered from target → detected by camera at certain pixel

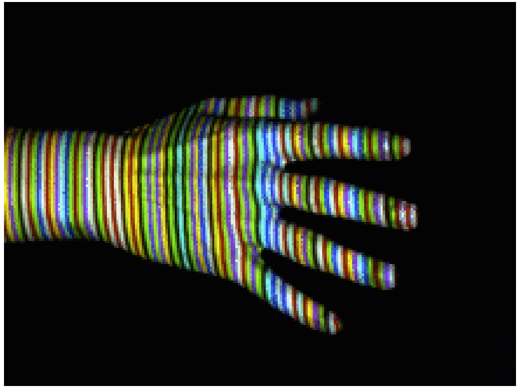


Localize the region of the object being imaged by illumination with a pattern. Tighter the pattern, better the ranging accuracy.

Improved Structured Light (Dynamic)

- Feature identification achieved by labeling the pattern, e.g., with color
- More complex light source and less sensitive detection
- Utilizing coded patterning can improve feature localization
- Sequential acquisition → motion artifacts, computational complexity and higher power consumption
- Performance at range quickly deteriorates due to beam divergence and quick fall-off of returning signal

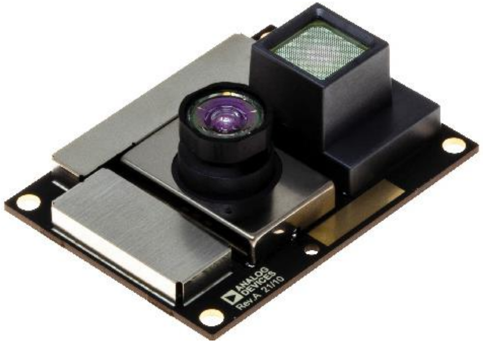




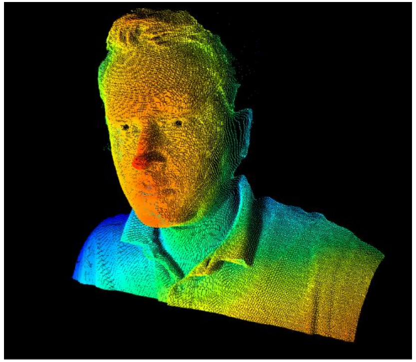
Source: Jason Geng, "Structured-light 3D surface imaging: a tutorial," Adv. Opt. Photon

Depth via Indirect Time-of-Flight (iToF)

- All ToF sensors use the fact that $d = \frac{ct}{2}$
- In iToF, the field is illuminated by RF-amplitude-modulated light
- By proper gating of a CCD device, one can calculate the phase shift of returning light, and from there the time-of-flight and distance
- In-pixel circuitry enables high-resolution 3D point clouds to get acquired



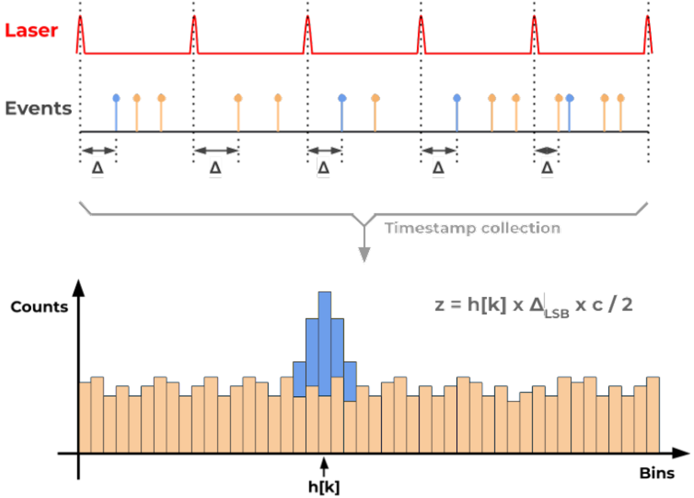
Analog Devices iToF Module

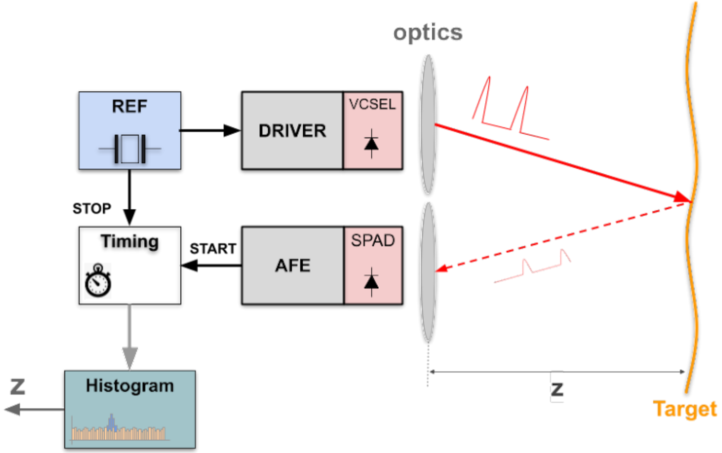


540k pixel ST iToF Sensor

Depth via Direct Time-of-Flight (dToF)

- Operates by sending pulses of light and measuring their round-trip time-of-flight
- Requires very short laser pulses, precise photon time-of-arrival detector (SPAD), circuitry to process timing info
- Generating ToF histograms is challenging and consumes a large silicon real estate
- Recent advances in 3D stacking technologies resulted in the first commercial dToF sensors





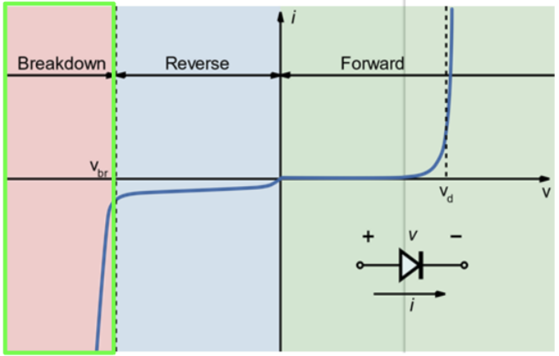
Meta

SPAD as ToF Detector

- Generates a digital pulse at the time of each incoming photon
- Unique in its ability to measure time-of-arrival of photons
- Digital detection means no read noise penalty & direct processing of individual events
- Ability to disambiguate and measure single photons
- No read noise enables extreme low light, digital measurement removes limitations of Full Well Capacity

Single incoming photon generates an avalanche of electrons → counted as a digital trigger.

Current is quenched and device is reset to detect the next photon



	Description	Typical Values
PDE	Probability to detect an event. Analogous to QE	10-35%
Deadtime	Time required to reset SPAD after event is detected	5-10 ns
VBD+Vex	Excess (Vex) supply voltage required above breakdown (Vbd). Drives array power	-21 V

Depth Technologies Comparison Matrix

embedded VISION SUMMIT

	Active Stereo	Structured Light	iToF	dToF
Range	Low	Low	Range & resolution need to be traded off due to low dynamic range	High
Resolution	Medium	High		Low to med; High possible as the SPAD pixel evolves over time
Point Cloud Quality	Poor	High	Medium	High
Power	High	High	High	Low
SoC/ISP Compute	High	Medium	High: Needs ISP	Low: On sensor processing
Cost	High (2 cameras + 1 projector)	Medium (1 camera + 1 projector)	Medium to high (depends on range)	Low to medium
Solution Size	Large	Large due to baseline requirements	Large for high range	Small

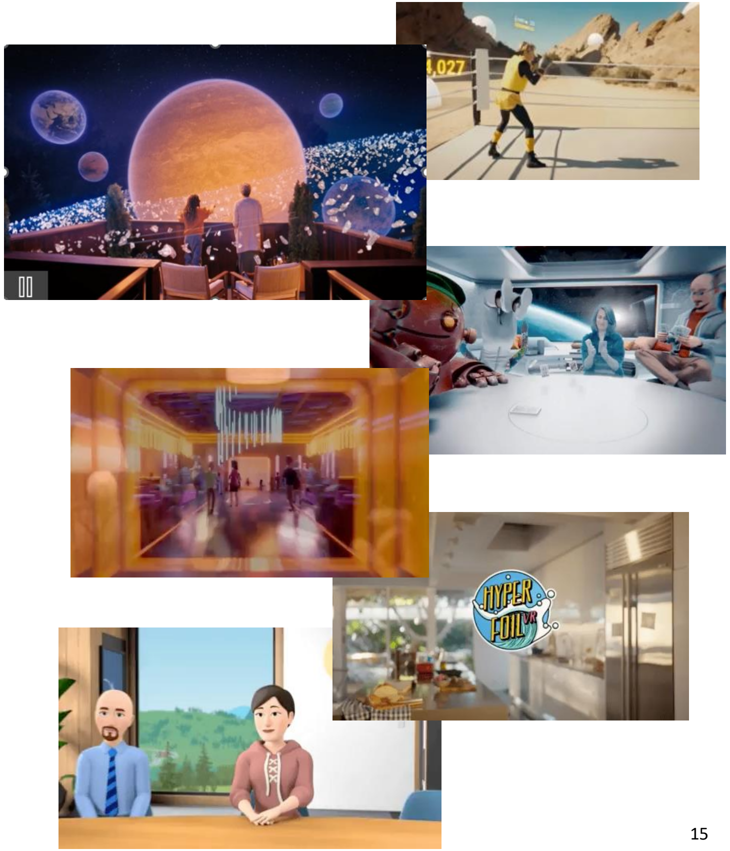
Meta

Depth Technologies Trends: Example Applications

	Active Stereo	Structured Light	iToF	dToF
AR & VR Systems	X	X	X	X
Healthcare & Dental		X		X
Surveillance			X	X
Authentication		X		
Autonomous Vehicle Navigation	X		X	X
Robotics/Vacuum Cleaners	X		X	X
Room Scanning				X
Photography	X			X

Summary

- There isn't one perfect solution
  - A system that already needs 2 RGB cameras might prefer a stereo solution
  - A self driving car enabling technology chooses time of flight
- dToF came about as a way to mitigate iToF limitations, offers a lot of positives but SPAD pixel is large
- Numerous interdisciplinary technological innovations still need to be developed – @ Meta, we see depth sensing as one key enabling technology in AR/VR



Resource Slide

- https://aivero.com/overview-of-depth-cameras/
- https://3d.pmdtec.com/en/ecosystem/blog/what-depth-sensing-technology-is-best-for-your-project/
- https://www.e-consystems.com/blog/camera/technology/what-are-depth-sensing-cameras-how-do-they-work/

Meta Inc links:

- https://developers.facebook.com/blog/post/2023/04/25/presence-platform-overview/
- https://developer.oculus.com/documentation/unity/unity-depthapi/